Stats 504 Assignment 2: Diabetic Retinopathy

Introduction

Diabetic retinopathy is a diabetes complication that can cause complications with eyes, resulting in loss of vision. This report compares the results of two laser coagulation treatments which attempt to delay the onset of diabetic retinopathy. The findings from this analysis are intended to enable the researchers from Michigan Medicine to quantify the efficacy of both treatments. The report also addresses specific questions of interest such as the influence of age and pre-existing clinical risk of loss of visual acuity on whether a patient loses vision. It is found that both treatments are effective in delaying the onset of diabetic retinopathy, but there is no statistically significant difference between the two treatments.

Methods

The goal of the analysis is to quantify efficacies of the Xenon and Argon laser treatments, as well as provide insights on the effect of other variables on a patient's chance of going blind. Considering that the data was "censored" (meaning that times at which each patient lost vision was not available), a survival analysis model seemed most appropriate in terms of methodology. Survival Analysis modeling attempts to explain the time it takes for an event to occur, which in this case, is the event of losing vision in an eye. More specifically, a Proportional Hazards Survival model is used, as this is able to describe the effect of a unit change in covariates (like treatment group, age, etc) in terms of a multiplicative, which is in line with the requirements of this problem space. However, proportional hazards modeling must be performed with care. In case of paired data such as the one used in this study where each subject provides data for two eyes, a frailty parameter must be added to account for the inter-dependence caused by the paired data. This frailty model is further detailed in the appendix.

As a visual aid (presented in the following section) of trends in patients' vision loss, a Kaplan-Meier curve can be fitted to the data. This depicts the "survival function" of the data, which is the probability of the event in question occurring at any given point in time. For this study, this translates to the probability of a patient losing vision at any given point in time. These curves can be plotted for both the Xenon and Argon treatments, and then contrasted with their respective control groups to get a graphical idea of which treatment works, and if applicable, which one is better. Furthermore, a log-rank test can also be performed to evaluate the statistical differences between the curves in the Kaplan-Meier figure.

Finally, the coefficients resulting from the modeling can be interpreted as multiplicative factors affecting the risk of losing vision, and this is described in the following section.

Results

The data provided by the researchers follows 197 clinic patients, with each patient contributing two rows to the data. Each row provides information about the type of laser treatment received, and the eye it was received on, along with other covariates described in Table 1. The data was clean and no pre-processing or data cleaning was necessary to prepare the data for analysis. The analysis was finally performed on 394 different rows (197 pairs) with each row having 7 different features which are further described below.

Feature	Median (IQR) / Percentage
ID	197 unique subject IDs
Laser	
Xenon	200 (51%)
Argon	194 (49%)
Age	16 (10, 30)
Age Type	
Juvenile	228 (58%)
Adult	166 (42%)
Treatment Group	
Status - Lost to Follow Up	143 (72.5%)
Status - Loss of Vision	54 (27.5%)
Control Group	
Status - Lost to Follow Up	101 (51%)
Status - Loss of Vision	96 (49%)
Follow Up Time	38.8 (13.9, 54.2)
Risk of Loss of Visual Acuity	10 (9, 11)

Table 1: Baseline table indicating summary statistics of data used in the analysis

For the Cox Proportional Hazard model, the response variable i.e the variable we are interested in, is the pair of the Follow Up Time, and the Status. These data points, in conjunction, inform whether a patient reported loss of vision, and if so, after how long in the study. In case the study ended, or the patient became deceased before losing vision, those rows were assigned the "Lost to Follow Up" status, and the Follow Up Time refers to the last follow up the patient had. The Age Type variable remained unused in this analysis, as it provides no new information (mathematically) due to the presence of the Age variable which encapsulates the same information. Finally, the Risk variable indicates the pre-existing risk of a patient losing vision (or visual acuity) in an eye. Owing to no further information about the specifics of this value, it is assumed reasonably that the higher this value, the higher the chance for the patient to lose visual acuity in that eye.

In Figure 1, the Kaplan-Meier curves indicate the probability of retaining vision across time, for each of the groups of interest. It is evident that the two treatment groups, for Xenon and Argon, are distinct from their corresponding control groups, thereby indicative of a positive effect from the laser treatments. This can further be confirmed and quantified by fitting the Cox Proportional Hazard model, and is depicted below.

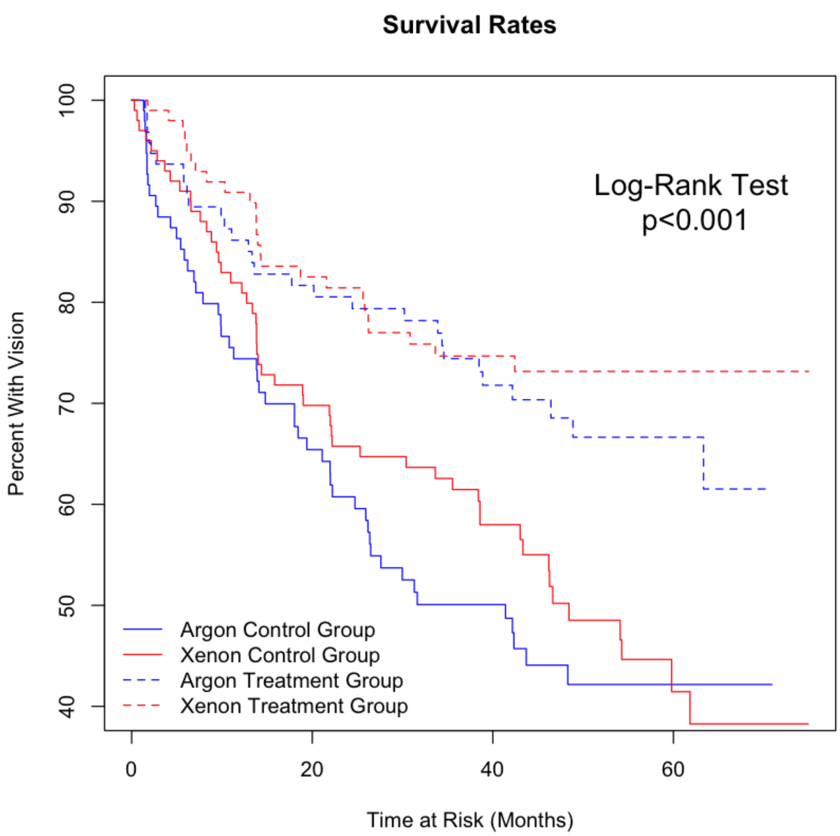


Figure 1: Kaplan-Meier curves indicate that the probability of retaining vision is higher for the two treatments, as compared to the control groups.

The log-rank test is a statistical test to check the validity of the Kaplan-Meier curves. A p-value of less than 0.05 (as observed in Figure 1) indicates that at least one of the curves is significantly different from the others. This could indicate one of two things. Either the significantly different curve is one of the treatment curves, as compared to a control curve, or it could be a statistically significant difference between the two treatment curves, indicating a difference in efficacies of the treatments. To confirm this hypothesis, the Cox Proportional Hazard model is fitted, and results are presented below.

Covariate	Hazard Rate 95% CI	p-value
Treatment (Either Xenon or Argon)	0.38, (0.24 0.62)	<0.001
Treatment Xenon compared to Argon	1.07, (0.54 2.11)	0.85
Risk	1.18, (1.03 1.36)	<0.01
Age	1.007, (0.99 1.02)	0.36

Table 2: Model parameters indicating the Hazard Rate for each covariate.

As the table suggests, Age is a statistically insignificant variable, implying that a change in age does not necessarily affect the time at which a patient loses visual acuity due to diabetic retinopathy. To similarly interpret other parameters, the pre-existing risk of loss of visual acuity is statistically significant, and has a hazard rate of 1.18. In other words, this means if all other covariates were the same between two eyes, an increase of 1 unit in the risk would result in an 18% higher chance of loss of vision. the hazard rates of the treatment effects can be interpreted.

To address the primary research question about the efficacy and quantification of the two treatments, the hazard rates can be interpreted as follows. Receiving either kind of laser treatment, implied a 62% lower chance of loss of vision, whereas the effect of receiving the Argon treatment as opposed to the Xenon treatment was statistically insignificant, and does not seem to affect the risk of losing visual acuity.

In other words, both treatments seem to work in delaying the onset of diabetic retinopathy. However, there is a lack of evidence to statistically claim either treatment performs better than the other.

Conclusion

This report presents the results of an analysis performed on clinic patients' eye treatment data, with the objective of trying to understand the efficacies of both the Argon and Xenon laser treatment. It also addresses the key questions posed by the client regarding specific features of interest like the Age of the patient, and their pre-existing risk of loss of visual acuity. The analysis reveals that Age is not a statistically significant factor, while the pre-existing risk increases chances of losing visual acuity by a multiplicative factor of 1.18. Also, while both the laser treatments show evidence of having the desired effect (reducing risk of losing vision by 62%), there is not enough evidence to suggest one of the laser treatments can be preferred over the other due to a higher efficacy.

Exploratory Analysis

In [13]:
```
import pandas as pd
```

In [14]:
```
data = pd.read_csv("diabeticVision.csv", index_col=0)
```

Look at summary statistics

In [15]:
```
data.describe()
```

Out[15]:

	id	age	trt	futime	status	risk	group
count	394.000000	394.000000	394.000000	394.000000	394.000000	394.000000	394.000000
mean	873.203046	20.781726	0.500000	35.579289	0.393401	9.697970	1.507614
std	495.523410	14.812074	0.500636	21.355896	0.489126	1.475033	1.119430
min	5.000000	1.000000	0.000000	0.300000	0.000000	6.000000	0.000000
25%	480.000000	10.000000	0.000000	13.977500	0.000000	9.000000	1.000000
50%	834.000000	16.000000	0.500000	38.800000	0.000000	10.000000	1.500000
75%	1296.000000	30.000000	1.000000	54.252500	1.000000	11.000000	3.000000
max	1749.000000	58.000000	1.000000	74.970000	1.000000	12.000000	3.000000

In [26]:
```
data[data["trt"] == 1]["status"].value_counts(normalize=0)
```

Out[26]:
```
0    143
1     54
Name: status, dtype: int64
```

In [25]:
```
data[data["trt"] == 0]["status"].value_counts(normalize=0)
```

Out[25]:
```
1    101
0     96
Name: status, dtype: int64
```

In [20]:
```
data["type"].value_counts()
```

Out[20]:
```
juvenile    228
adult       166
Name: type, dtype: int64
```

In [4]:
```
data.isna().sum()
```

Out[4]:
```
id          0
laser       0
eye         0
age         0
```

```
type       0
trt        0
futime     0
status     0
risk       0
group      0
dtype: int64
```

In [6]:
```
data.groupby(["trt", "status"])["futime"].mean()
```

Out[6]:
```
trt  status
0    0         46.321771
     1         18.948515
1    0         46.668112
     1         18.222407
Name: futime, dtype: float64
```

In [5]:
```
data.groupby(["trt", "status", "laser"])["futime"].mean()
```

Out[5]:
```
trt  status  laser
0    0       argon    43.247234
             xenon    49.270816
     1       argon    16.125000
             xenon    21.716667
1    0       argon    45.927500
             xenon    47.339600
     1       argon    20.004828
             xenon    16.154800
Name: futime, dtype: float64
```

In [48]:
```
data.pivot(index=["id", "trt"], columns=["status"])["futime"][0].dropna()
```

Out[48]:
```
id    trt
5     0      46.23
      1      46.23
14    1      42.50
16    0      42.27
      1      42.27
             ...
1717  1      51.60
1727  1      49.97
1746  1      45.90
1749  0      41.93
      1      41.93
Name: 0, Length: 239, dtype: float64
```

Look at various sub-groups

In [31]:
```
treat = data[data["trt"] == 1]
ctrl = data[data["trt"] == 0]
treat1 = treat[treat["laser"] == "argon"]
treat2 = treat[treat["laser"] == "xenon"]
```

In [ ]:
```
treat[treat["status"] == 1]
```

In [15]:
```
data.groupby("id")["eye"].sum().unique()
```

Out[15]: array(['leftleft', 'rightright'], dtype=object)

In [17]:

```
data[data["futime"] == 46.23]
```

Out[17]:

	id	laser	eye	age	type	trt	futime	status	risk
1	5	argon	left	28	adult	1	46.23	0	9
2	5	argon	left	28	adult	0	46.23	0	9
195	832	argon	right	5	juvenile	1	46.23	0	12
196	832	argon	right	5	juvenile	0	46.23	0	12

In [24]:

```
ctrl.describe()
```

Out[24]:

	id	age	trt	futime	status	risk
count	101.000000	101.000000	101.0	101.000000	101.0	101.000000
mean	801.207921	23.079208	0.0	18.948515	1.0	9.970297
std	481.609869	15.532342	0.0	15.735833	0.0	1.465984
min	14.000000	1.000000	0.0	0.300000	1.0	6.000000
25%	409.000000	11.000000	0.0	6.530000	1.0	9.000000
50%	722.000000	19.000000	0.0	13.900000	1.0	10.000000
75%	1205.000000	37.000000	0.0	26.470000	1.0	11.000000
max	1746.000000	56.000000	0.0	61.830000	1.0	12.000000

In [25]:

```
treat1.describe()
```

Out[25]:

	id	age	trt	futime	status	risk
count	29.000000	29.000000	29.0	29.000000	29.0	29.000000
mean	822.344828	18.206897	1.0	20.004828	1.0	9.931034
std	497.132368	14.639426	0.0	17.418952	0.0	1.251600
min	100.000000	1.000000	1.0	1.500000	1.0	6.000000
25%	357.000000	9.000000	1.0	5.770000	1.0	9.000000
50%	866.000000	13.000000	1.0	13.330000	1.0	10.000000
75%	1184.000000	23.000000	1.0	34.370000	1.0	10.000000
max	1649.000000	53.000000	1.0	63.330000	1.0	12.000000

In [26]:

```
treat2.describe()
```

Out[26]:

	id	age	trt	futime	status	risk

	id	age	trt	futime	status	risk
count	25.000000	25.000000	25.0	25.000000	25.0	25.000000
mean	812.560000	18.600000	1.0	16.154800	1.0	9.720000
std	438.838531	13.044795	0.0	10.425035	0.0	1.369915
min	127.000000	3.000000	1.0	1.770000	1.0	6.000000
25%	503.000000	10.000000	1.0	7.070000	1.0	9.000000
50%	778.000000	13.000000	1.0	13.830000	1.0	10.000000
75%	1017.000000	25.000000	1.0	25.630000	1.0	11.000000
max	1688.000000	50.000000	1.0	42.430000	1.0	12.000000

In [10]:

```
def assign_group(row):
    if row["trt"] == 0:
        if row["laser"] == "argon":
            return 0
        return 1
    elif row["trt"] == 1:
        if row["laser"] == "argon":
            return 2
        return 3
```

Identify and label Laser and Treatment as an interaction term

In [11]:

```
data["group"] = data[["laser", "trt"]].apply(assign_group, axis=1).astype("categ
data
```

Out[11]:

	id	laser	eye	age	type	trt	futime	status	risk	group
1	5	argon	left	28	adult	1	46.23	0	9	2
2	5	argon	left	28	adult	0	46.23	0	9	0
3	14	argon	right	12	juvenile	1	42.50	0	8	2
4	14	argon	right	12	juvenile	0	31.30	1	6	0
5	16	xenon	right	9	juvenile	1	42.27	0	11	3
...	...	...	...	...	...	...	...	...	...	...
390	1727	argon	right	33	adult	0	2.90	1	10	0
391	1746	argon	right	3	juvenile	1	45.90	0	10	2
392	1746	argon	right	3	juvenile	0	1.43	1	10	0
393	1749	argon	right	32	adult	1	41.93	0	9	2
394	1749	argon	right	32	adult	0	41.93	0	9	0

394 rows × 10 columns

In [24]:
```
data["group"].describe()
```

Out[24]:
```
count     394
unique      3
top         2
freq      197
Name: group, dtype: int64
```

In [12]:
```
data.to_csv("diabeticVision.csv")
```

In [ ]:

In [75]:

```
library(survival)
```

Appropriate data manipulations

In [83]:

```
dat <- read.csv("diabeticVision.csv")
dat$trt = factor(dat$trt)
dat$laser = factor(dat$laser)
dat$type = factor(dat$type)
dat$group = factor(dat$group)
dat
```

A data.frame: 394 × 11

X	id	laser	eye	age	type	trt	futime	status	risk	group
<int>	<int>	<fct>	<chr>	<int>	<fct>	<fct>	<dbl>	<int>	<int>	<fct>
1	5	argon	left	28	adult	1	46.23	0	9	2
2	5	argon	left	28	adult	0	46.23	0	9	0
3	14	argon	right	12	juvenile	1	42.50	0	8	2
4	14	argon	right	12	juvenile	0	31.30	1	6	0
5	16	xenon	right	9	juvenile	1	42.27	0	11	3
6	16	xenon	right	9	juvenile	0	42.27	0	11	1
7	25	argon	left	9	juvenile	1	20.60	0	11	2
8	25	argon	left	9	juvenile	0	20.60	0	11	0
9	29	xenon	left	13	juvenile	1	38.77	0	9	3
10	29	xenon	left	13	juvenile	0	0.30	1	10	1
11	46	xenon	right	12	juvenile	1	65.23	0	9	3
12	46	xenon	right	12	juvenile	0	54.27	1	9	1
13	49	argon	right	8	juvenile	1	63.50	0	8	2
14	49	argon	right	8	juvenile	0	10.80	1	6	0
15	56	xenon	right	12	juvenile	1	23.17	0	8	3
16	56	xenon	right	12	juvenile	0	23.17	0	9	1
17	61	argon	right	16	juvenile	1	1.47	0	9	2
18	61	argon	right	16	juvenile	0	1.47	0	10	0
19	71	argon	right	21	adult	1	58.07	0	9	2
20	71	argon	right	21	adult	0	13.83	1	9	0
21	100	argon	left	23	adult	1	46.43	1	9	2
22	100	argon	left	23	adult	0	48.53	0	9	0
23	112	argon	right	44	adult	1	44.40	0	11	2
24	112	argon	right	44	adult	0	7.90	1	12	0

X	id	laser	eye	age	type	trt	futime	status	risk	group
<int>	<int>	<fct>	<chr>	<int>	<fct>	<fct>	<dbl>	<int>	<int>	<fct>
25	120	xenon	left	47	adult	1	39.57	0	11	3
26	120	xenon	left	47	adult	0	39.57	0	6	1
27	127	xenon	right	48	adult	1	30.83	1	6	3
28	127	xenon	right	48	adult	0	38.57	1	10	1
29	133	argon	right	26	adult	1	66.27	0	10	2
30	133	argon	right	26	adult	0	14.10	1	9	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
365	1619	xenon	left	20	adult	1	74.97	0	9	3
366	1619	xenon	left	20	adult	0	61.83	1	12	1
367	1627	xenon	left	10	juvenile	1	6.57	1	10	3
368	1627	xenon	left	10	juvenile	0	66.97	0	12	1
369	1636	argon	right	16	juvenile	1	38.87	1	6	2
370	1636	argon	right	16	juvenile	0	68.30	0	6	0
371	1640	xenon	left	10	juvenile	1	42.43	1	11	3
372	1640	xenon	left	10	juvenile	0	46.63	1	9	1
373	1643	xenon	right	11	juvenile	1	67.07	0	9	3
374	1643	xenon	right	11	juvenile	0	67.07	0	9	1
375	1649	argon	right	1	juvenile	1	2.70	1	10	2
376	1649	argon	right	1	juvenile	0	2.70	0	12	0
377	1666	argon	left	17	juvenile	1	63.80	0	6	2
378	1666	argon	left	17	juvenile	0	63.80	0	8	0
379	1672	argon	left	7	juvenile	1	32.63	0	9	2
380	1672	argon	left	7	juvenile	0	32.63	0	9	0
381	1683	xenon	right	29	adult	1	62.00	0	10	3
382	1683	xenon	right	29	adult	0	62.00	0	8	1
383	1688	xenon	left	5	juvenile	1	13.10	1	11	3
384	1688	xenon	left	5	juvenile	0	54.80	0	10	1
385	1705	xenon	left	1	juvenile	1	8.00	0	8	3
386	1705	xenon	left	1	juvenile	0	8.00	0	8	1
387	1717	argon	left	22	adult	1	51.60	0	12	2
388	1717	argon	left	22	adult	0	42.33	1	11	0
389	1727	argon	right	33	adult	1	49.97	0	9	2
390	1727	argon	right	33	adult	0	2.90	1	10	0
391	1746	argon	right	3	juvenile	1	45.90	0	10	2

X	id	laser	eye	age	type	trt	futime	status	risk	group
<int>	<int>	<fct>	<chr>	<int>	<fct>	<fct>	<dbl>	<int>	<int>	<fct>
392	1746	argon	right	3	juvenile	0	1.43	1	10	0
393	1749	argon	right	32	adult	1	41.93	0	9	2
394	1749	argon	right	32	adult	0	41.93	0	9	0

In [28]:

```
levels(dat$group)
```

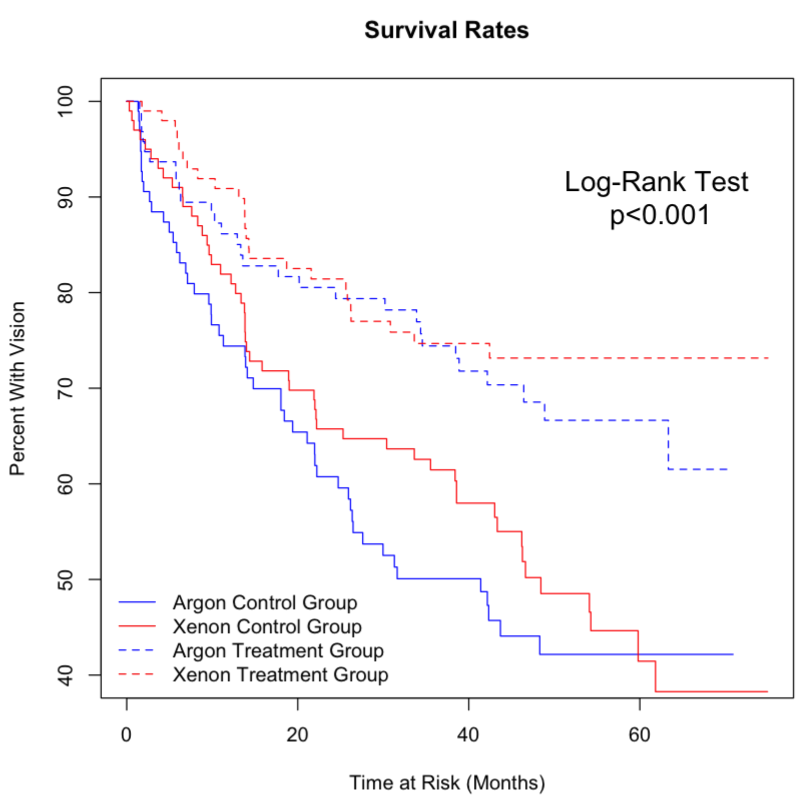
'0' · '1' · '2' · '3'

Kaplan-Meier Curves and Log Rank

In [76]:

```
survobj <- with(dat, Surv(futime, status))

# ph$compas <- cut(ph$decile_score, breaks=c(0,3,6,10))
fitc <- survfit(survobj~group, data=dat)
plot(fitc, xlab="Time at Risk (Months)",
   ylab="Percent With Vision", yscale=100, ylim=c(1, 0.4),
   main="Survival Rates",
   col = c('blue', 'red', 'blue', 'red'),
   lty = c('solid', 'solid', 'dashed', 'dashed')
)
legend_text = c('Argon Control Group', 'Xenon Control Group', 'Argon Treatment G
legend('bottomleft', legend=legend_text, bty='n',
       col=c('blue', 'red', 'blue', 'red'), lty=c('solid', 'solid', 'dashed', 'd
text(62, 0.9, 'Log-Rank Test\n p<0.001', cex=1.4)
```



In [6]:

```
dat$group = relevel(dat$group, ref="2")
```

In [77]:

```
finalfit(dat, "Surv(futime, status)", c("trt*laser", "age", "risk", "frailty(id)
```

A data.frame.ff: 9 × 5

	Dependent: Surv(futime, status)		all	HR (univariable)	HR (multivariable)
	<chr>	<chr>	<chr>	<chr>	<chr>
8	trt	0	197 (50.0)	-	-
9		1	197 (50.0)	-	-
3	laser	argon	194 (49.2)	-	-
4		xenon	200 (50.8)	0.85 (0.57-1.25, p=0.412)	0.75 (0.46-1.22, p=0.247)
1	age	Mean (SD)	20.8 (14.8)	1.00 (0.99-1.01, p=0.604)	1.01 (0.99-1.02, p=0.358)
5	risk	Mean (SD)	9.7 (1.5)	1.16 (1.04-1.29, p=0.009)	1.19 (1.03-1.36, p=0.014)
2	frailty(id)			-	-
6	NA	NA	NA	0.47 (0.30-0.74, p=0.001)	0.38 (0.24-0.62, p<0.001)
7	trt:laserxenon	Interaction	NA	0.95 (0.49-1.83, p=0.869)	1.07 (0.54-2.11, p=0.854)

In [21]:

```
finalfit(dat, "Surv(futime, status)", c("risk"))
```

A data.frame.ff: 1 × 5

	Dependent: Surv(futime, status)		all	HR (univariable)	HR (multivariable)
	<chr>	<chr>	<chr>	<chr>	<chr>
1	risk	Mean (SD)	9.7 (1.5)	1.16 (1.04-1.29, p=0.009)	1.16 (1.04-1.29, p=0.009)

Interpret coefficients of Interaction Term

In [84]:

```
summary(coxph(survobj~trt*laser + age + risk + frailty(id), data=dat))
```

```
Call:
coxph(formula = survobj ~ trt * laser + age + risk + frailty(id),
    data = dat)

  n= 394, number of events= 155

                  coef      se(coef) se2      Chisq   DF    p
```

```
trt1             -0.954928 0.243161 0.238255  15.42  1.0 8.6e-05
laserxenon       -0.291249 0.251438 0.204767   1.34  1.0 2.5e-01
age               0.006777 0.007375 0.005625   0.84  1.0 3.6e-01
risk              0.169901 0.069378 0.059500   6.00  1.0 1.4e-02
frailty(id)                                  107.86 79.5 1.9e-02
trt1:laserxenon   0.064234 0.348640 0.342208   0.03  1.0 8.5e-01

                 exp(coef) exp(-coef) lower .95 upper .95
trt1                0.3848     2.5985    0.2389    0.6198
laserxenon          0.7473     1.3381    0.4565    1.2233
age                 1.0068     0.9932    0.9924    1.0215
risk                1.1852     0.8437    1.0345    1.3578
trt1:laserxenon     1.0663     0.9378    0.5384    2.1118

Iterations: 6 outer, 31 Newton-Raphson
     Variance of random effect= 0.7990444   I-likelihood = -846.8
Degrees of freedom for terms=  1.0  0.7  0.6  0.7 79.5  1.0
Concordance= 0.838  (se = 0.016 )
Likelihood ratio test= 202  on 83.4 df,   p=8e-12
```

Model without Interaction Term

In [79]:
```
summary(coxph(survobj~group + age + risk + frailty(id), data=dat))
```

```
Call:
coxph(formula = survobj ~ group + age + risk + frailty(id), data = dat)

  n= 394, number of events= 155

             coef      se(coef) se2      Chisq  DF   p
group1       -0.291249 0.251438 0.204767   1.34  1.0 2.5e-01
group2       -0.954928 0.243161 0.238255  15.42  1.0 8.6e-05
group3       -1.181943 0.288699 0.250297  16.76  1.0 4.2e-05
age           0.006777 0.007375 0.005625   0.84  1.0 3.6e-01
risk          0.169901 0.069378 0.059500   6.00  1.0 1.4e-02
frailty(id)                              107.86 79.5 1.9e-02

       exp(coef) exp(-coef) lower .95 upper .95
group1    0.7473     1.3381    0.4565    1.2233
group2    0.3848     2.5985    0.2389    0.6198
group3    0.3067     3.2607    0.1742    0.5400
age       1.0068     0.9932    0.9924    1.0215
risk      1.1852     0.8437    1.0345    1.3578

Iterations: 6 outer, 31 Newton-Raphson
     Variance of random effect= 0.7990444   I-likelihood = -846.8
Degrees of freedom for terms=  2.5  0.6  0.7 79.5
Concordance= 0.838  (se = 0.016 )
Likelihood ratio test= 202  on 83.32 df,   p=7e-12
```

In [14]:
```
library(finalfit)
# install.packages("finalfit")
```

In [ ]: