STATS 504 HW2

Statistical Analysis for Diabetic Retinopathy

February 2022

1 Introduction

Diabetic Retinopathy is a diabetes complication that affects the eyes. It is caused by damage to the blood vessels at the retina or back of the eye. Diabetic retinopathy may have no symptoms or only mild vision problems at first. However, it can lead to blindness in the end. This can happen to anyone who has type 1 or 2 diabetes. Moreover, the longer the length of the diabetes history and the less controlled the blood sugar is, the chance of diabetic retinopathy is higher.
Our statistical analysis report will focus on two laser treatments: argon and xenon. We need to identify and quantify the efficacy of treatment type on visual acuity and the improvement between eyes by treatment type. Also, we need to understand the potential impact that age at diagnosis and clinical risk of diabetic retinopathy have on visual acuity. In this way, our analysis result can help our client get a better result on treatment to delay diabetic retinopathy.
The results of our models shows that:

- Both treatment types are similar in their efficacy on visual acuity.
- Higher clinical risk is associated with a higher risk of vision loss.

2 Methods

2.1 Purpose

There are two purposes for our analysis. The main purpose is to determine the efficacy of the treatment type of visual acuity and quantify the improvement based on each treatment type. Moreover, the secondary purpose is to determine the potential impact of age at diagnosis and clinical risk of diabetic retinopathy on the loss of vision.

2.2 Statistical Method

2.2.1 Survival Analysis

The first statistical method we will use is survival analysis. Survival analysis is used to analyze the time until some event occurs especially when follow up is not complete for each observation.
Since we are interested in the visual acuity in each eye dropping below 5/200 for two visits in a row, indicating a loss of vision in the eye. We use this to examine the efficacy of the argon and xenon laser treatments. However, this outcome occurs over time, and the data set is censored, meaning that the exact event time is unknown. In our case, the time of vision loss is unknown for some subjects. When following a subject through the study, the subject can drop out, pass away, or the study ends before the event of interest occurs; thus, the

subject is lost to follow-up, leading to an unknown event time. Due to our data and outcome, we use survival analysis.
In this case, we model the time until the loss of vision. We model this in two parts: the survival function and the hazard. The survival function gives the probability of not having an event before time t while the gives the instantaneous probability of an event at time t given survival until time t.
For the mathematical theory, assume that T is the time until an event randomly occurs.

$$T \sim f$$

That is, $f(t)$ is the probability density function (pdf) of T where t is time $F(t) = P(T < t) = \int_0^t f(x)dx$ is cumulative distribution function (cdf) of T.

- Survival function: $S(t) = P(T > t) = 1 - F(t)$
- Hazard function: $\lambda(t) = \lim_{h\to 0} \frac{P(T\leq t+h|T>t)}{h} = \lim_{h\to 0} \frac{P(t<T\leq t+h)}{hP(T>t)} = \frac{f(t)}{S(t)} = -\frac{d\log S(t)}{dt}$
- Cumulative Hazard function: $\Lambda(t) = \int_0^t \lambda(x)dx = -\int_0^t d\log S(x) = -\log S(t)$

2.2.2 Kaplan-Meier (K-M) curve

There are two types of survival analysis methods in our analysis. The first is Kaplan-Meier (K-M) curve. The Kaplan-Meier curve is used to estimate the survival function. It serves as a visual representation for the probability that an event will or will not occur over some time. This method uses a survival object as the response variable consisting of a time and a status.
Our models use the lag-corrected time to loss of vision and the status of vision together as the survival object. In our analysis, the K-M curve is first used to examine the survival function for vision loss across the different treatment types and the control groups.
For the mathematical theory, under this method, we have that:

- Approximated survival function: $\hat{S}(t) = \prod_{j=1}^{J} \left(1 - \frac{d_j}{n_j}\right)^{I\left(t_{(j)}\leq t\right)}$.
- Variance of the estimated survival function: $\hat{V}[\hat{S}(t)] = \hat{S}(t)^2 \sum_{j:t_0\leq t} \frac{d_j}{n_j(n_j-d_j)}$

2.2.3 Cox proportional-hazards model

Our model's second survival analysis method is the Cox proportional-hazards model. The Cox proportional-hazards model is a regression model used to examine the hazard ratio and relative risk. The Cox proportional-hazards model can be used to analyze which factors are potentially important to the event of interest. This method also uses a survival object as the response variable consisting of a time and a status.
In our case, after examining the K-M curve, we will use the Cox proportional-hazards model to quantify the effect of the different treatments. We also use the Cox proportional-hazards model to understand the importance of age at the time of diabetes diagnosis and the clinical risk of diabetic retinopathy to visual acuity.
Besides the Cox proportional-hazards model, we also included the frailty model in our analysis. From our data, we know that each subject contributes two data values to the data set, one for each eye. This could result in some association between the two data values. For example, the subject may have a condition that affects both the right and the left eye. Thus information about the right eye could provide information about the left eye. Based on the nature of our data, we consider the frailty model for dealing with this potential association.

For the mathematical theory, assuming that each individual hazard function is proportional to some common baseline hazard function makes the problem workable:

$$\lambda\left(t \mid x_i\right) = \lambda_0(t) \exp\left(\beta x_i\right)$$

where x_i is the covariate vector for participant i and β is the parameter vector to be estimated $\lambda_0(t)$ is the hazard function for $x_i = (0, \ldots, 0)$ $\exp\left(\beta x_i\right)$ explains proportional differences in hazards as x_i changes as in parametric regression Then the probability that individual j experiences an event at $t_{(j)}$ given survival until $t_{(j)}$ is

$$\lambda\left(t_{(j)} \mid x_{(j)}\right) = \lambda_0\left(t_{(j)}\right) \exp\left(x_{(j)}\beta\right)$$

The total probability within the sample of an event occurring at time $t_{(j)}$ given those who have survived until $t_{(j)}$ is

$$\sum_{k:t_k \geq t_{(j)}} \lambda\left(t_{(j)} \mid x_k\right) = \sum_{k:t_k \geq t_{(j)}} \lambda_0\left(t_{(j)}\right) \exp\left(x_k\beta\right)$$

Then probability of an event occurring at $t_{(j)}$ conditioning on covariates $x_{(j)}$ (partial likelihood) is

$$\tilde{L}_j(\beta) = P\left[(j) \text{ fails } \mid 1 \text{ failure from those at risk at } t_{(j)}\right] = \frac{P[(j) \text{ fails } \mid \text{ still at risk }]}{\sum_{k:t_k \geq t_{(j)}} P(k \text{ fails } \mid \text{ still at risk })}$$

$$= \frac{\lambda\left(t_{(j)} \mid x_{(j)}\right)}{\sum_{k:t_k \geq t_{(j)}} \lambda\left(t_{(j)} \mid x_k\right)} = \frac{\lambda_0\left(t_{(j)}\right) \exp\left(x_{(j)}\beta\right)}{\sum_{k:t_k \geq t_{(j)}} \lambda_0\left(t_{(j)}\right) \exp\left(x_k\beta\right)} = \frac{\exp\left(x_{(j)}\beta\right)}{\sum_{k:t_k \geq t_{(j)}} \exp\left(x_k\beta\right)}$$

3 Results

3.1 Data Overview

We first introduce the data set. The nine variables present in our data set are respectively: subject ID, type of laser treatment used (argon or xenon), eye that received treatment (left or right), age (at time of diabetes diagnosis in years), type (adult or juvenile at time of diabetes diagnosis), treated eye or control eye, lag-corrected time to loss of vision or last follow-up in months, status of vision or follow-up, and clinical risk of visual acuity loss.

This data includes 197 subjects and nine variables, one being the subject ID which serves as a unique identifier for each subject. Each subject contributes two rows of data, one for each eye. Thus, while there are 197 subjects, there are a total of 394 entries in the data set. There are no missing values for any of the variables. In Table 1 and 2, we present the summary statistics for the variables used in our analysis.

The variables eye and type are omitted, since

- type is directly correlated with the variable age at the time of diabetes diagnosis
- treated eye is the left or right eye should not have any effect on our results

Numerical Variables	
Lag-Corrected Time (months)	Mean: $35.580, \text{SD} : 21.356$
Age of Diabetes Diagnosis (years)	Mean: $20.780, \text{SD} : 14.812$
Clinical Risk of Vision Loss (unknown units)	Mean: $9.698, \text{SD} : 1.475$

Table 1: Summary Statistics of Numerical Variables

There are three numerical variables present in the data set, all other variables are categorical variables. From Table 2 and our knowledge of the data set, we know that 97 eyes were treated with argon and 100 were treated

Categorical Variables	
Variable	**Summary Statistics**
Status	Lost to follow-up: 239, Loss of vision: 155
Laser Treatment	Argon: 194, Xenon: 200
Treated Eye	Treated eye: 197, Control eye: 197

Table 2: Summary Statistics of Categorical Variables

with xenon, while the other 197 eyes were left as control eyes. Due to the limited number of observations in the data set and the fact that there were no visible issues with the data set, we use the entire data set for our analysis.

3.2 Model Findings

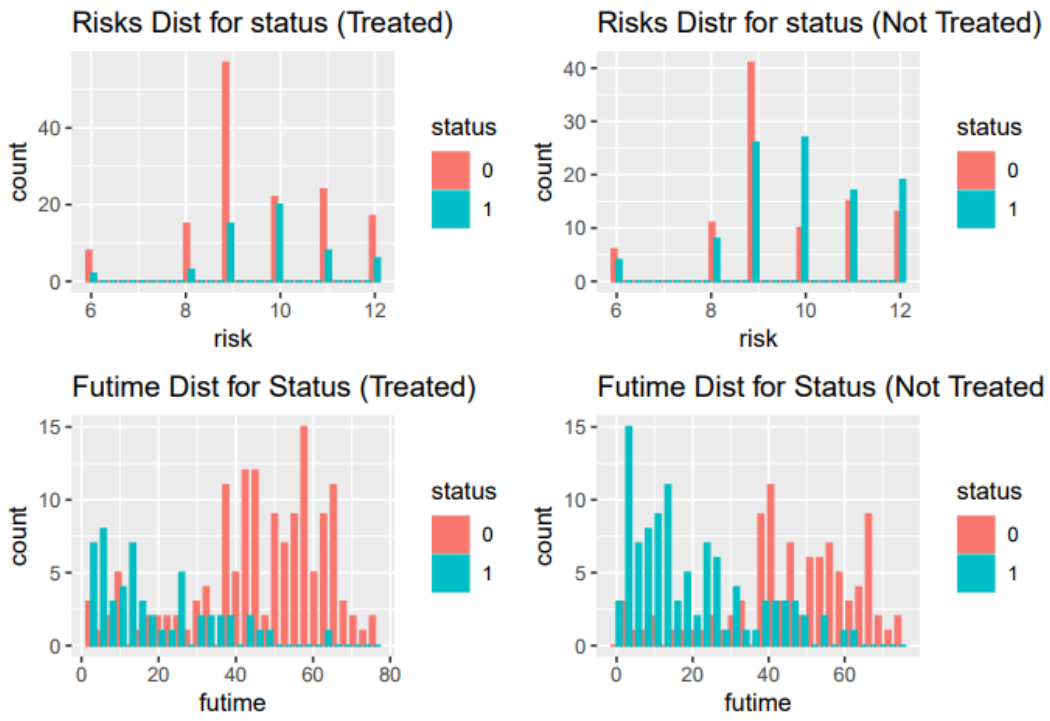


Figure 1: Distribution Comparison for Treated or not

Figure 1 displays the distribution examination for different variables under condition treated or not. From the distribution plots, we can have the intuitive that the distribution for the variables are different under the treatment, this might shows the effect of the treatments.

Figure 2 displays the K-M curve for the probability of maintaining vision over time. We display four different curves total: two for the two different types of laser treatment, argon and xenon, and two for their respective control groups, in which no laser treatment is used. From the distribution, we can see that as time progresses, no matter which treatment is used, the probability of maintaining a vision for the treated eye is approximately 20% higher than that of the control.

We also see that between treated groups for argon and xenon, it seems that overall the two probabilities are quite similar; however, xenon seems to perform slightly better than argon as time progresses. Besides, based on Table 3, the proportion of xenon is smaller, which might also indicate xenon's advantage. On the other hand, for the control groups, the xenon control group seems to have a higher probability of maintaining vision at

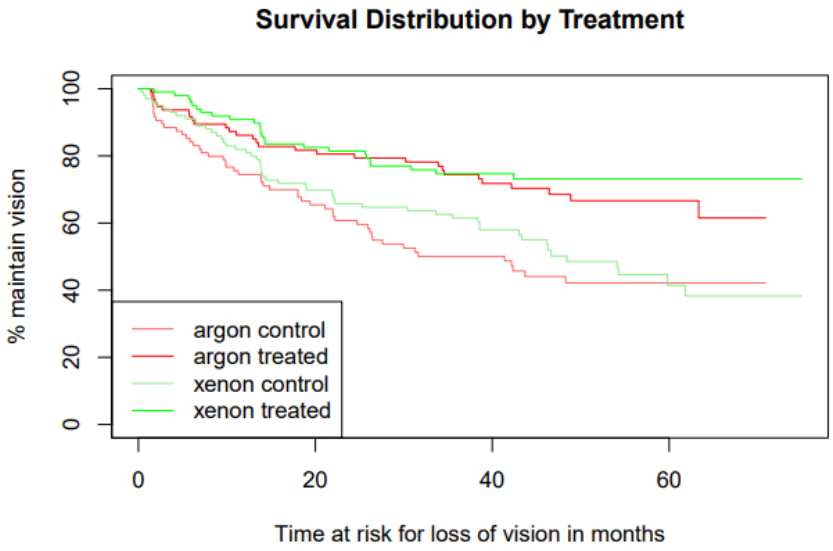


Figure 2: K-M curve of the probability of maintaining vision estimation

Treated but Lost Vision Cases	
argon	xenon
53.7%	46.3%

Table 3: Proportion of Different Laser Method in Treated but Lost Vision Cases

first; however, both control groups level off to achieve approximately the same probability as time progresses. With this being said, we cannot draw any conclusions directly from this distribution. Thus, we turn to the Cox proportional-hazards model.

Both the Cox proportional-hazards model and the frailty model give similar results regarding the inferences we can perform on the models. Both present the same rates and very similar confidence intervals. The significance of the factors considered in the model is also very similar. Due to the similarity in the results, we present the findings from the simpler Cox proportional-hazards model. Based on the rates presented in Table 4, we can interpret the variables. We see that at the 5% significance level, the two variables, treated eye and clinical risk, are statistically significant. Thus, interpreting these variables, we can see a 53.6% decrease in the expected risk of vision loss for eyes that receive treatment compared to eyes that do not receive treatment while holding all other variables constant. Similarly, for a one-unit increase in the clinical risk of acuity loss, there is a 15.8% increase in the expected risk of vision loss in the eye while holding all other variables constant.

Variable	Rate (95% CI)	p-value
Argon vs. Xenon (control)	$0.844(0.571, 1.249)$	0.397
Control Eye vs. Treated Eye	$0.464(0.294, 0.735)$	0.001
Age of Diabetes Diagnosis (years)	$1.005(0.995, 1.016)$	0.330
Clinical Risk	$1.158(1.038, 1.292)$	0.008
Argon vs. Xenon interaction with Treated Eye	$0.950(0.490, 1.842)$	0.879

Table 4: Rates and 95% confidence intervals of the risk of vision loss for the variable assessed along with their p-values.

Since the treated eye is significant and results in a decrease in vision loss, we can conclude that receiving treatment, whether argon or xenon treatment, is effectively associated with maintaining vision and delaying diabetic retinopathy. However, we see that between xenon laser treatment and argon laser treatment, the expected risk of vision loss is not significant. Thus, we cannot conclude that the two treatment types differ in

their efficacy on visual acuity. As we also saw, clinical risk of visual acuity loss is another significant variable that corresponds to an increase in vision loss. Thus the potential impact of clinical risk of acuity loss seems to be negative about vision loss. On the other hand, the age at the time of diabetes diagnosis does not appear to be significant, meaning that the age at the time of diabetes diagnosis does not seem to have much association with vision loss, which is most likely not very impactful to the later vision loss.

4 Conclusion

The purpose of this analysis was to determine the efficacy of the treatment type of visual acuity and quantify the improvement based on each treatment type. We also wanted to understand the potential impact of age at diagnosis and clinical risk of diabetic retinopathy on the loss of vision.

Our results found that while laser treatment was effective in delaying vision loss, the effect of argon and xenon treatments were not significantly different.

And for the secondary research goal, we found that the higher the clinical risk, the higher the probability of vision loss. While we cannot identify if there is direct causation between the two, we can see that the two are significantly associated. Age of diagnosis did not appear to be significantly associated with vision loss and thus is probably not influential to vision loss from diabetic retinopathy.

A potential limitation of this analysis is the small number of variables in the data set. While we see that treatment and clinical risk are significant in our model, we have not examined other variables that may also be factors associated with vision loss from diabetic retinopathy.

The inclusion of more factors could allow us to perform a deeper dive into the efficacy of the two treatments. Also, we assumed that treating the left or the right eye did not have an effect on the results; however, it might be the case that one eye is naturally better than the other. In the future, we hope to examine other factors that could lead to vision loss from diabetic retinopathy and the difference in the eyes. With this being said, the results from our analysis can inform us of the effectiveness of laser treatment and the potential impact of clinical risk on vision loss.

Appendix

This file details our data analysis and model fitting process, examining the relationship between different factors and vision loss from diabetic retinopathy. It also produces plots and tables for our statistical report.

First, we load the data and do the overall checking. There are no NaN values.

```
library(ggplot2)
# load the required data
data <- read.csv("diabeticVision.csv")
# check the dimensions of data
dim(data)
```

```
## [1] 394  10
```

```
# check the head of data
head(data)
```

```
##   X id laser   eye age     type trt futime status risk
## 1 1  5 argon  left  28    adult   1  46.23      0    9
## 2 2  5 argon  left  28    adult   0  46.23      0    9
## 3 3 14 argon right  12 juvenile   1  42.50      0    8
## 4 4 14 argon right  12 juvenile   0  31.30      1    6
## 5 5 16 xenon right   9 juvenile   1  42.27      0   11
## 6 6 16 xenon right   9 juvenile   0  42.27      0   11
```

Now, we can start the data analysis process. First we check the overall distribution of the data.

```
# check the length of unique id
length(unique(data$id))
```

```
## [1] 197
```

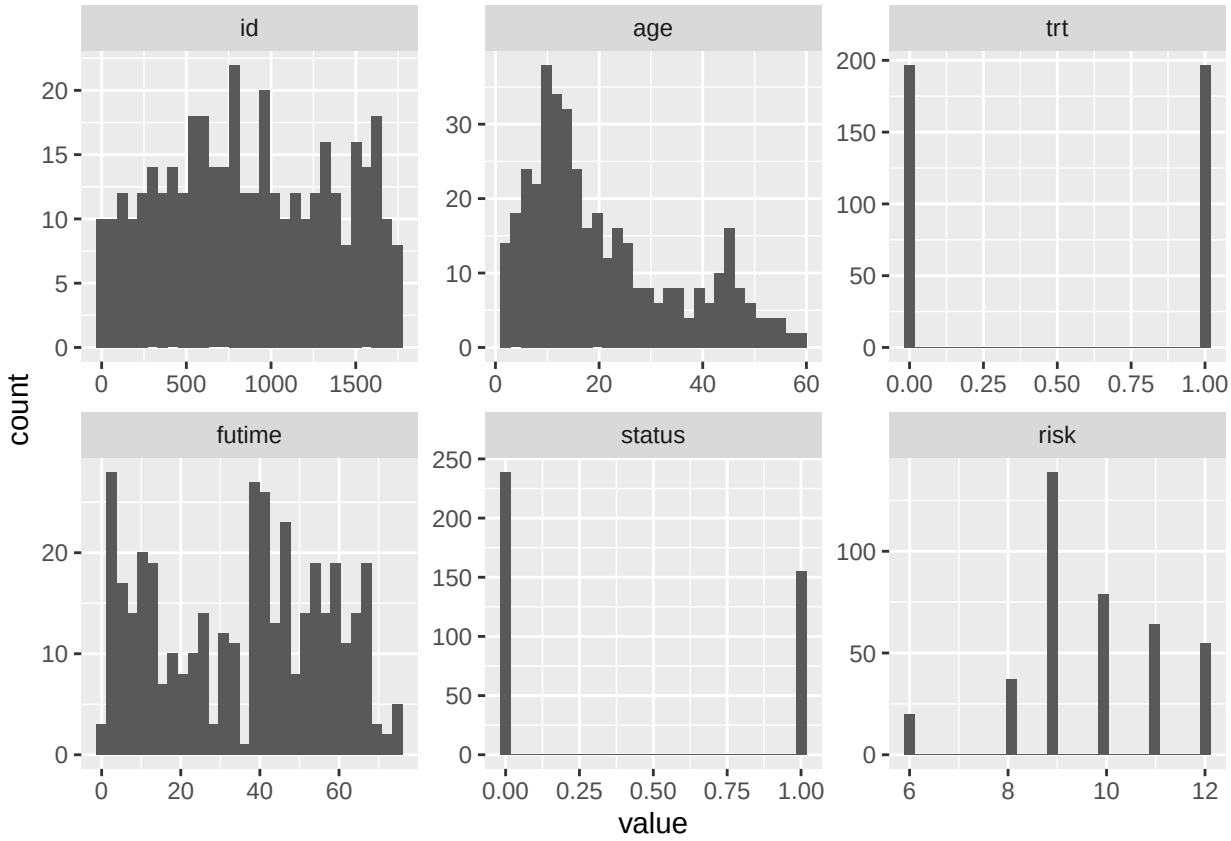
```
# distribution for laser
table(data$laser)/sum(!is.na(data$laser))*100
```

```
##
##    argon    xenon
## 49.23858 50.76142
```

```
# overall distribution
library(reshape2)
ggplot(melt(data[,-1]),  aes(x=value))+ geom_histogram(position='dodge') +
facet_wrap(~variable, scales ="free")
```

```
## Using laser, eye, type as id variables
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Above are the histograms of the data variables.

Some of the variables are categorical, and we change their type into categorical data. Then, we do the summary of the data.

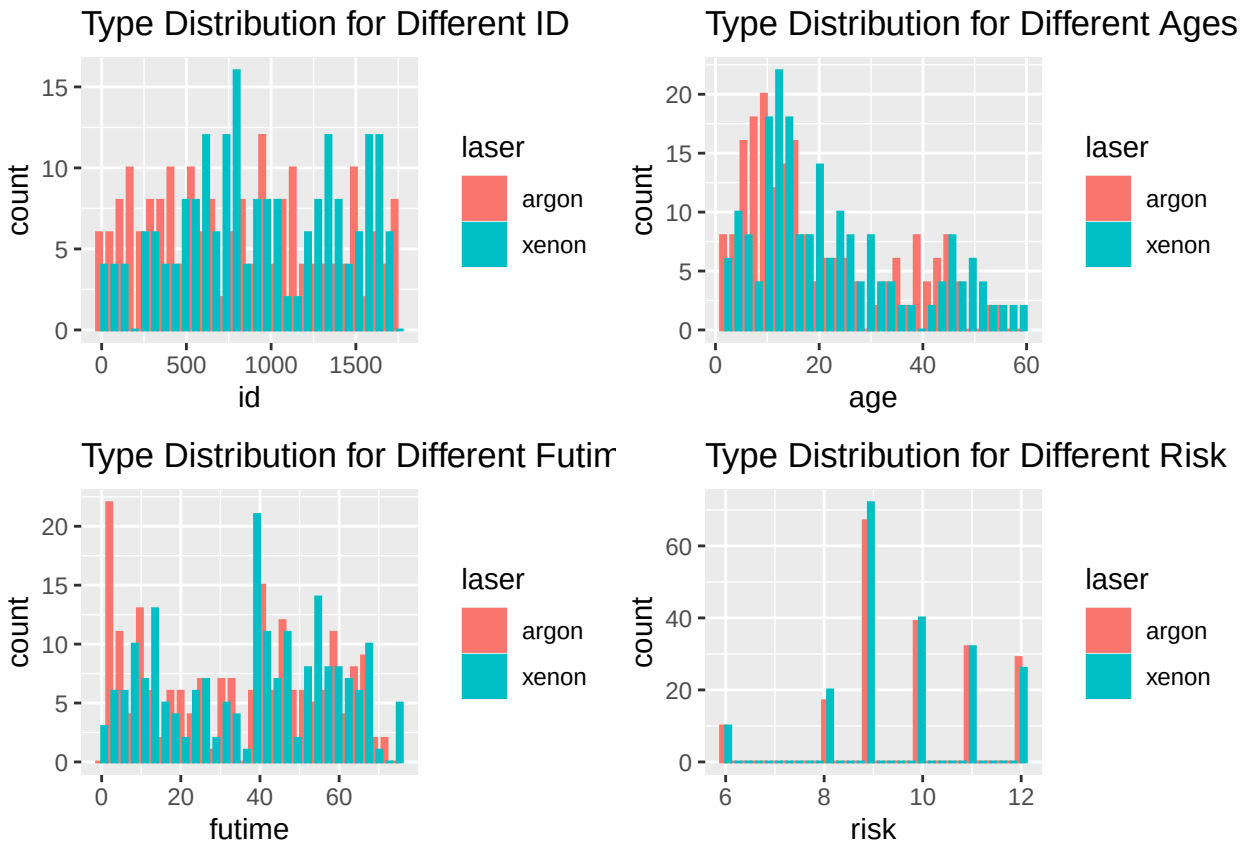
```
# factor reasonable datas
data$status <- factor(data$status)
data$laser <- factor(data$laser)
data$eye <- factor(data$eye)
data$type <- factor(data$type)
data$trt <- factor(data$trt)
summary(data[, -1])
```

```
##        id           laser        eye           age             type
##  Min.   :   5.0   argon:194   left :216   Min.   : 1.00   adult   :166
##  1st Qu.: 480.0   xenon:200   right:178   1st Qu.:10.00   juvenile:228
##  Median : 834.0                           Median :16.00
##  Mean   : 873.2                           Mean   :20.78
##  3rd Qu.:1296.0                           3rd Qu.:30.00
##  Max.   :1749.0                           Max.   :58.00
##  trt         futime      status       risk
##  0:197   Min.   : 0.30   0:239   Min.   : 6.000
##  1:197   1st Qu.:13.98   1:155   1st Qu.: 9.000
##          Median :38.80           Median :10.000
##          Mean   :35.58           Mean   : 9.698
##          3rd Qu.:54.25           3rd Qu.:11.000
##          Max.   :74.97           Max.   :12.000
```

Now we check the distribution split by treatment type.

```
library(gridExtra)
hist1 <- ggplot(data, aes(x=id, color=laser, fill=laser)) +
  geom_histogram(position="dodge") +
  ggtitle("Type Distribution for Different ID")

hist2 <- ggplot(data, aes(x=age, color=laser, fill=laser)) +
  geom_histogram(position="dodge") +
  ggtitle("Type Distribution for Different Ages")

hist3 <- ggplot(data, aes(x=futime, color=laser, fill=laser)) +
  geom_histogram(position="dodge") +
  ggtitle("Type Distribution for Different Futime")

hist4 <- ggplot(data, aes(x=risk, color=laser, fill=laser)) +
  geom_histogram(position="dodge") +
  ggtitle("Type Distribution for Different Risk")

grid.arrange(hist1, hist2, hist3, hist4,
             nrow = 2, ncol = 2)
```

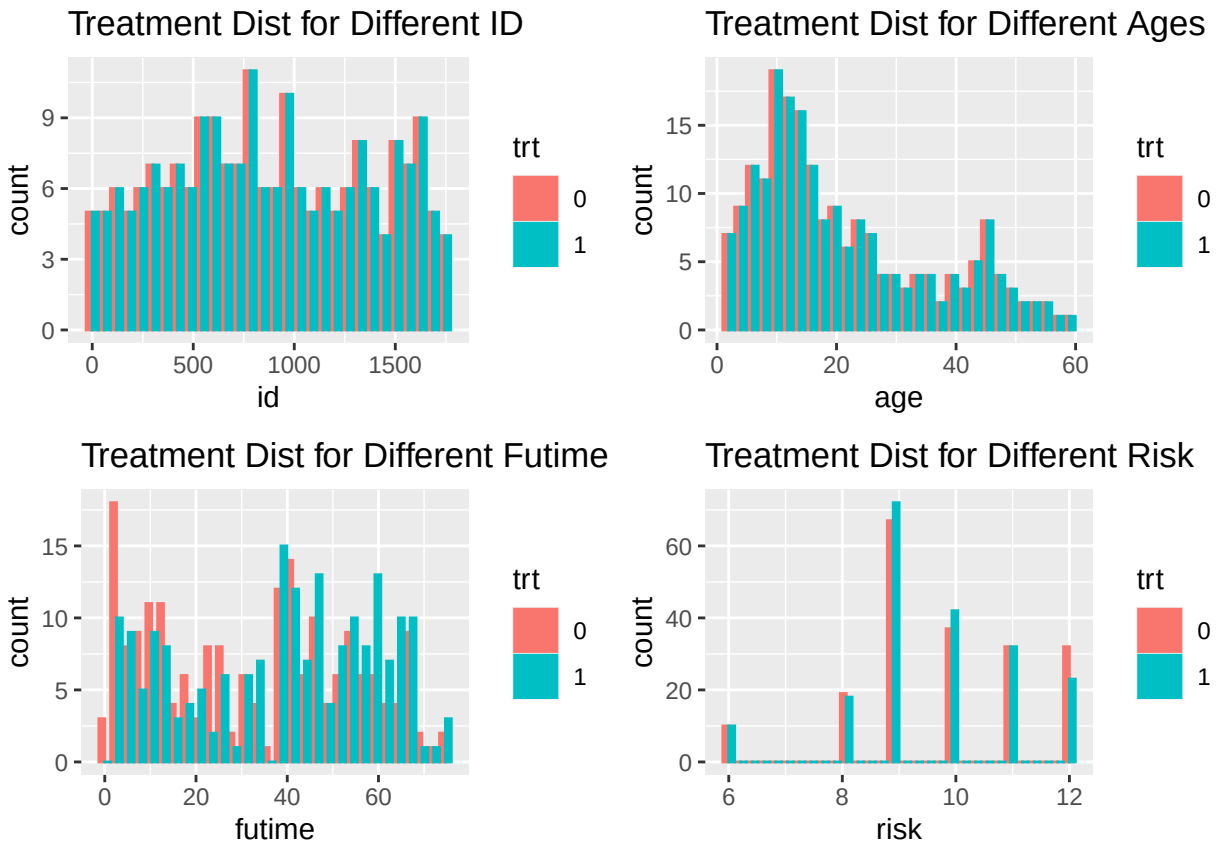
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Now we check the distribution split by treated eye and controled eye.

```
hist1 <- ggplot(data, aes(x=id, color=trt, fill=trt)) +
  geom_histogram(position="dodge") +
  ggtitle("Treatment Dist for Different ID")

hist2 <- ggplot(data, aes(x=age, color=trt, fill=trt)) +
  geom_histogram(position="dodge") +
  ggtitle("Treatment Dist for Different Ages")

hist3 <- ggplot(data, aes(x=futime, color=trt, fill=trt)) +
  geom_histogram(position="dodge") +
  ggtitle("Treatment Dist for Different Futime")

hist4 <- ggplot(data, aes(x=risk, color=trt, fill=trt)) +
  geom_histogram(position="dodge") +
  ggtitle("Treatment Dist for Different Risk")

grid.arrange(hist1, hist2, hist3, hist4,
             nrow = 2, ncol = 2)
```

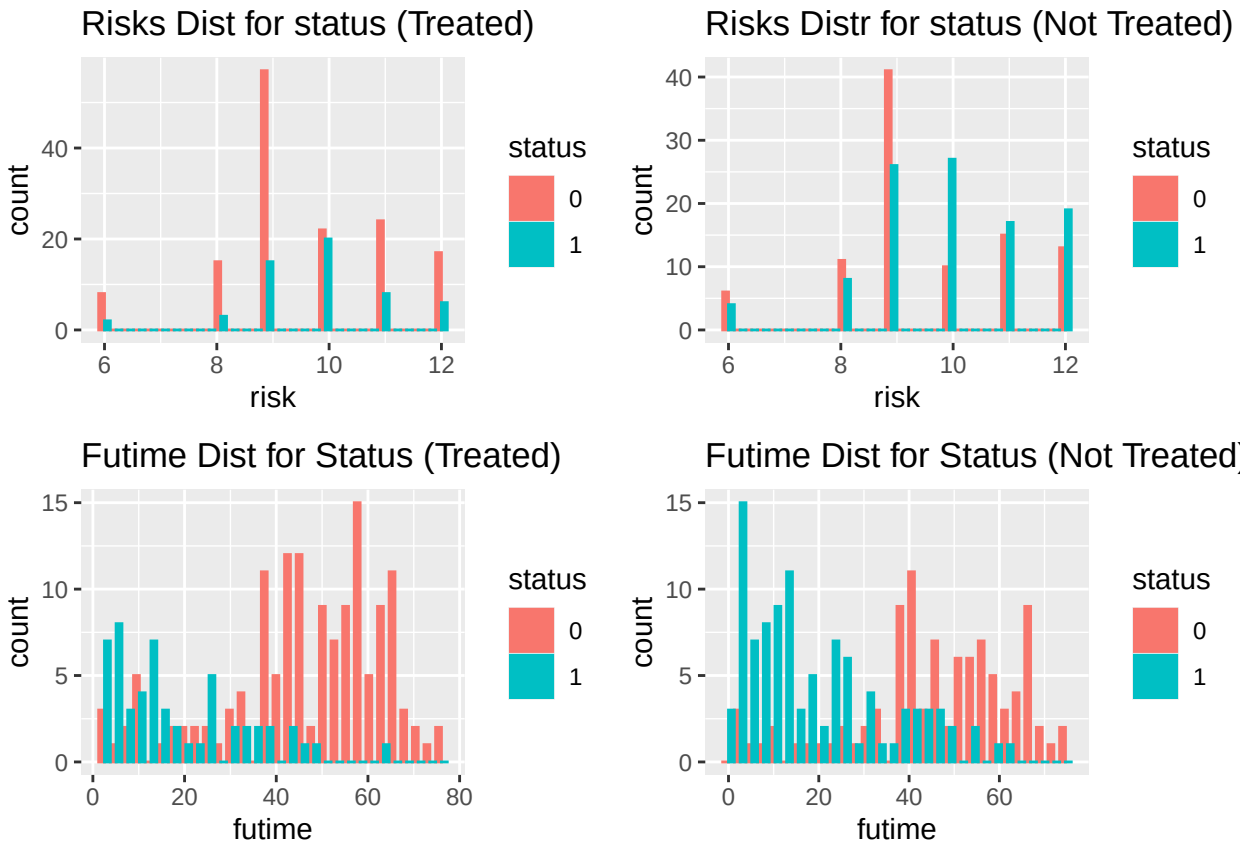
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Now we check the distribution first split by treated eye and controled eye and then split by status.

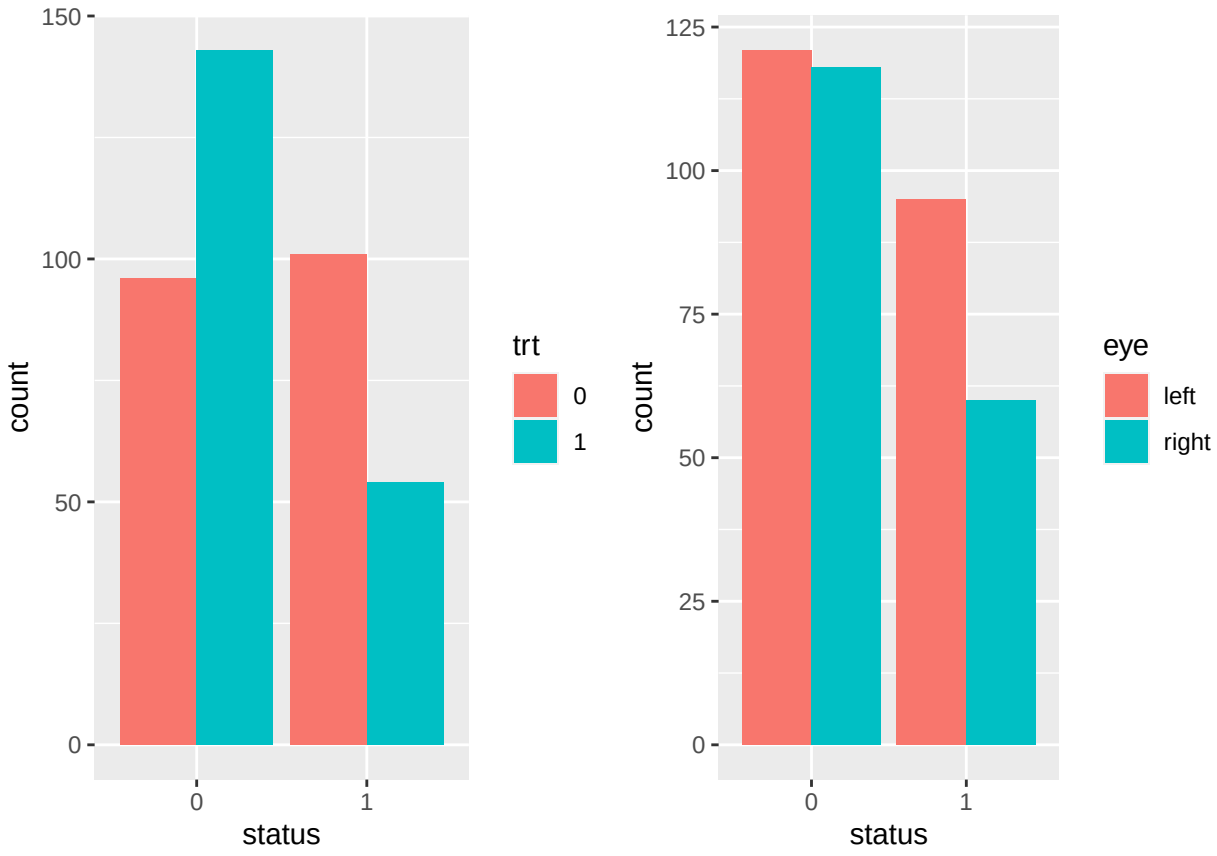
```
data_t <- data[data$trt==1,]
data_nt <- data[data$trt==0,]
hist5 <- ggplot(data_t, aes(risk, color=status, fill=status)) +
  geom_histogram(position="dodge")+
  ggtitle("Risks Dist for status (Treated)")
hist7 <- ggplot(data_t, aes(futime, color=status, fill=status)) +
  geom_histogram(position="dodge")+
  ggtitle("Futime Dist for Status (Treated)")
hist6 <- ggplot(data_nt, aes(risk, color=status, fill=status)) +
  geom_histogram(position="dodge")+
  ggtitle("Risks Distr for status (Not Treated)")
hist8 <- ggplot(data_nt, aes(futime, color=status, fill=status)) +
  geom_histogram(position="dodge")+
  ggtitle("Futime Dist for Status (Not Treated)")

# combined plot
grid.arrange(hist5, hist6, hist7, hist8,
             nrow = 2, ncol = 2)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Now we check the distribution across eyes of treatment and treatment or not.

```
p1 <- ggplot(data, aes(x=status, fill=trt)) + geom_bar(position = "dodge")
p2 <- ggplot(data, aes(x=status, fill=eye)) + geom_bar(position = "dodge")
grid.arrange(p1, p2, nrow = 1, ncol = 2)
```



Here is the pair plot of the variables.

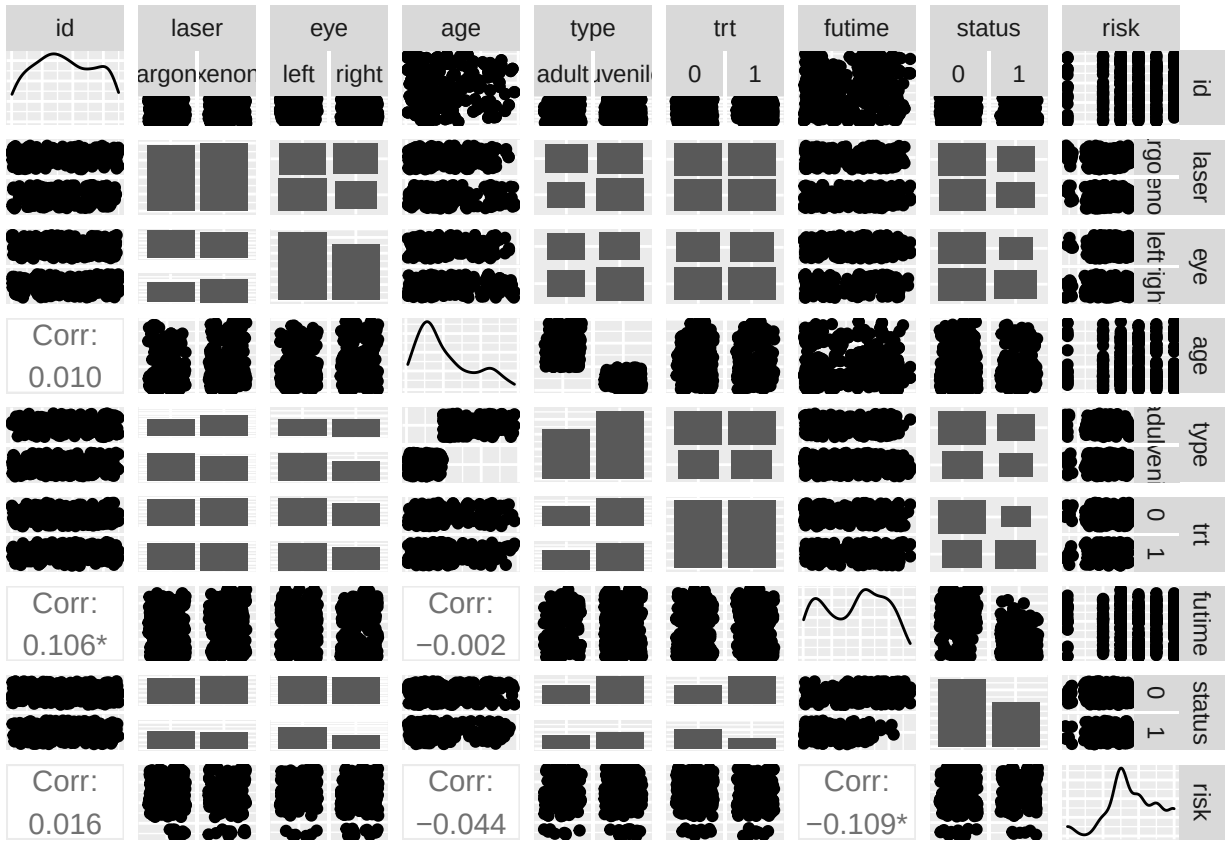
```
library(corrplot)
```

```
## corrplot 0.91 loaded
```

```
library(GGally)
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
ggpairs(data[,-1], axisLabels = 'none',
upper = list(continuous = 'points', combo = 'dot'),
lower = list(continuous = 'cor', combo= 'dot'),
diag =list(continuous = 'densityDiag'))
```




```
data_ts <- data_t[data_t$status==1,]
# distribution for laser in treated but lost vision eye
table(data_ts$laser)/sum(!is.na(data_ts$laser))*100
```

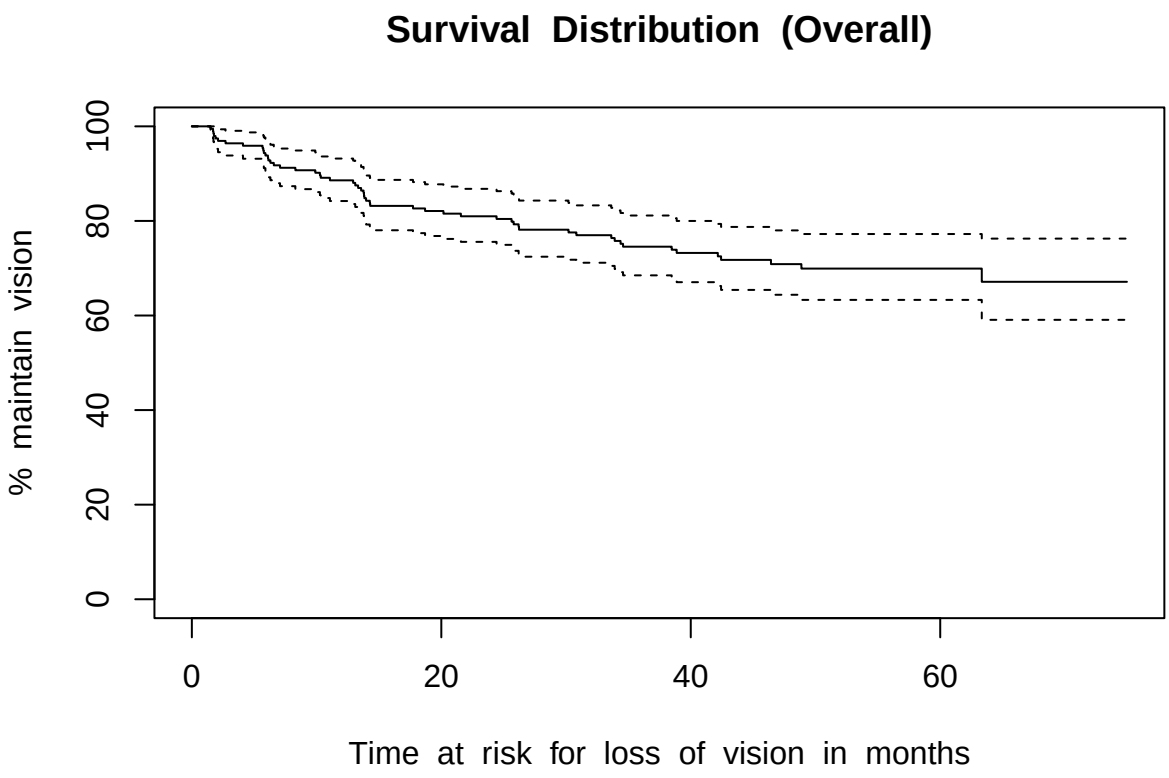
```
##
##   argon   xenon
## 53.7037 46.2963
```

From the above distribution, we can see that xenon type seems have less people loss their eyesight.

Plot Kaplan-Meier Curves

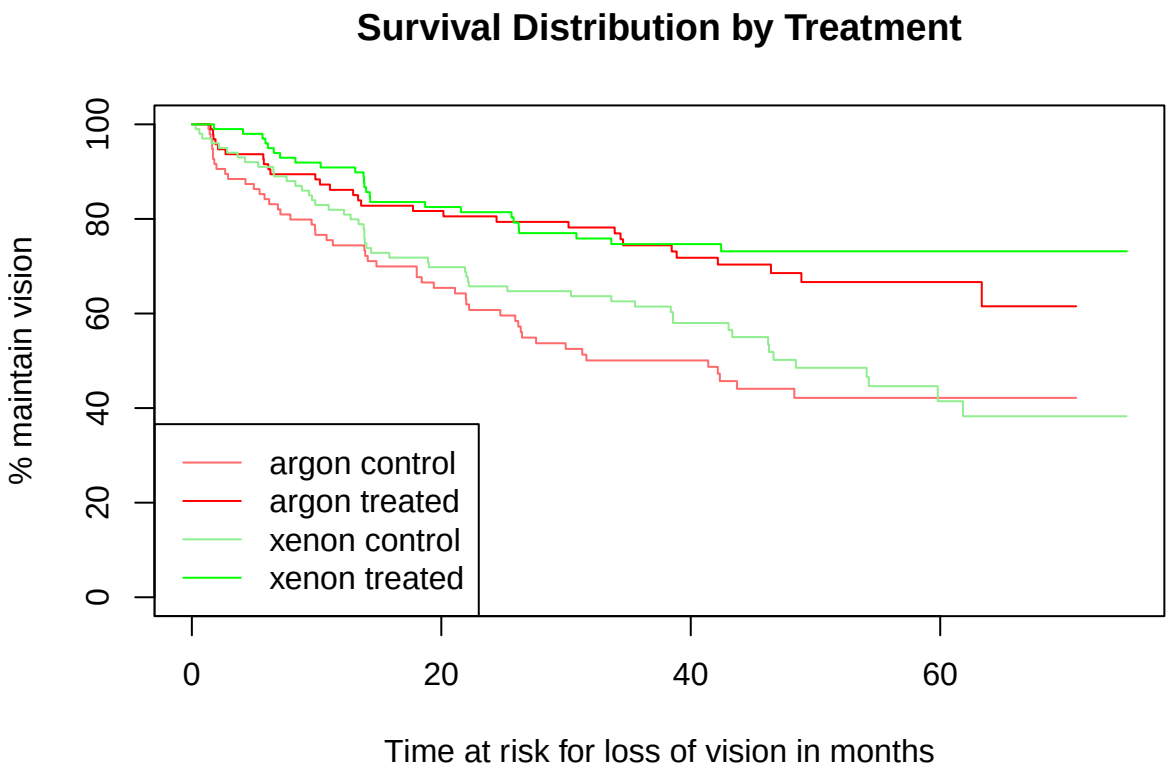
Here we examine the overall K-M curve across all different groups just to gain a general picture.

```
library(survival)
data <- read.csv("diabeticVision.csv")
survobj <- with(data[data$trt == 1, ], Surv(futime, status))
fit0 <- survfit(survobj ~ 1, data = data[data$trt == 1, ])
plot(fit0, xlab="Time at risk for loss of vision in months",
ylab= "% maintain vision", yscale=100,
main = "Survival Distribution (Overall)")
```



Here we examine the survival distribution for each of the different types of laser treatment along with their respective control group.

```
#   create   new   variable   for   the   4   different   "treatment   types"
data$treatment <- as.character(data$laser)
data$treatment <- ifelse(data$trt == 0,paste(data$treatment, "control"), paste(data$treatment, "treated
survobj <- with(data, Surv(futime, status))
fitr <- survfit(survobj   ~   treatment, data = data)
plot(fitr, xlab="Time at risk for loss of vision in months",
ylab="% maintain vision", yscale=100,
main="Survival Distribution by Treatment", col = c('indianred1', 'red', 'lightgreen', 'green'))
legend('bottomleft', legend=levels(as.factor(data$treatment)),
col = c('indianred1', 'red', 'lightgreen', 'green'),
lty=1)
```



Fit Cox Proportional-Hazards Models

First fit the model without accounting for association between observations.

```
mod <- coxph(Surv(futime, status) ~ laser*trt + age + risk, data)
summary(mod)
```

```
## Call:
## coxph(formula = Surv(futime, status) ~ laser * trt + age + risk,
##     data = data)
##
##   n= 394, number of events= 155
##
##                     coef exp(coef)  se(coef)      z Pr(>|z|)
## laserxenon     -0.169242  0.844304  0.199811 -0.847  0.39699
## trt            -0.766822  0.464487  0.234227 -3.274  0.00106 **
## age             0.005334  1.005349  0.005473  0.975  0.32976
## risk            0.146871  1.158204  0.055728  2.635  0.00840 **
## laserxenon:trt -0.051551  0.949755  0.338028 -0.153  0.87879
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
##                exp(coef) exp(-coef) lower .95 upper .95
## laserxenon        0.8443     1.1844    0.5707    1.2490
```

```
## trt               0.4645     2.1529    0.2935    0.7351
## age               1.0053     0.9947    0.9946    1.0162
## risk              1.1582     0.8634    1.0384    1.2919
## laserxenon:trt    0.9498     1.0529    0.4896    1.8422
##
## Concordance= 0.626  (se = 0.023 )
## Likelihood ratio test= 31.48  on 5 df,   p=8e-06
## Wald test            = 29.99  on 5 df,   p=1e-05
## Score (logrank) test = 31.29  on 5 df,   p=8e-06
```

Fit the model while accounting for association between observations using the cluster.

```
mod_c <- coxph(Surv(futime, status) ~ laser*trt + age + risk + cluster(id), data)
summary(mod_c)
```

```
## Call:
## coxph(formula = Surv(futime, status) ~ laser + trt + age + risk +
##     laser:trt, data = data, cluster = id)
##
##   n= 394, number of events= 155
##
##                     coef exp(coef)  se(coef) robust se      z Pr(>|z|)
## laserxenon     -0.169242  0.844304  0.199811  0.196118 -0.863 0.388159
## trt            -0.766822  0.464487  0.234227  0.210918 -3.636 0.000277 ***
## age             0.005334  1.005349  0.005473  0.006100  0.875 0.381832
## risk            0.146871  1.158204  0.055728  0.058597  2.506 0.012195 *
## laserxenon:trt -0.051551  0.949755  0.338028  0.297940 -0.173 0.862632
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                exp(coef) exp(-coef) lower .95 upper .95
## laserxenon        0.8443     1.1844    0.5749    1.2400
## trt               0.4645     2.1529    0.3072    0.7023
## age               1.0053     0.9947    0.9934    1.0174
## risk              1.1582     0.8634    1.0325    1.2992
## laserxenon:trt    0.9498     1.0529    0.5297    1.7030
##
## Concordance= 0.626  (se = 0.023 )
## Likelihood ratio test= 31.48  on 5 df,   p=8e-06
## Wald test            = 30.72  on 5 df,   p=1e-05
## Score (logrank) test = 31.29  on 5 df,   p=8e-06,   Robust = 29.82  p=2e-05
##
##   (Note: the likelihood ratio and score tests assume independence of
##      observations within a cluster, the Wald and robust score tests do not).
```

Fit the model while accounting for association between observations using the frailty options.

```
mod_f <- coxph(Surv(futime, status) ~ laser*trt + age + risk + frailty(id), data)
summary(mod_f)
```

```
## Call:
## coxph(formula = Surv(futime, status) ~ laser * trt + age + risk +
```

```
##     frailty(id), data = data)
##
##   n= 394, number of events= 155
##
##                 coef      se(coef) se2      Chisq  DF   p
## laserxenon     -0.291249 0.251438 0.204767   1.34  1.0 2.5e-01
## trt            -0.954928 0.243161 0.238255  15.42  1.0 8.6e-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
## laserxenon:trt  0.064234 0.348640 0.342208   0.03  1.0 8.5e-01
##
##                 exp(coef) exp(-coef) lower .95 upper .95
## laserxenon         0.7473     1.3381    0.4565    1.2233
## trt                0.3848     2.5985    0.2389    0.6198
## age                1.0068     0.9932    0.9924    1.0215
## risk               1.1852     0.8437    1.0345    1.3578
## laserxenon:trt     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=  0.7  1.0  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
```

Check to ensure that we do in fact have proportional hazards.

```
test.ph <- cox.zph(coxph(Surv(futime, status) ~ laser*trt + age + risk, data))
test.ph
```

```
##           chisq df    p
## laser     0.663  1 0.42
## trt       0.657  1 0.42
## age       0.355  1 0.55
## risk      1.606  1 0.21
## laser:trt 0.390  1 0.53
## GLOBAL    3.731  5 0.59
```