Stats 504, Assignment 2: Treating diabetic retinopathy

November 09, 2021

Introduction

Diabetic retinopathy is an eye condition that can cause vision loss and blindness in people who have diabetes. It is very likely to develop such symptom if a patient does not control their blood sugar properly. While diabetic retinopathy may cause no symptoms or only mild vision problems in the early stage, it can lead to blindness. Moreover, having high blood sugar levels for a long time can increase the probability of causing damage to the blood vessels of the light-sensitive tissue at the retina, or back of the eye, and thus resulting in more serious condition. Therefore, in order to delay diabetic retinopathy, researchers consider laser coagulation as an option for treatment. In this study, we are going to investigate the efficacy of treatment type on visual acuity and to quantify the improvement. Also, we would like to investigate the potential affect some factors have on visual acuity. Those factors include the age at diagnosis and clinical risk of diabetic retinopathy.

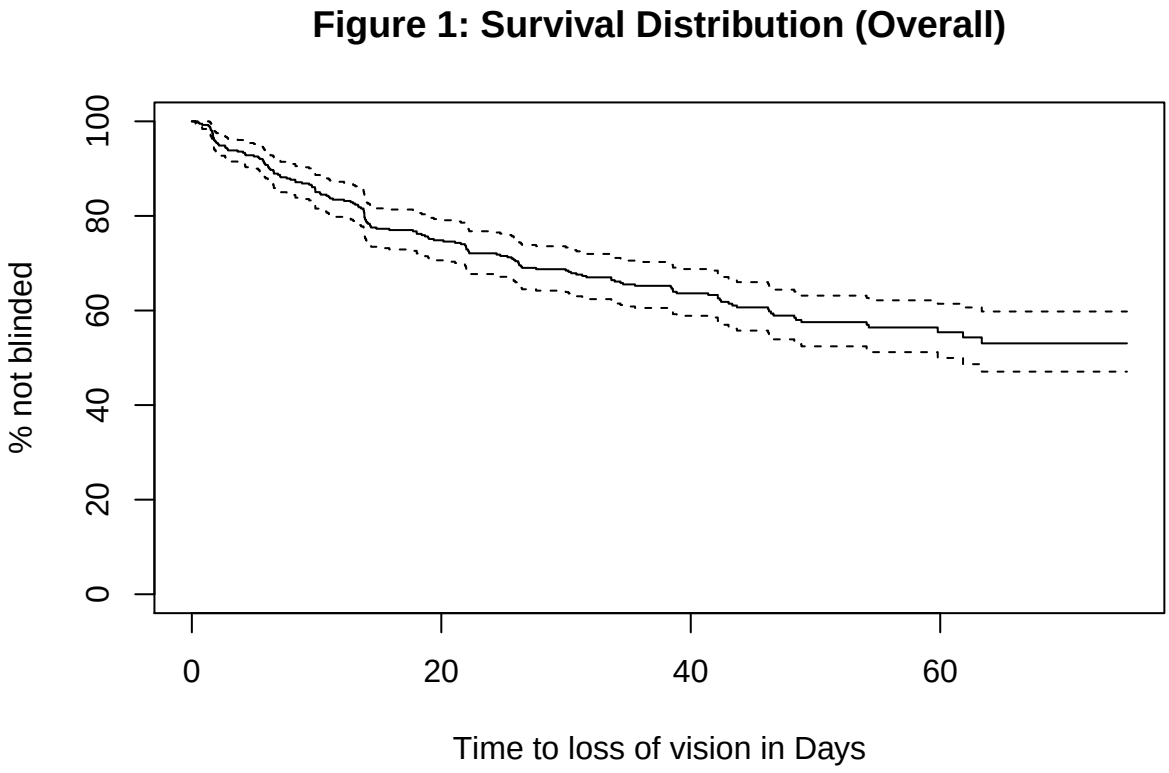
Exploratory Data Analysis

This data set records the condition of both eyes of patients, with one eye treated with the experimental treatment and the other being used as control. Note that, the treatment consists of two types of laser: xenon and argon. Besides, the data set contains 197 patients and is therefore composed of 394 observations, two eyes per patient. Column `futime` stands for particular time to loss of vision, and the whole data set is right censored, which means that meaning that the exact event time of vision loss is unknown for some subjects. These subject can either dropout, pass away, or the study ends before the event of interest occurs. Therefore, we can consider a survival model for this data.

Note that we have 80 patients having both eyes censored, and we have a total of 155 events. Among those patients who had an event, 16 patients having an event for the treated eye and not for the control eye, 63 patients with the opposite situations, and 38 patients with an event for both eye. Since each patient contributes to two rows in the data(one for each eye), these data are somewhat correlated, or more precisely, clustered, and having cluster size equal to 2.

Therefore, we should take factors mentioned above into account to make sure our model can describe the structure that is behind the data. Variable `eye` and `type` are removed from our model, because based on a assumption that two eyes are roughly the same, `eye` is irrelative to our analysis. The `type` variable is correlated with `age`, and since we include `age` here, we no longer keep `type`.

A simple fit with respect to time only is plotted below to give a basic understanding of how the survival function behaves like. This curve is also known as "Kaplan-Meier Curves", which is used to estimate the survival function and serves as a visual representation for the probability that an event will occur over a period of time. Here specifically, in Figure 1, the plot shows the probability of loss of vision in days.



Model Description

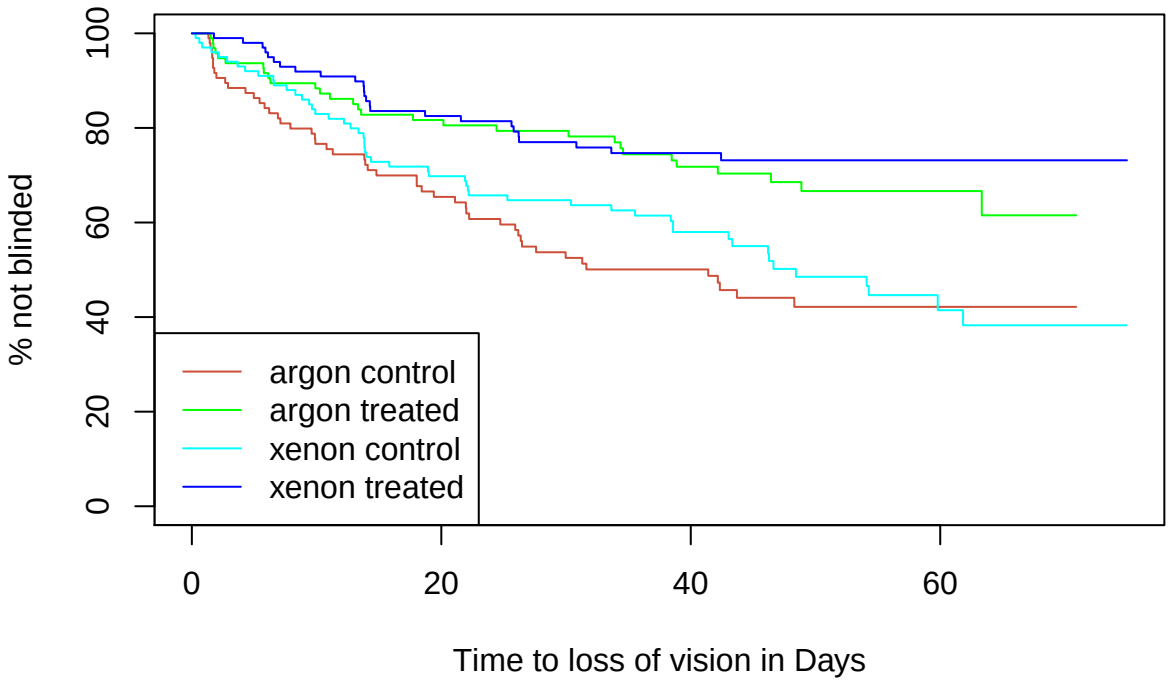
As we said in the previous section, we will use a survival model for this data. A survival model consists of two part, survival function and the hazard function. The survival function examines the probability the event has not occurred at a specific time while the hazard function is the instantaneous rate of occurrence of the event given that it has not occurred in the specified time before.

We already gave Kaplan-Meier (K-M) curve above in our analysis, which corresponds to a simple but intuitive survival analysis method that provides us an overview of the survival function. Now, we can make a fit with respect to laser treatment to examine the treatment effect given controlled eyes by a naive survival model.

The results are plotted below in Figure 2. From the Figure 2, we see that first of all, the curve of treated eyes is always above the controlled eyes, regardless of laser type, which indicates that the treatment is effective at delaying diabetic retinopathy. Then, comparing Argon treated with Xenon treated, we see that the blue curve, at most of the time, especially after 35 days, is above the green curve. This indicates that the efficacy of laser Xenon may slightly outweigh that of laser Argon, simply because more people are not blinded if treated with Xenon. However, in order to better quantify such behavior, we may need to consider using a more sophisticated model to determine such judgment in a statistical sense.

Therefore, next, let us consider cox proportional-hazards model, which takes the effect of factors into account properly, meaning that more information is involved into our model. Particular, from our dataset, we will consider the effect of the age at diagnosis and clinical risk of diabetic retinopathy. More generally, we view the Cox proportional-hazards model a regression model which can take in covariates and be used to examine the hazard ratio and relative risk. Based on this, we can further add frailty to fit the data, which is used to account for the cluster effect we mentioned in the previous section.

Figure 2: Survival Distribution by Treatment (Overall)



After multiple comparison, our final model is determined to be semi-parametric frailty model with a gamma frailty. This model can also be viewed as a generalized proportional hazards regression model since we include frailty.

Here's how frailty is taking account into our data, basically by doing so, we can fit each cluster case by case. More specifically, under the frailty model, the hazard function at time t of observation j of cluster i is given by

$$\lambda_{ij}\left(t|X_{ij},\omega_i\right)=\omega_i\lambda_0(t)\exp(\beta^T X_{ij}),\quad j=1,2,\cdots,m_i,\quad i=1,2,\cdots,n$$

where ω_i is an unobservable frailty variate of cluster i, $\lambda_0(t)$ is the unknown common baseline hazard function, β is the unknown regression coefficient vector, and X_{ij} is the observed vector of covariates of observation j in cluster i. The frailty variates $\omega_1,\omega_2,\cdots,\omega_n$, are independent and identically distributed with known density $f(\bullet;\theta)$ and unknown parameter θ.

So by doing this, we mainly includes two parts into our model:

1. Based on naive survival analysis, we add factors of interest to our model.
2. We add frailty to describe the cluster in the data.

Model Results

Treatment Types and Effects

The regression formula for our model is shown below:

$$survobj\sim laser*trt+age+risk+frailty(id)$$

Now we can address the following question in a more statistical way:

What is the efficacy of treatment type on visual acuity?

From the Table 1, both treatment and id cluster are significant here in the model, which means that regardless of laser types, the laser treatment itself is effective at delaying diabetic retinopathy, and the cluster structure is also significant in the data. This conclusion is consistent with our previous judgment made from Figure 2. More specifically, the exponentiated coefficient of Treatment is 0.385, which means that treatment effect is about the same with a hazard ratio of 0.385(less than 1). More straightforwardly, treatment can have a significant reduction effect(38.5%) in time to loss of vision for the treated eyes compared to the control eyes.

Then for the treatment type, we see that the laser xenon is not statistically significant at $\alpha = 0.05$ level in the model, so laser xenon or laser argon may have similar performance in terms of delaying diabetic retinopathy, but they are all effective. Also, although from the p value we can not conclude that laser xenon is better than laser argon, the rate of Xenon(less than 1 vs. Argon) does indicate that there is a possibility that the Xenon can slightly outweigh the Argon in terms of treatment effect. This discovery is actually the same as what we see from Figure 2, as at most of the time the blue curve is above the green one. Moreover, the interaction between laser type and treatment eyes is not significant. Therefore, this indicates that we may need further investigation of the treatment effect between laser types.

Table 1: Rates and 95% confidence intervals of the risk of vision loss for the variables and factors assessed along with their p-values.

Variable	Rates(95% CI)	P Value
LaserXenon vs. LaserArgon	0.747(0.457, 1.223)	0.247
Treatment: treated eyes vs. controlled eyes	0.385(0.239, 0.620)	0.000
Age	1.007(0.992, 1.021)	0.358
Clinical Risk	1.185(1.035, 1.358)	0.014
Interaction:Xenon vs. Argon ~ Treatment	1.066(0.538, 2.112)	0.854
Frailty(id)	-	0.019

Adjusting for covariates

Now, in order to answer the second question, which ask for the potential impact that age at diagnosis and clinical risk of diabetic retinopathy have on visual acuity, we adjust our model for those two variable: *age* and *risk*.

Now let's focus on the covariate terms in the regression table. The age at diagnosis seems to have little impact because the rates is very close to 1, which means that there is no significant treatment effect difference among people of different ages. Meanwhile, the p value is much more larger than a 0.05 cutoff, so age may be irrelavent to the event of lossing vision. However, the clinical risk of diabetic retinopathy has an obvious impact on the event. Note that, the exponentiated coefficient of *risk* is 1.185 and the p value is less than cutoff 0.05, indicating that higher the risk, higher the probability of developing blindness under same condition, and more specifically, the probability is increased by 18.5%.

Conclusion

Our purpose of this analysis is to simply explain and quantify the efficacy of treatment type on visual acuity, and to determine the effect covariates like age at diagnosis and clinical risk have on loss of vision. While the model tells both Xenon and Argon are effective in delaying vision loss, one laser treatment type does not significantly outweigh the other, which means that the efficacy of those two laser type are roughly the same. However, from the coefficient of the model, we still can see that the laser Xenon still has a slightly better effect against laser argon, and this is also being shown on our figure. So we may need further investigation on this. Besides, the clinical risk of diabetic retinopathy plays an important role in disease developing, as people with higher clinical risk are more likely to develop such disease, and the probability is increased by 18.5%. Meanwhile, the age of diabetes diagnosis seems to be irrelevant to the whole process, which means there is no significant difference among people with different ages. Since each participant contributed to two observations in our data, we also introduce frailty term to the regression model to account for cluster

structure caused by eyes from the same person. The interaction effect between laser type and treatment is also trivial. Finally we see that the model characterize the survival process well.

References

[1] Monaco, Vinnie & Gorfine, Malka & Hsu, Li. (2018). General Semiparametric Shared Frailty Model Estimation and Simulation with frailtySurv. Journal of statistical software. 86. 10.18637/jss.v086.i04.

[2] Legrand, Catherine. (2021). Advanced Survival Models. 10.1201/9780429054167.

Appendix

The whole analysis was done in R programming language.

```
library(tidyverse)
library(survival)
path = "./"
retino_file = sprintf("%s/diabeticVision.csv", path)
retino = read_csv(file = retino_file) %>%
  select(-X1) %>%
  mutate(treatment = ifelse(trt == 0,
                            paste(laser, "control"),
                            paste(laser, "treated")
  ) )

survobj = with(retino, Surv (futime, status))

## KM plot wrt to time only:-------------------------------------------------
fit0 = survfit(survobj~1, data = retino)
# summary(fit0)
plot(fit0, xlab = "Time to loss of vision in Days",
     ylab = "% not blinded", yscale = 100,
     main = "Survival Distribution (Overall)")

## KM plot wrt to laser: -----------------------------------------------------
fit1 = survfit(survobj~treatment, data = retino)
plot(fit1, xlab = "Time to loss of vision in Days",
     ylab = "% not blinded", yscale = 100,
     main = "Figure 2: Survival Distribution by Treatment (Overall)",
     col = c('tomato3', 'green', 'cyan', 'blue'))
legend('bottomleft',
       legend = levels(as.factor(retino$treatment)),
       col = c('tomato3', 'green', 'cyan', 'blue'),
       lty = 1)

## ---------------------------------------------------------------------------
## -------------------------Comparing several models --------------------------
## ---------------------------------------------------------------------------
## We compare model from simple to complicate by adding terms to coxph model:--
## Model 1 Treatment only
trt_fit = coxph (survobj ~ trt, data = retino)

## Model 2 laser + treatment + age + clincal risk
indep_fit = coxph(survobj~laser + trt + age + risk, data = retino)

## Model 3 add cluster, but not consider factors
trt_fit = coxph(survobj ~ laser + trt + frailty (id), data = retino)
results = summary(trt_fit)

results$conf.int %>%
  as_tibble() %>%
  mutate(Variable = c("LaserXenon", "Treatment"),
         `Rates(95% CI)` = sprintf("%.3f(%.3f, %.3f)",
                                   `exp(coef)`, `lower .95`, `upper .95`)) %>%
  select(Variable, `Rates(95% CI)`) %>%
```

```
  add_row(Variable = "Frailty(id)", `Rates(95% CI)` = "-") %>%
  mutate(`P Value` = sprintf("%.3f", results$coefficients[,6]))

results$coefficients[,6]

## Model 4, consider the factors and the cluster
cluster_fit = coxph(survobj ~ laser*trt + age + risk + frailty (id),
                    data = retino)
results2 = summary(cluster_fit)
results2$conf.int %>%
  as_tibble() %>%
  mutate(Variable = c("LaserXenon vs. LaserArgon",
                      "Treatment: treated eyes vs. controlled eyes", "Age", "Clinical Risk",
                      "Interaction:Xenon vs. Argon ~ Treatment"),
         `Rates(95% CI)` = sprintf("%.3f(%.3f, %.3f)",
                                   `exp(coef)`, `lower .95`, `upper .95`)) %>%
  select(Variable, `Rates(95% CI)`) %>%
  add_row(Variable = "Frailty(id)", `Rates(95% CI)` = "-") %>%
  mutate(`P Value` = sprintf("%.3f", results2$coefficients[,6][c(1,2,3,4,6,5)])) %>%
  knitr::kable(caption = cap_tab1)
```