

Survival Analysis of Patients with Recurrent Bladder Cancer

Srijan Mallick

Dipan Banik

(under the supervision of Dr. Sudipta Das)

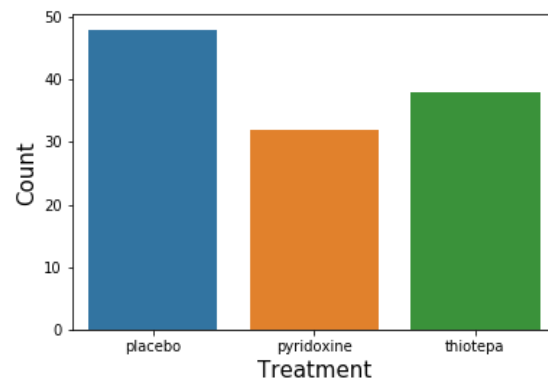
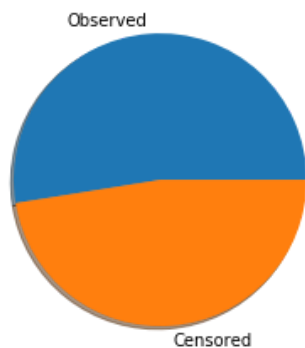
In this project, we have tried to estimate the survival and hazard functions of 118 patients with bladder cancer, who have entered a study where they received 3 different kinds of treatments :

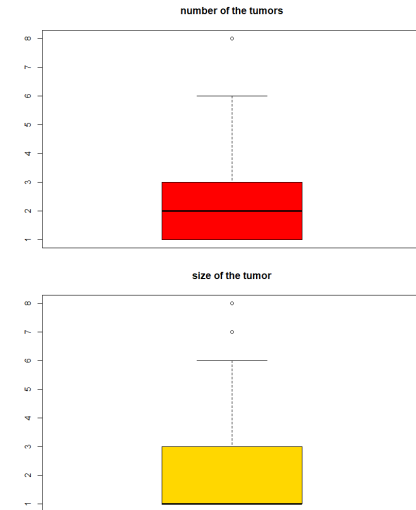
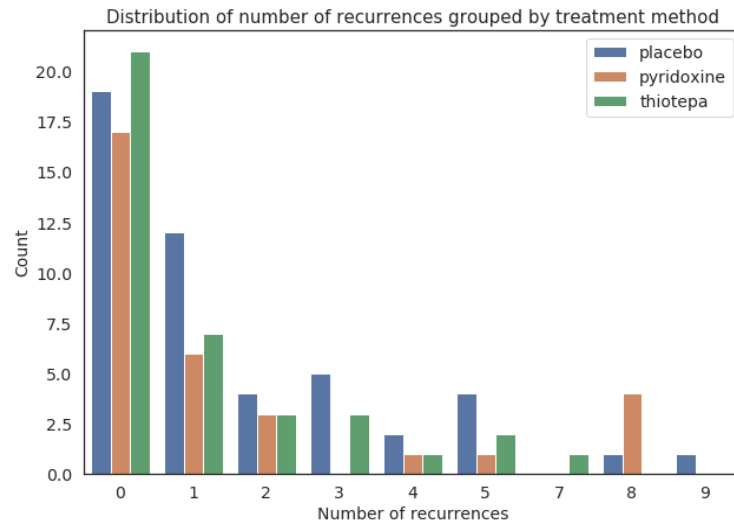
- placebo
- pyridoxine
- thiotepa

Further down the study, some of the patients have observed multiple recurrences (upto 9), some have died, and the rest have been censored. In order to deal with this data, we have divided the entire dataset into 2 parts :

- 118 patients with only their first event/censoring times (independent)
- the same 118 patients with all recurrence times (dependent)

Exploratory Data Analysis :



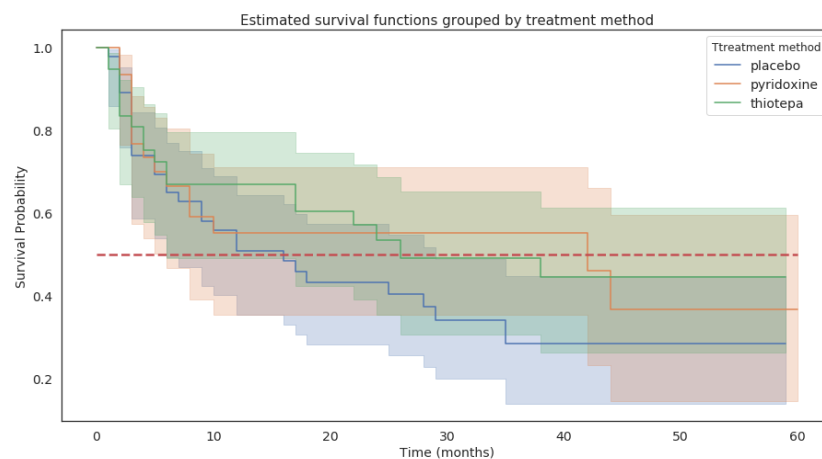


FIRST GROUP ANALYSIS

Kaplan-Meier Estimator :

The Kaplan–Meier estimator is a non-parametric statistic used to estimate the survival function from time-to-event data. The following is the Kaplan-Meier estimator of the patients grouped by the 3 treatment methods mentioned before. From the graph, we can conclude that

1. the group on pyridoxine has the highest median survival time among all patients
2. patients on pyridoxine have a median survival time approximately 4 times greater than the patients on placebo



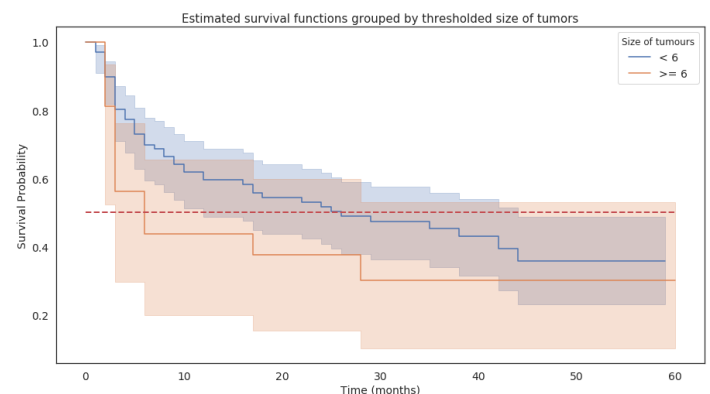
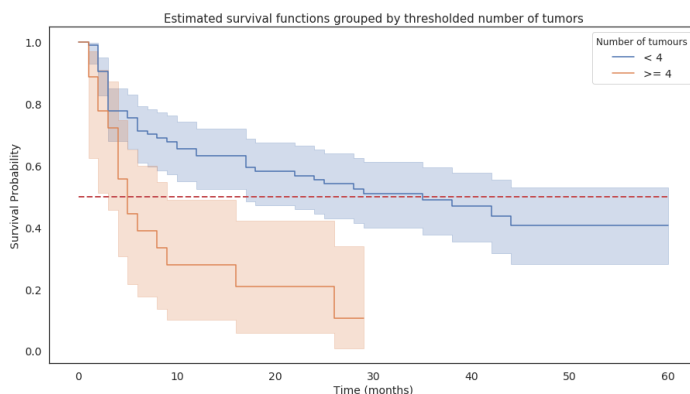
The pairwise log-rank test between these 3 groups yield the following results:

Log-rank test between treatments		Test statistic	p-value
Placebo	Pyridoxine	1.326279	0.249468
	Thiotepa	1.520945	0.217477
Pyridoxine	Thiotepa	0.001407	0.970077

The high p-value from the log-rank test between pyridoxine and thiotepa is indicative of the fact that their survival curves are not quite dissimilar, which is also clear from the graph above. The low p-value between placebo and the other 2 groups supports the fact that their survival curves deviate completely in later time points.

Next we have thresholded on the number and size of tumours respectively, on the threshold for which the pairwise log-rank test gives the lowest p-value, which indicates that the two groups are the most dissimilar in terms of survival and hazard functions. The following are the Kaplan-Meier estimators of the patients grouped by the thresholded number and size of tumours respectively. From the graphs below, we can conclude that

1. patients with less than 4 tumours have a median survival time approximately 7 times greater than those with greater than or equal to 4 tumours
2. patients whose size of the largest initial tumour was less than 6 cm have a median survival time approximately 4 times greater than the other group

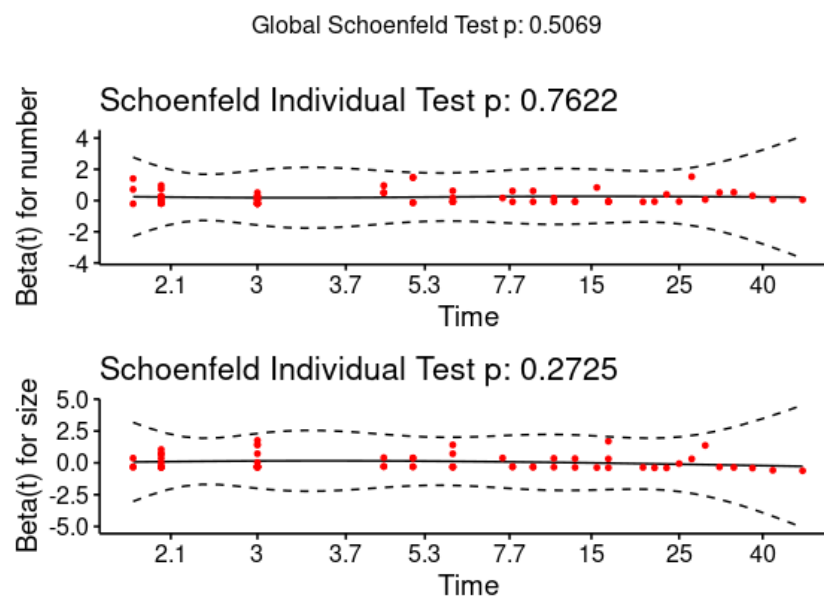


Cox Proportional Hazard model :

In the previous section we have worked with the Kaplan-Meier estimator which does not take any covariates into consideration. In this section, we shall be modelling the survival functions by taking into account the 2 covariates - number and size of tumour into the model itself.

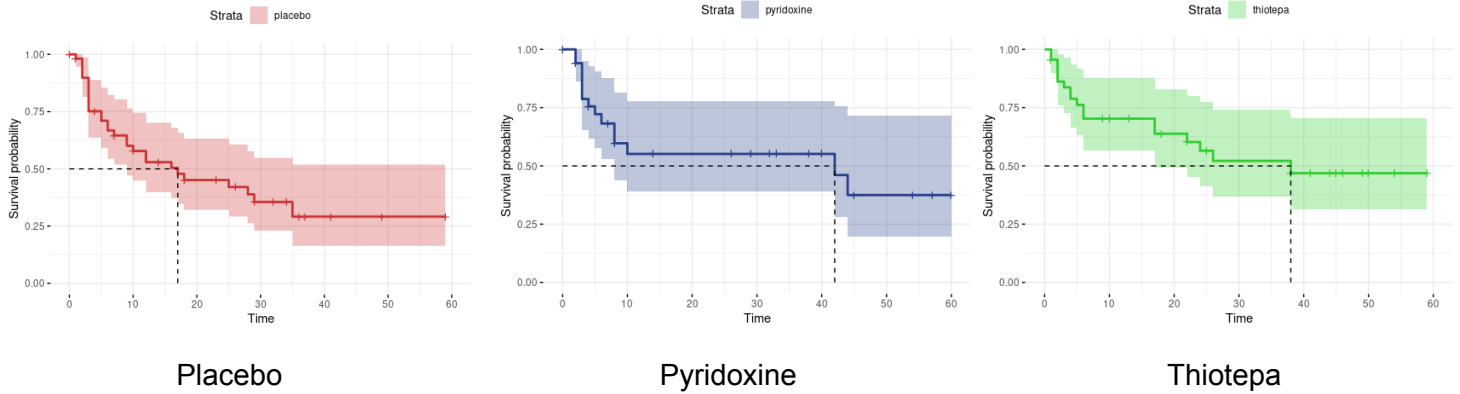
Cox proportional hazards assumptions:

Before modelling, we must check whether the covariates follow the Cox proportional hazard assumptions - that the hazards are proportional. This is achieved with the Schoenfeld residuals test, which plots the residual values of the covariates against time to look for time dependency.



As we can clearly see, the covariate residuals do not exhibit any pattern, and are completely random. Also, the p-values for both the covariates are greater than 0.05, hence both the covariates satisfy the proportional hazard assumption, and we can go ahead with modelling.

Model :



SECOND GROUP ANALYSIS

Unlike the first group, the second group has got multiple recurrences for individual patients.

X	id	treatment	number	size	recur	start	stop	status	rtumor	rsize	enum	delta	survtime
1	0	1 placebo	1	1	0	0	0	3 .	.	.	1	0	0
2	1	2 placebo	1	3	0	0	1	3 .	.	.	1	0	1
3	2	3 placebo	2	1	0	0	4	0 .	.	.	1	0	4
4	3	4 placebo	1	1	0	0	7	0 .	.	.	1	0	7
5	4	5 placebo	5	1	0	0	10	3 .	.	.	1	0	10
6	5	6 placebo	4	1	1	0	6	1 1	1	.	1	1	6
7	6	6 placebo	4	1	1	6	10	3 .	.	.	2	0	4
8	7	7 placebo	1	1	0	0	14	0 .	.	.	1	0	14
9	8	8 placebo	1	1	0	0	18	0 .	.	.	1	0	18
10	9	9 placebo	1	3	1	0	5	1 2	4	.	1	1	5
11	10	9 placebo	1	3	1	5	18	3 .	.	.	2	0	13
12	11	10 placebo	1	1	2	0	12	1 2	2	.	1	1	12
13	12	10 placebo	1	1	2	12	16	1 3	.	.	2	1	4
14	13	10 placebo	1	1	2	16	18	3 .	.	.	3	0	2
15	14	11 placebo	3	3	0	0	23	0 .	.	.	1	0	23
16	15	12 placebo	1	3	2	0	10	1 6	1	.	1	1	10
17	16	12 placebo	1	3	2	10	15	1 3	1	.	2	1	5
18	17	12 placebo	1	3	2	15	23	0 .	.	.	3	0	8
19	18	13 placebo	1	1	3	0	3	1 8	1	.	1	1	3
20	19	13 placebo	1	1	3	3	16	1 8	.	.	2	1	13
21	20	13 placebo	1	1	3	16	23	1 8	.	.	3	1	7

Cox proportional Hazard models are unable to capture the dependencies in between the data points (Here in between two recurrence data which comes from the same patient).

So, the Shared Frailty model has been used to model the hazard function of different treatment groups here. It has been assumed that the random variable Z follows Gamma Distribution with finite variance and mean 1, as we have used the multiplicative frailty modeling technique.

```

> summary(gam)
Call:
emfrail(formula = Surv(start, stop, delta) ~ number + size +
  treatment + cluster(id), data = group22)

Regression coefficients:

```

	coef	exp(coef)	se(coef)	adj. se	z	p
number	0.3181	1.3745	0.1203	0.1204	2.6426	0.01
size	0.0446	1.0456	0.1342	0.1343	0.3324	0.74
treatmentpyridoxine	0.0922	1.0965	0.3573	0.3573	0.2579	0.80
treatmentthiotepa	-0.6704	0.5115	0.3599	0.3600	-1.8622	0.06

```

Estimated distribution: gamma / left truncation: FALSE

Fit summary:
Commenges-Andersen test for heterogeneity: p-val 5.71e-11
no-frailty Log-likelihood: -663.099
Log-likelihood: -636.731
LRT: 1/2 * pchisq(52.7), p-val 1.91e-13

Frailty summary:

```

	estimate	lower 95%	upper 95%
Var[Z]	1.296	0.710	2.222
Kendall's tau	0.393	0.262	0.526
Median concordance	0.396	0.259	0.541
E[logZ]	-0.773	-1.435	-0.395
Var[logZ]	2.431	1.016	5.917
theta	0.772	0.450	1.409

```

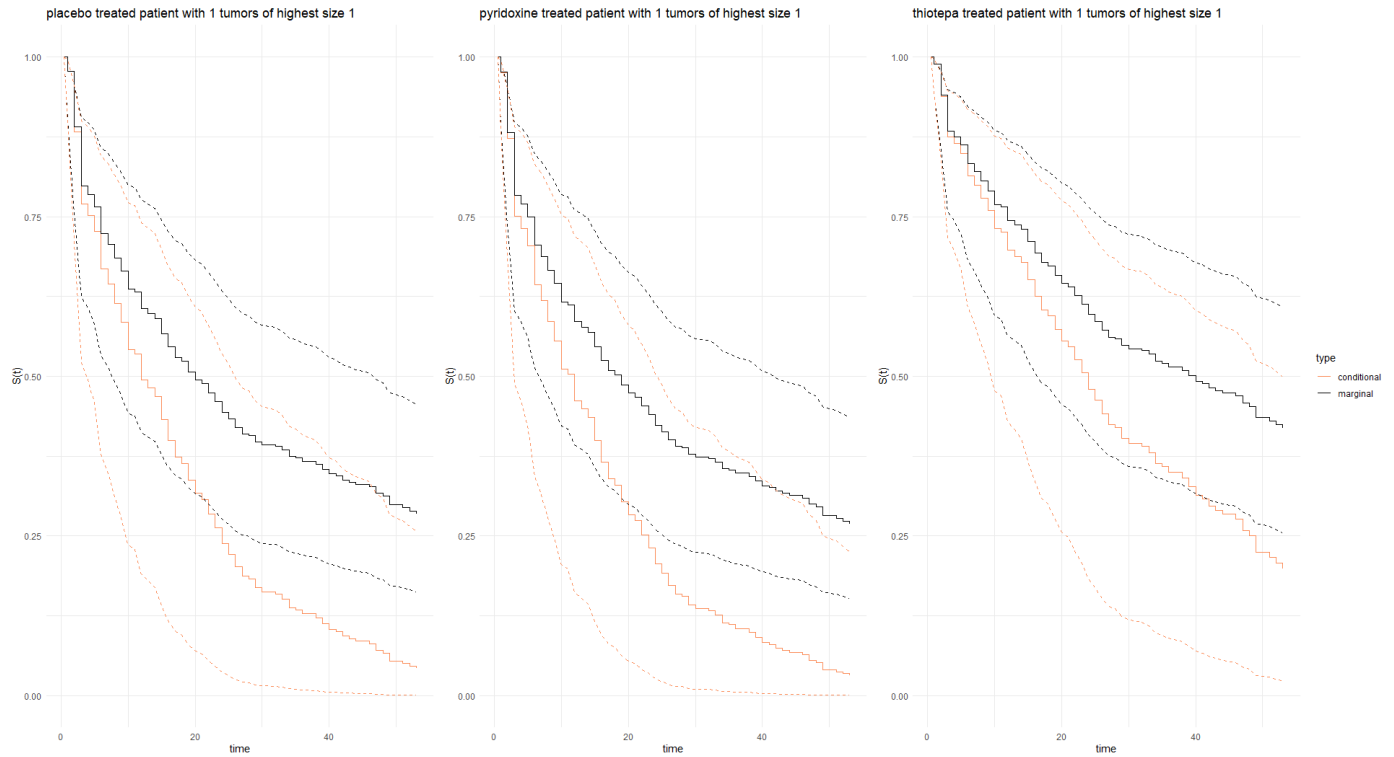
Confidence intervals based on the likelihood function
> |

```

Here we can see the summary of the fitted gamma shared frailty model. The first part i.e. **Regression Coefficients** , gives us :

1. Unit increase in the number of tumors and the maximum size of the tumors in the patient with all other covariates unchanged, increases the hazard 1.3745 times and 1.0456 times respectively.
2. A patient with a constant number of tumors with constant maximum size is 1.0965 times as likely to die in Pyridoxine treatment as the same in Placebo treatment.
3. A patient with a constant number of tumors with constant maximum size is 0.5115 times as likely to die in Thiotepa treatment as the same in Placebo treatment.

In the Commenges-Andersen Test, we check for heterogeneity in the data. And due to a very low p-value we can reject the null hypothesis that is There is no heterogeneity in the data i.e. the frailty variable Z does not follow a mixed distribution.



These are the conditional and marginal survival curves (i.e. the expected survival over Z), predicted by the fitted gamma shared frailty model, for an individual with one tumor of size 1 cm with respect to all three different treatments. We can instantly conclude :

- Median Life Time for the patient in Placebo treatment : 20 months (approx)
- Median Life Time for the patient in Pyridoxine treatment : 18 months (approx)
- Median Life Time for the patient in Thiotepa treatment : 40 months (approx)

These values correspond to the exponential of coefficients quite clearly, and this gives us the conclusion that, “The treatment Thiotepa is almost 2 times more effective for Bladder cancer than Pyridoxine or Placebo.”

Thank you.