



Survival Analysis of Patients with Recurrent Bladder Cancer



Srijan Mallick Dipan Banik
(under the supervision of Dr. Sudipta Das)

2nd year, M.Sc. Big Data Analytics
Ramakrishna Mission Vivekananda Educational and Research Institute

Project objective

- Estimate and compare survival curves of patients undergoing different treatment methods
 - taking covariates into consideration
 - without taking covariates into consideration
- Frailty modelling for multiple recurrences (dependent time-to-event)

DATA

```
graph TD; DATA[DATA] --> G1[Group 1 - upto the first event time (118)]; DATA --> G2[Group 2 - all event times (294)]; G1 --> G1_Obs[Observed (62)]; G1 --> G1_Cens[Censored (56)]; G2 --> G2_Obs[Observed (191)]; G2 --> G2_Cens[Censored (103)];
```

Group 1 - upto the first event time
(118)

Observed
(62)

Censored
(56)

Group 2 - all event times
(294)

Observed
(191)

Censored
(103)

Group 1 - upto the first event time (118)

id	treatment	number	size	Tis	time_to_first_event	delta
1	placebo	1	1	[[0, 3]]	0	0
2	placebo	1	3	[[1, 3]]	1	0
3	placebo	2	1	[[4, 0]]	4	0
4	placebo	1	1	[[7, 0]]	7	0
5	placebo	5	1	[[10, 3]]	10	0
6	placebo	4	1	[[6, 1), (4, 3]]	6	1
7	placebo	1	1	[[14, 0]]	14	0
8	placebo	1	1	[[18, 0]]	18	0
9	placebo	1	3	[[5, 1), (13, 3]]	5	1
10	placebo	1	1	[[12, 1), (4, 1), (2, 3]]	12	1
11	placebo	3	3	[[23, 0]]	23	0
12	placebo	1	3	[[10, 1), (5, 1), (8, 0]]	10	1
13	placebo	1	1	[[3, 1), (13, 1), (7, 1)]	3	1
14	placebo	3	1	[[3, 1), (6, 1), (12, 1), (2, 2)]	3	1
15	placebo	2	3	[[7, 1), (3, 1), (6, 1), (8, 1)]	7	1
16	placebo	1	1	[[3, 1), (12, 1), (10, 1)]	3	1

Group 2 - all event times (294)

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

Group 1 - upto the first event time
(118)

id	treatment	number	size	Tis	time to first event	delta
1	placebo	1	1	[[0, 3]]	0	0
2	placebo	1	3	[[1, 3]]	1	0
3	placebo	2	1	[[4, 0]]	4	0
4	placebo	1	1	[[7, 0]]	7	0
5	placebo	5	1	[[10, 0]]	10	0
6	placebo	1	3	[[10, 1], (4, 3)]	6	1
7	placebo	1	1	[[14, 0]]	14	0
8	placebo	1	1	[[16, 0]]	16	0
9	placebo	1	3	[[5, 1], (13, 3)]	5	1
10	placebo	1	1	[[21, 1], (10, 0)]	11	1
11	placebo	3	3	[[23, 0]]	23	0
12	placebo	1	3	[[10, 1], (5, 1), (8, 0)]	10	1
13	placebo	1	1	[[3, 1], (10, 1), (7, 1)]	3	1
14	placebo	3	1	[[3, 1], (6, 1), (12, 1), (2, 2)]	3	1
15	placebo	2	3	[[7, 1], (3, 1), (6, 1), (8, 1)]	7	1
16	placebo	1	1	[[3, 1], (12, 1), (10, 1)]	3	1

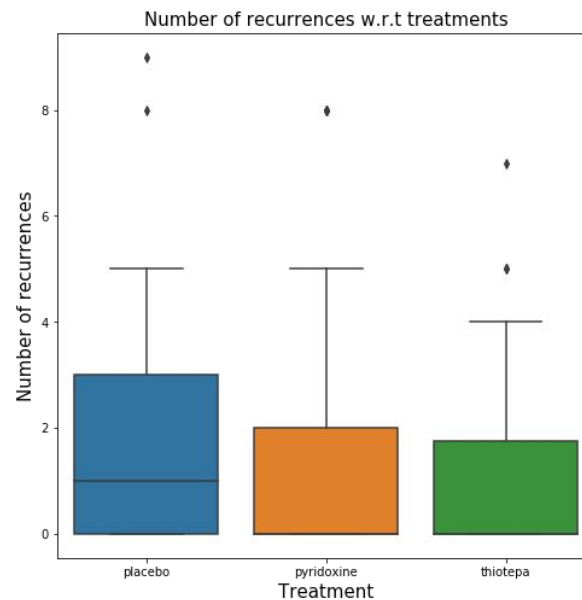
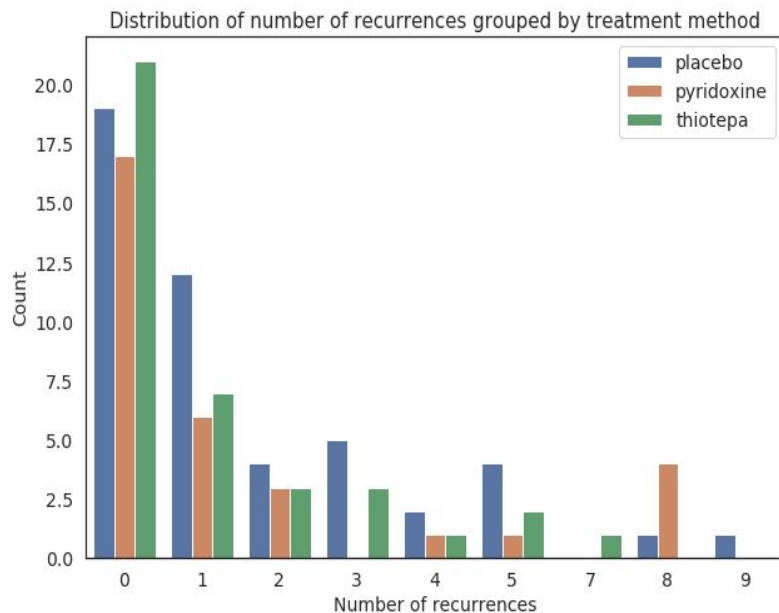
This data contains
only the first
recurrence times,
hence independent
time points

Group 2 - all event times
(294)

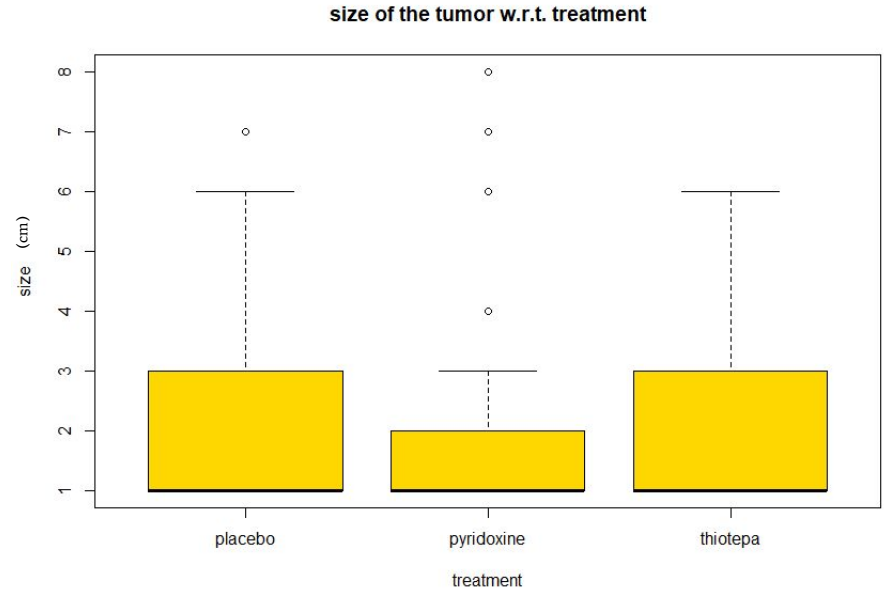
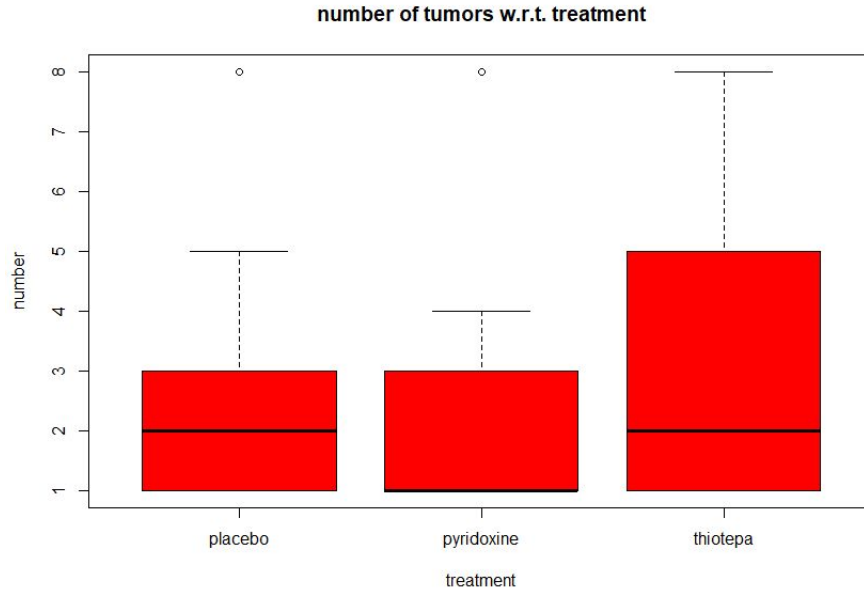
id	treatment	number	size	recur	start	stop	status	rtumor	rsize	enum	delta	survtime
1	placebo	1	1	0	0	0	3	.	.	1	0	0
2	placebo	1	3	0	0	1	3	.	.	1	0	1
3	placebo	2	1	0	0	4	0	.	.	1	0	4
4	placebo	1	1	0	7	0	.	.	.	1	0	7
5	placebo	1	1	0	10	0	3	.	.	1	0	10
6	placebo	4	1	1	0	6	1	1	1	1	1	6
7	placebo	1	1	1	6	10	3	.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

This data contains
multiple recurrence
times, hence
dependent time
points

Treatment-wise distribution of recurrences

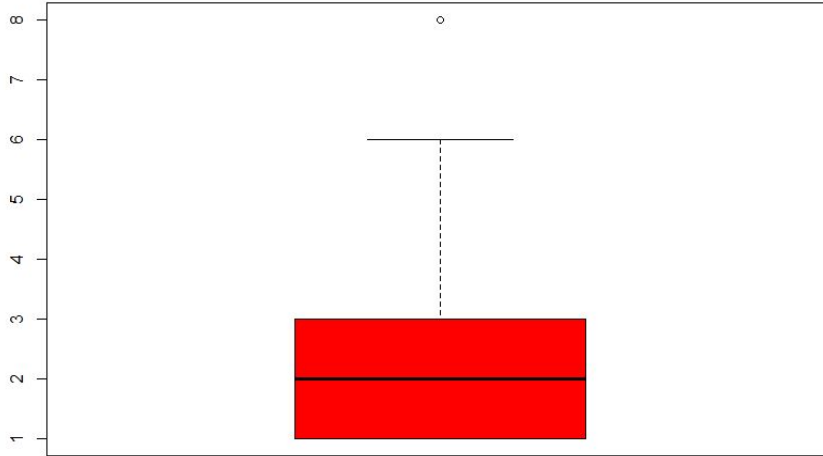


Outlier Detection in the Covariates

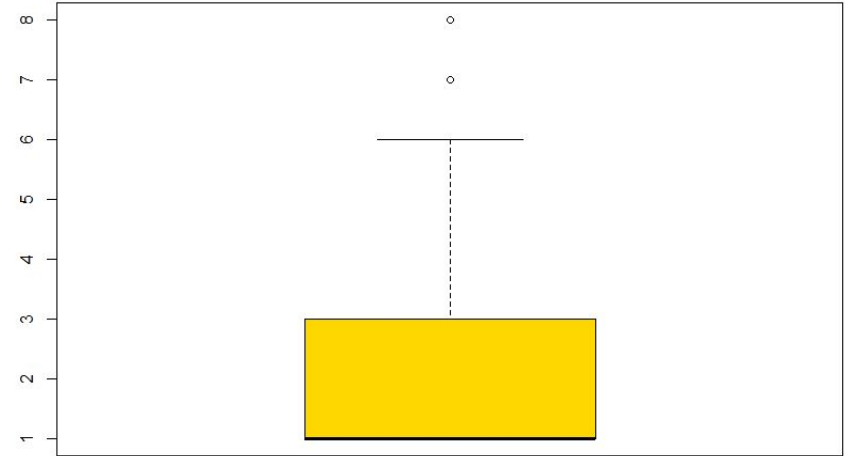


Outlier Detection in the Covariates

number of the tumors



size of the tumor



Kaplan-Meier Estimation (no covariates into consideration)

Kaplan-Meier estimator

The Kaplan–Meier estimator is a non-parametric statistic used to estimate the survival function from time-to-event data.

$$\hat{S}(t) = \prod_{t_i \leq t, i=1}^n \left(1 - \frac{d_i}{Y_i}\right)$$

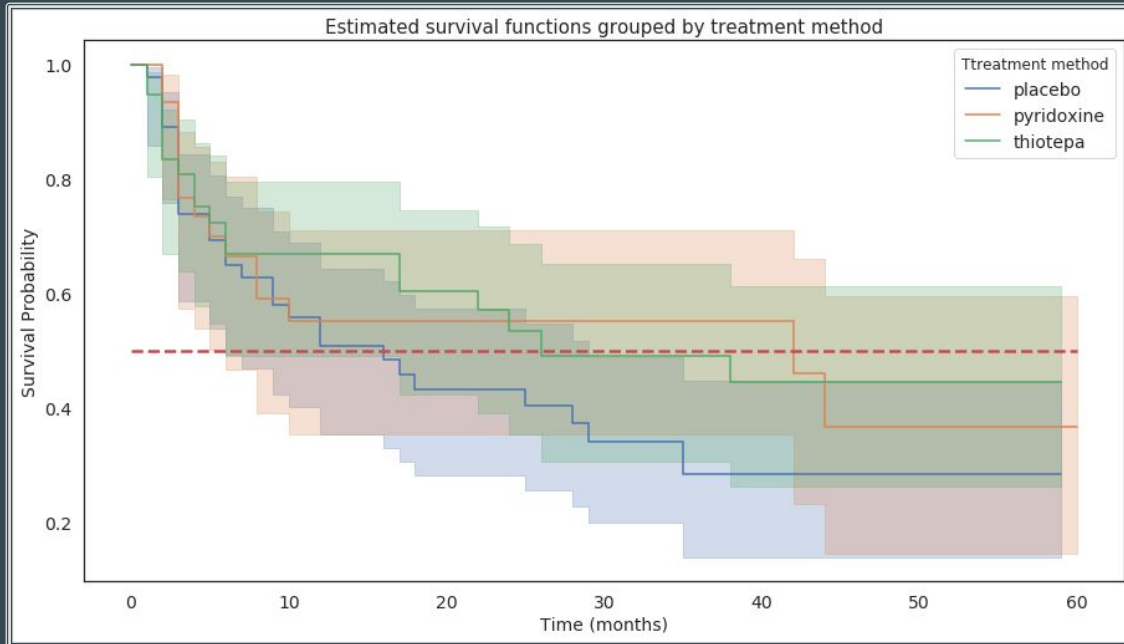
Log-rank test

The log-rank test is a non-parametric hypothesis test to compare survival distributions from two samples.

The null hypothesis is that the two groups have identical hazard functions, ie :

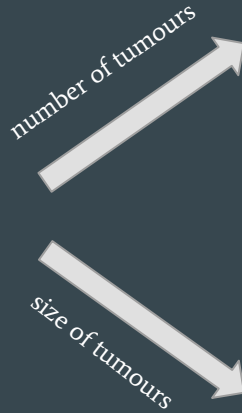
$$H_0 : h_1(t) = h_2(t)$$

Kaplan-Meier estimator of survival functions of the 3 treatment groups



		test_statistic	p	-log2(p)
placebo	pyridoxine	1.326279	0.249468	2.003072
	thiotepa	1.520945	0.217477	2.201068
pyridoxine	thiotepa	0.001407	0.970077	0.043828

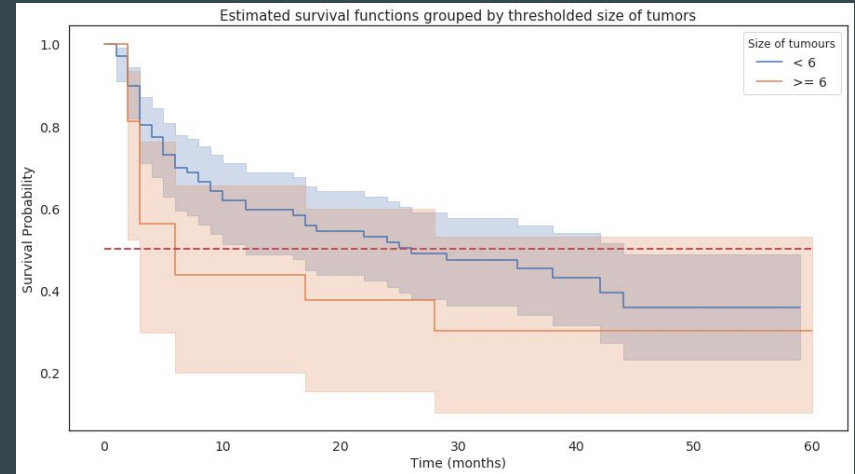
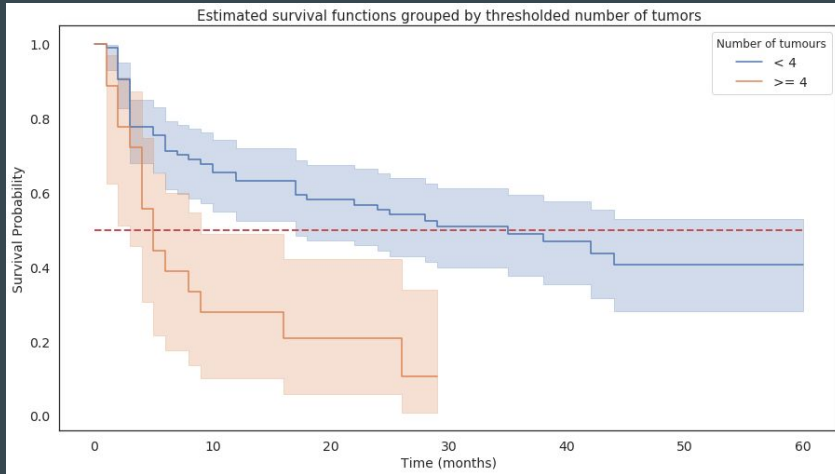
Thresholding on



```
p-value when threshold on number of tumours is 2 : 0.0017939330834322697
p-value when threshold on number of tumours is 3 : 0.007020162822710038
p-value when threshold on number of tumours is 4 : 0.0002594167683235225
p-value when threshold on number of tumours is 5 : 0.0026688648730418267
p-value when threshold on number of tumours is 6 : 0.0010422478423726767
p-value when threshold on number of tumours is 7 : 0.02283357735020892
p-value when threshold on number of tumours is 8 : 0.02283357735020892
```

```
p-value when threshold on size of tumours is 2 : 0.5199524395300987
p-value when threshold on size of tumours is 3 : 0.3024758137141664
p-value when threshold on size of tumours is 4 : 0.23865366312441283
p-value when threshold on size of tumours is 5 : 0.1682752207760559
p-value when threshold on size of tumours is 6 : 0.1570410142894633
p-value when threshold on size of tumours is 7 : 0.9146316972988796
p-value when threshold on size of tumours is 8 : 0.3025414266710924
```

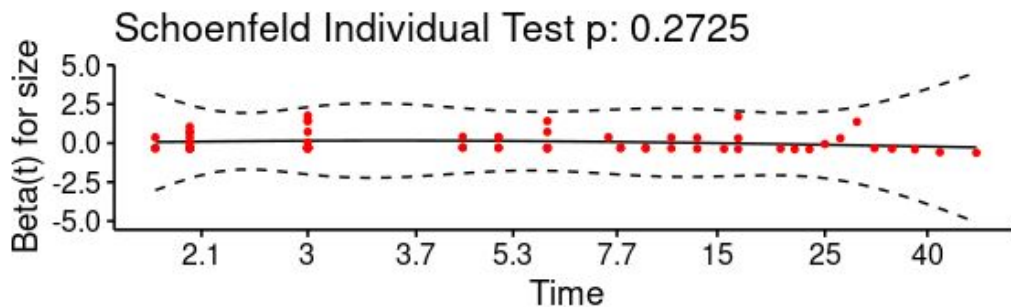
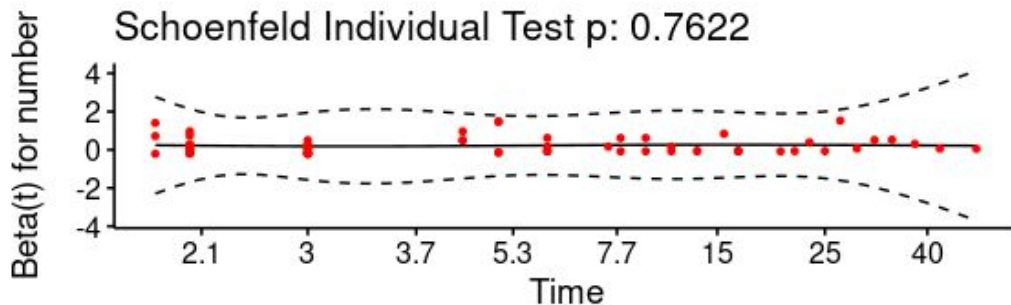
Kaplan-Meier estimator of survival functions of the thresholded groups



Cox Proportional Hazard Model (taking covariates into consideration)

Proportional Hazard Assumption : Schoenfeld Test

Global Schoenfeld Test p: 0.5069



H_0 : Hazards are proportional

H_1 : Hazards are not proportional

	chisq	df	p
number	0.0915	1	0.76
size	1.2041	1	0.27
GLOBAL	1.3589	2	0.51

Cox Proportional Hazard Model

```
> summary(cox)
Call:
coxph(formula = survobj1 ~ treatment + number + size, data = group22,
      cluster = id)

n= 254, number of events= 162

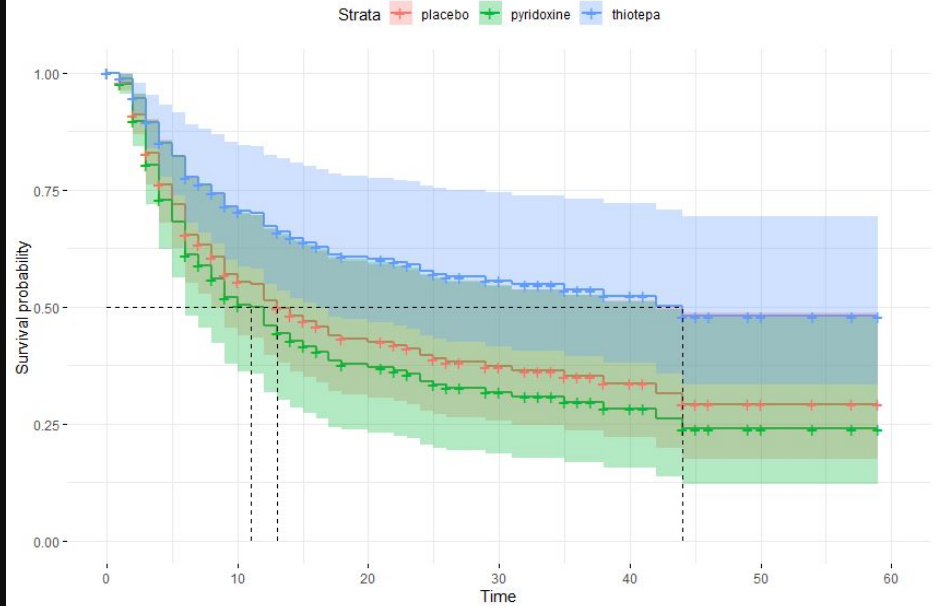
              coef exp(coef) se(coef) robust se      z Pr(>|z|)
treatmentpyridoxine  0.14967   1.16145  0.18418   0.24514  0.611 0.541516
treatmentthiotepa   -0.51959   0.59476  0.21593   0.26281 -1.977 0.048036 *
number              0.26014   1.29712  0.06319   0.07655  3.398 0.000678 ***
size                0.06854   1.07094  0.07729   0.09234  0.742 0.457983
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
treatmentpyridoxine  1.1614   0.8610   0.7183   1.8779
treatmentthiotepa    0.5948   1.6813   0.3553   0.9955
number              1.2971   0.7709   1.1164   1.5071
size                1.0709   0.9338   0.8936   1.2834

Concordance= 0.602 (se = 0.025 )
Likelihood ratio test= 23.47 on 4 df,  p=1e-04
Wald test              = 22.79 on 4 df,  p=1e-04
Score (logrank) test = 23.67 on 4 df,  p=9e-05, Robust = 12.36 p=0.01

(Note: the likelihood ratio and score tests assume independence of
observations within a cluster, the Wald and robust score tests do not).
```

survival curves for a patient with 1 tumors of highest size 1



Shared Frailty Modelling

- Unobserved covariates, (U):

$$h(t) = h_0(t)e^{x'\beta + u'\beta^*}$$

$$h(t | z) = zh_0(t)e^{x'\beta}$$

where $z = e^{u'\beta^*}$

- Conditional survival function:

$$S(t|z) = \exp(-zH_0(t))e^{x'\beta}$$

where $H_0(t)$ integrated hazard.

- Unconditional survival function:

$$S(t) = E_z \left[e^{-zH_0(t)e^{x'\beta}} \right] = L_z \left[H_0(t)e^{x'\beta} \right]$$

Commenges-Anderson Test :

```
> ca_test(cox)
      tstat      var      pval
9.377964e+01 5.482473e+02 6.197441e-05
> |
```

Gamma Frailty Model :

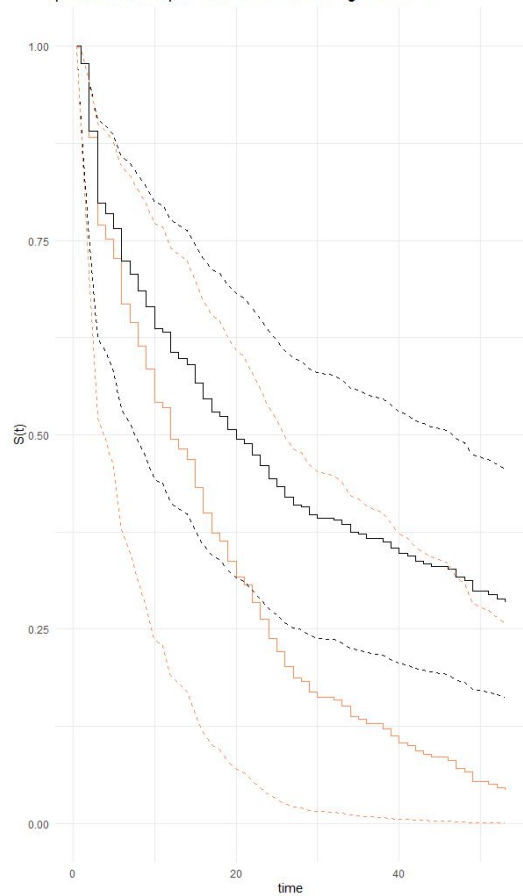
```
> summary(gam)
Call:
emfrail(formula = Surv(start, stop, delta) ~ number + size +
  treatment + cluster(id), data = group22)

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

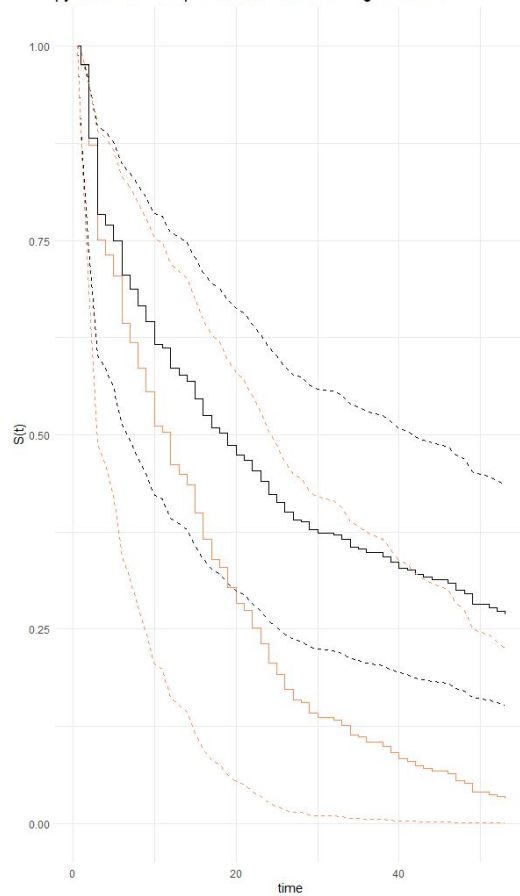
Fit summary:
Commenges-Andersen test for heterogeneity: p-val 5.71e-11
no-frailty Log-likelihood: -663.099
Log-likelihood: -636.731
LRT: 1/2 * pchisq(52.7), p-val 1.91e-13

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
> |
```

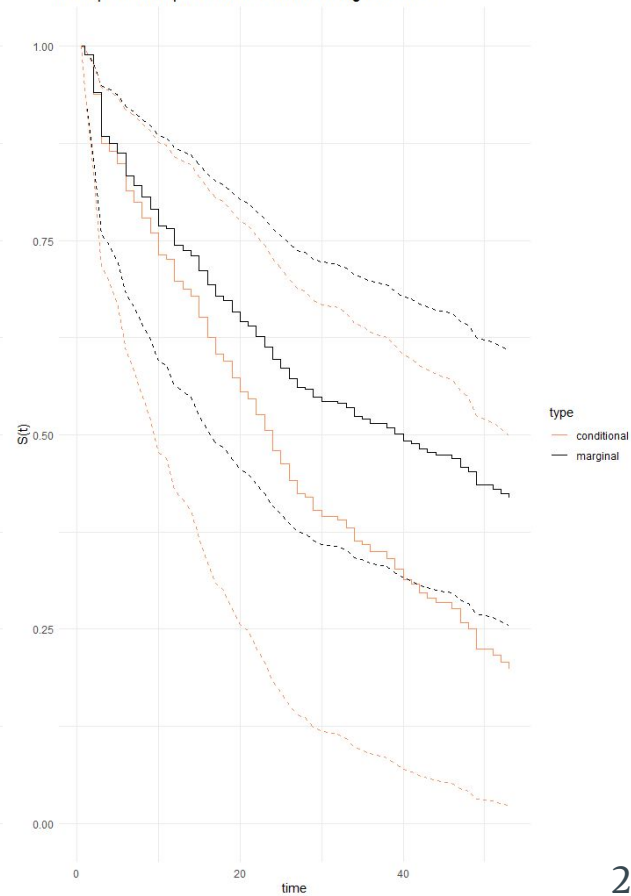
placebo treated patient with 1 tumors of highest size 1



pyridoxine treated patient with 1 tumors of highest size 1



thiotepa treated patient with 1 tumors of highest size 1



Thank You !