BIOST 537 Homework 4

Ivy Zhang

2/21/2022

Problem 1

(a)

```
library(survival)
library(flexsurv)
library(KMsurv)
library(readr)
meth <- read_csv("~/Desktop/R hw/methadone-1.csv") [,-1]
s.meth = with(meth, Surv(time, event))
#Change clinic and prison into dummy variable
meth$clinic = as.factor(meth$clinic)
meth$prison = as.factor(meth$prison)

aft.weib = flexsurvreg(s.meth ~ dose + clinic+prison, data = meth,dist = "weibull")
aft.weib
```

```
## Call:
## flexsurvreg(formula = s.meth ~ dose + clinic + prison, data = meth,
##     dist = "weibull")
##
## Estimates:
##          data mean  est        L95%       U95%       se         exp(est)
## shape           NA    1.37019    1.20047    1.56392    0.09245         NA
## scale           NA  123.20954   71.95892  210.96189   33.80725         NA
## dose      60.39916    0.02443    0.01546    0.03339    0.00458    1.02473
## clinic2    0.31513    0.70904    0.40089    1.01719    0.15722    2.03204
## prison1    0.46639   -0.22947   -0.46620    0.00727    0.12079    0.79496
##          L95%       U95%
## shape           NA         NA
## scale           NA         NA
## dose       1.01558    1.03396
## clinic2    1.49315    2.76543
## prison1    0.62738    1.00729
##
## N = 238,  Events: 150,  Censored: 88
## Total time at risk: 95812
## Log-likelihood = -1084.477, df = 5
## AIC = 2178.953
```

Based on the previous summary:

We estimated the average survival time of individuals with certain daily maintenance methadone dose is approximately 2.5% greater (meaning the time ratio equals to 1.025) than that of individuals with 1 mg/day less in daily maintenance methadone dose, but in same clinic and have same status of history of previous incarceration. The 95% confidence interval of the time ratio is [1.016, 1.034].

We estimated the average survival time of individuals who are in clinic 2 is approximately 103.2% greater (meaning the time ratio equals to 2.032) than that of individuals in clinic 1, but with same daily maintenance methadone dose and have same status of history of previous incarceration. The 95% confidence interval of the time ratio is [1.493, 2.765].

We estimated the average survival time of individuals who have history of previous incarceration is approximately 20.5% smaller (meaning the time ratio equals to 0.795) than that of individuals who don't have history of previous incarceration, but with same daily maintenance methadone dose and in same clinic. The 95% confidence interval of the time ratio is [0.627, 1.007].

Then we try to see whether this finding agree qualitatively with output from a Cox model:

```
cox.meth.1 = coxph(s.meth ~ dose + clinic+prison, data = meth)
summary(cox.meth.1)
```

```
## Call:
## coxph(formula = s.meth ~ dose + clinic + prison, data = meth)
##
##   n= 238, number of events= 150
##
##              coef exp(coef)  se(coef)      z Pr(>|z|)
## dose    -0.035369  0.965249  0.006379 -5.545 2.94e-08 ***
## clinic2 -1.009896  0.364257  0.214889 -4.700 2.61e-06 ***
## prison1  0.326555  1.386184  0.167225  1.953   0.0508 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##         exp(coef) exp(-coef) lower .95 upper .95
## dose       0.9652     1.0360    0.9533    0.9774
## clinic2    0.3643     2.7453    0.2391    0.5550
## prison1    1.3862     0.7214    0.9988    1.9238
##
## Concordance= 0.665  (se = 0.025 )
## Likelihood ratio test= 64.56  on 3 df,   p=6e-14
## Wald test            = 54.12  on 3 df,   p=1e-11
## Score (logrank) test = 56.32  on 3 df,   p=4e-12
```

If we compare this result with the previous summary we have from the AFT Weibull modek, we can find out that they are agree equalitatively. The cox proportional model is saying in people in same clinic and have same history status of previous incarceration, people who have higher dose of daily maintenance methadone dose have smaller hazard, meaning they have longer average time of survival, which is fitting what we have from the AFT Weibull model. The cox proportional model also is saying people having same dose of daily maintenance methadone and same history status of previous incarceration, people who are in clinic 2 have smaller hazard, meaning they have longer average time of survival, which is also fitting what we have from the AFT Weibull model. The cox proportional model also is implying for people who have same daily maintenance methadone dose and in the same clinic, people who have hisotry of previous incarceration will be estimated to have higher hazerd compared to people who have not, meaning they should have shorter average time of survival, which is also fitting the results from AFT Weibull model. Therefore, overall, I think findings of AFT Weibull model agree qualitively with out from a corresponding Cox model.

Then we will fit generalized gamma baseline distribution insted of Weibull distribution:

```
aft.gengamma = flexsurvreg(s.meth ~ dose +clinic+prison, data = meth,dist = "gengamma")
aft.gengamma
```

```
## Call:
## flexsurvreg(formula = s.meth ~ dose + clinic + prison, data = meth,
##     dist = "gengamma")
##
## Estimates:
##          data mean  est       L95%      U95%      se        exp(est)  L95%
## mu             NA    5.44901   4.84320   6.05481   0.30909        NA        NA
## sigma          NA    0.46428   0.28770   0.74926   0.11337        NA        NA
## Q              NA    2.09667   0.84671   3.34664   0.63775        NA        NA
## dose     60.39916    0.01771   0.00977   0.02565   0.00405   1.01787   1.00982
## clinic2   0.31513    0.70860   0.41063   1.00658   0.15203   2.03115   1.50776
## prison1   0.46639   -0.19145  -0.37218  -0.01073   0.09221   0.82576   0.68923
##          U95%
## mu             NA
## sigma          NA
## Q              NA
## dose       1.02598
## clinic2    2.73622
## prison1    0.98933
##
## N = 238,  Events: 150,  Censored: 88
## Total time at risk: 95812
## Log-likelihood = -1081.132, df = 6
## AIC = 2174.265
```

```
(exp(aft.gengamma$coefficients[c(4,5,6)]) -
    exp(aft.weib$coefficients[c(3,4,5)]))/
  exp(aft.weib$coefficients[c(3,4,5)]) * 100 #Calculated in percentage
```

```
##        dose     clinic2     prison1
## -0.66934026 -0.04404812  3.87456770
```

Based on the previous output, we estimated that the time ratio estimate change for the dose, clinic and prison parameter is -0.67%, -0.04% and 3.87%.

(b)

Suppose Z_1 corresponds to the participants methadone dosage. Z_2 corresponds to the clinic of the participant. $Z_2 = 0$ corresponds to the participant is in clinic 1 and $Z_2 = 1$ corresponds to the participant is in clinic 2. $Z_3 = 0$ corresponds to the participant does not have a history of incarceration and $Z_3 = 1$ corresponds to the participant had a history of incarceration. Then we want to estimate the ratio of mean time until exit from maintenance comparing individuals from clinic 2 without a history of incarceration administered dosage of 40 mg/day to individuals from clinic 1 with a history of incarceration administered a dosage of 100 mg/day.

Based on the Weibull AFT model from part (a), we are trying to estimate

$$\begin{aligned}\frac{E[T|Z_1 = 40, Z_2 = 1, Z_3 = 0]}{E[T|Z_1 = 100, Z_2 = 0, Z_3 = 1]} &= \frac{A(Z_1 = 40, Z_2 = 1, Z_3 = 0)^{-1} \times E[T|Z_1 = 0, Z_2 = 0, Z_3 = 0]}{A(Z_1 = 100, Z_2 = 0, Z_3 = 1)^{-1} \times E[T|Z_1 = 0, Z_2 = 0, Z_3 = 0]} \\ &= \frac{e^{\phi_1 \times 40 + \phi_2 \times 1 + \phi_3 \times 0}}{e^{\phi_1 \times 100 + \phi_2 \times 0 + \phi_3 \times 1}} \\ &= e^{\phi_1 \times (-60) + \phi_2 - \phi_3}\end{aligned}$$

Then we can used the previous fitted model estiamtion to calculated the ratio, also we can use the delta method to calculate the 95% confidence interval:

```
library(msm)
ratio.est = sum((-60)*aft.weib$coefficients[3],aft.weib$coefficients[4], (-1)*aft.weib$coefficients[5])
ratio.se = deltamethod(g = ~((-60)*x1+x2-x3),
                       mean = coef(aft.weib)[c(3,4,5)],
                     cov = vcov(aft.weib)[c(3,4,5),
                                          c(3,4,5)])
exp(ratio.est+c(-1,0,1)*1.96*ratio.se)
```

```
## [1] 0.3048388 0.5903651 1.1433289
```

Based on the previous calculation, we estimated that the ratio of mean time until exit from maintenance comparing individuals from clinic 2 without a history of incarceration administered dosage of 40 mg/day to individuals from clinic 1 with a history of incarceration administered a dosage of 100 mg/day is 0.590, and the 95% confidence interval of the ratio is [0.305, 1.143].

Then we try to estimate model-based estimate and 95% confidence interval of the median time until exit from maintenance for each of these two subpopulations of individuals.

For individuales from clinic 2 without a history of incarceration administered dosage of 40 mg/day, we have $Z_1 = 40$, $Z_2 = 1$ and $Z_3 = 0$, then we can use the chapter 2 slides to calculate the median time by also using the shape and scale parameter:

```
median.1 = exp(40*coef(aft.weib)[3]+coef(aft.weib)[4])*log(2)^(1/exp(coef(aft.weib)[1]))/
  (1/exp(coef(aft.weib)[2]))
median.se.1 = deltamethod(g=~log(2)^(1/exp(x1))/(1/exp(x2))*exp(40*x3+x4),
                        mean = coef(aft.weib)[c(1,2,3,4)],
                        cov = vcov(aft.weib)[c(1,2,3,4),
                                             c(1,2,3,4)])
median.1+c(-1,0,1)*1.96*median.se.1
```

```
## [1] 334.3971 509.0026 683.6080
```

Although from the class slides, we should not include exponential function into the delta method. However, it seems hard for us to distract exponential calculation from the median time caculation. Therefore, I decided to put the whold calculation into delta method calculation.

Based on the previous output, we conclude that we estimated the median time of individuals from clinic 2 without a history of incarceration administered dosage of 40 mg/day is 509.0026 and the 95% confidence interval of the estimated median time is [334.39, 683.61].

For individuales clinic 1 with a history of incarceration administered a dosage of 100 mg/day, we have $Z_1 = 100$, $Z_2 = 0$ and $Z_3 = 1$, then we can use the chapter 2 slides to calculate the median time by also using the shape and scale parameter:

```
median.2 = exp(100*coef(aft.weib)[3]+coef(aft.weib)[5])*log(2)^(1/exp(coef(aft.weib)[1]))/
  (1/exp(coef(aft.weib)[2]))
median.se.2 = deltamethod(g=~log(2)^(1/exp(x1))/(1/exp(x2))*exp(100*x3+x4),
                        mean = coef(aft.weib)[c(1,2,3,5)],
                        cov = vcov(aft.weib)[c(1,2,3,5),
                                             c(1,2,3,5)])
median.2+c(-1,0,1)*1.96*median.se.2
```

```
## [1]  510.9328  862.1826 1213.4324
```

Same as the previous calculation, I decided to put all calculation into delta method.

Based on the previous output, we conclude that we estimated the median time of individuals in clinic 1 with

a history of incarceration administered a dosage of 100 mg/day is 862.18 and the 95% confidence interval of the estimated median time is [510.93, 1213.43].

(c)

We first try to fit another Weibull AFT model to estimate the association between risk of exit from maintenance and methadone dosage adjusting for both clinic and history of previous incarceration as dummy variables and allowing this association to differ based on history of previous incarceration.

```
aft.weib2 = flexsurvreg(s.meth ~ dose*prison + clinic, data = meth,dist = "weibull")
aft.weib2
```

```
## Call:
## flexsurvreg(formula = s.meth ~ dose * prison + clinic, data = meth,
##     dist = "weibull")
##
## Estimates:
##                data mean  est        L95%       U95%       se         exp(est)
## shape                NA    1.37047    1.20081    1.56410    0.09241         NA
## scale                NA  107.27772   55.22968  208.37547   36.33939         NA
## dose            60.39916    0.02689    0.01545    0.03833    0.00584    1.02725
## prison1          0.46639    0.14467   -0.94812    1.23747    0.55756    1.15566
## clinic2          0.31513    0.69819    0.38873    1.00764    0.15789    2.01011
## dose:prison1    28.34034   -0.00638   -0.02452    0.01177    0.00926    0.99364
##                L95%       U95%
## shape                NA         NA
## scale                NA         NA
## dose            1.01557    1.03907
## prison1         0.38747    3.44688
## clinic2         1.47511    2.73914
## dose:prison1    0.97578    1.01184
##
## N = 238,  Events: 150,  Censored: 88
## Total time at risk: 95812
## Log-likelihood = -1084.24, df = 6
## AIC = 2180.481
```

Then we try to provide an estimate and 95% confidence interval for the time ratio comparing, among patients with a history of previous incarceration and from the same clinic, patients administered 80 mg/day to patients administered 60 mg/day.

Suppose Z_1 corresponds to the participants methadone dosage. Z_2 corresponds to the clinic of the participant. $Z_2 = 0$ corresponds to the participant is in clinic 1 and $Z_2 = 1$ corresponds to the participant is in clinic 2. $Z_3 = 0$ corresponds to the participant does not have a history of incarceration and $Z_3 = 1$ corresponds to the participant had a history of incarceration.

Then we are trying to estimate:

$$\begin{aligned}\frac{E[T|Z_1 = 80, Z_2 = z_2, Z_3 = 1]}{E[T|Z_1 = 60, Z_2 = 0, Z_3 = 1]} &= \frac{A(Z_1 = 80, Z_2 = z_2, Z_3 = 1)^{-1} \times E[T|Z_1 = 0, Z_2 = 0, Z_3 = 0]}{A(Z_1 = 60, Z_2 = z_2, Z_3 = 1)^{-1} \times E[T|Z_1 = 0, Z_2 = 0, Z_3 = 0]} \\ &= \frac{e^{\phi_1 \times 80 + \phi_2 \times z_2 + \phi_3 \times 1 + \phi_4 \times 80 \times 1}}{e^{\phi_1 \times 60 + \phi_2 \times z_2 + \phi_3 \times 1 + \phi_4 \times 60 \times 1}} \\ &= e^{\phi_1 \times 20 + \phi_4 \times 20}\end{aligned}$$

Then we can use the fitted model to estimate the previously mentioned ratios:

```
ratio.est2 = sum(20*aft.weib2$coefficients[3], 20*aft.weib2$coefficients[6])
ratio.se2 = deltamethod(g = ~20*x1+20*x2,
                        mean = coef(aft.weib2)[c(3,6)],
                      cov = vcov(aft.weib2)[c(3,6),
                                            c(3,6)])
exp(ratio.est2+c(-1,0,1)*1.96*ratio.se2)
```

```
## [1] 1.134531 1.507250 2.002418
```

Based on the previous output, we estimated the time ratio comparing, among patients with a history of previous incarceration and from the same clinic, patients administered 80 mg/day to patients administered 60 mg/day is 1.507 and the 95% confidence interval is [1.135, 2.002].

Then we are trying to estimate time ratio comparint, amont patients without a history of previous incarceration and from the same clinic, patients administered 80 mg/day to patients administered 60 mg/day.

In other words, we are trying to estimate:

$$\begin{aligned}\frac{E[T|Z_1=80,Z_2=z_2,Z_3=0]}{E[T|Z_1=60,Z_2=0,Z_3=0]} &= \frac{A(Z_1=80,Z_2=z_2,Z_3=0)^{-1}\times E[T|Z_1=0,Z_2=0,Z_3=0]}{A(Z_1=60,Z_2=z_2,Z_3=0)^{-1}\times E[T|Z_1=0,Z_2=0,Z_3=0]}\\ &= \frac{e^{\phi_1\times 80+\phi_2\times z_2+\phi_3\times 0+\phi_4\times 80\times 0}}{e^{\phi_1\times 60+\phi_2\times z_2+\phi_3\times 0+\phi_4\times 60\times 0}}\\ &= e^{\phi_1\times 20}\end{aligned}$$

Then we can use the fitted model to estimate the previously mentioned ratios:

```
ratio.est3 = sum(20*aft.weib2$coefficients[3])
ratio.se3 = deltamethod(g = ~20*x1,
                        mean = coef(aft.weib2)[c(3)],
                      cov = vcov(aft.weib2)[c(3),
                                            c(3)])
exp(ratio.est3+c(-1,0,1)*1.96*ratio.se3)
```

```
## [1] 1.362104 1.712240 2.152380
```

Based on the previous output, we estimated the time ratio comparing, among patients without a history of previous incarceration and from the same clinic, patients administered 80 mg/day to patients administered 60 mg/day is 1.712 and the 95% confidence interval is [1.362, 2.152].

To see whethether these two subgroup-specific time ratios significant different each other, meaning we are interested in whether ϕ_4 is significantly differs from 0. Since when $\phi_4 = 0$, we have

$$\frac{e^{\phi_1\times 20+\phi_4\times 20}}{e^{\phi_1\times 20}} = \frac{e^{\phi_1\times 20}}{e^{\phi_1\times 20}} = 1$$

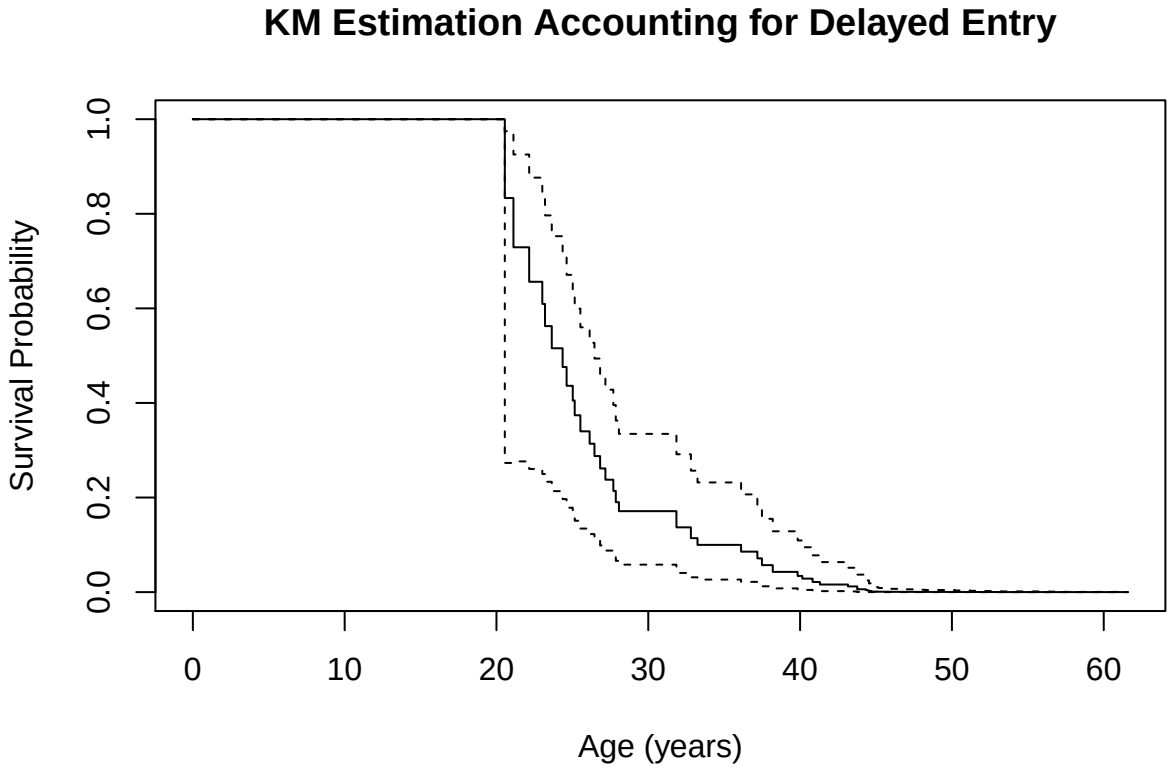
meaning these two subgroup-specific time ratios equal to each other. Based on the previous output, the 95% confidence interval of ϕ_4 is [-0.025, 0.011] which including 0, therefore we cannot reject the null hypothesis that the $\phi_4 = 0$ at the significance level of 0.05. In other words, these two subgroup-specific time ratios are not statistical significantly different each other at the significance level of 0.05.

Problem 2

(a)

We are trying to estimate nonparametrically the distribution of age at blindness due to diabetic retinopathy using data from untreated eyes only accounting for delayed entry at age at diagnosis. We will plot the KM curves with 95% confidence initervals to do that.

```
diabetes <- read_csv("~/Desktop/R hw/diabetes.csv")[,-1]
diabetes$followup = diabetes$age + diabetes$time / 12
untreated = diabetes[diabetes$treat == 0,]
surv.dia.untreated.delayed =Surv(untreated$age,untreated$followup, untreated$status)
km.dia.untreated.delayed = survfit(surv.dia.untreated.delayed~1, conf.type = "log-log")
plot(km.dia.untreated.delayed, conf.int = 0.95, xlab = "Age (years)", ylab = "Survival Probability",
     main = "KM Estimation Accounting for Delayed Entry")
```



Then we try to provide a point estimate and 95% confidence interval for the median age at blindness due to diabetic retinopathy based on the previous KM model.

```
km.dia.untreated.delayed
```

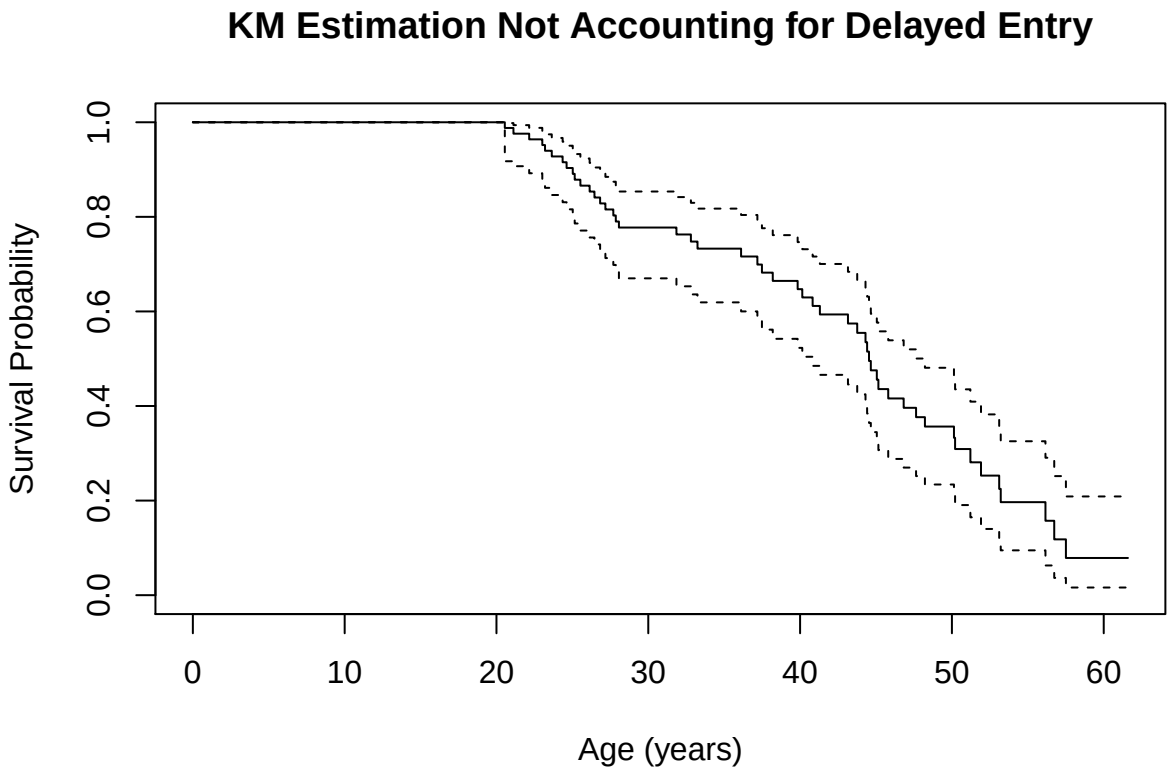
```
## Call: survfit(formula = surv.dia.untreated.delayed ~ 1, conf.type = "log-log")
##
## records   n.max n.start  events  median 0.95LCL 0.95UCL
##    83.0    14.0     6.0    50.0    24.4    20.5    26.5
```

Based on the previous output, we have a point estimate for the median age at blind due to diabetic retinopathy based on the KM model account for the delayed entry is 24.4 and the 95% confidence interval of the point estimate is [20.5, 26.5].

Then we try to estimate nonparametrically the distribution of age at blindness due to diabetic retinopathy

using data from untreated eyes only not accounting for delayed entry at age at diagnosis. We will plot the KM curves with 95% confidence initervals to do that.

```
surv.dia.untreated =Surv(untreated$followup, untreated$status)
km.dia.untreated = survfit(surv.dia.untreated~1, conf.type = "log-log")
plot(km.dia.untreated, conf.int = 0.95, xlab = "Age (years)", ylab = "Survival Probability",
     main = "KM Estimation Not Accounting for Delayed Entry")
```



Then we try to provide a point estimate and 95% confidence interval for the median age at blindness due to diabetic retinopathy based on the previous KM model.

```
km.dia.untreated
```

```
## Call: survfit(formula = surv.dia.untreated ~ 1, conf.type = "log-log")
##
##       n  events  median 0.95LCL 0.95UCL
##    83.0    50.0    44.5    40.8    48.2
```

Based on the previous output, we have a point estimate for the median age at blind due to diabetic retinopathy based on the KM model not account for the delayed entry is 44.5 and the 95% confidence interval of the point estimate is [40.8, 48.2].

(b)

I think the results from KM model account of delayed entry is not plausible but the results from KM model not account of delayed entry is plausible. Based on the KM model account of delayed entry, we are expecting 50% of the hign risk patients will become blind before they turn to 25 years old, and only 10% of the high risk patients are not blind when they reach to 40 years old (which is younger than the estimated median age of the KM model not account of the delayed entry). However, based on my google researching and the huge difference in the median age estimation between the previous two models, although the research study is focusing on the high-risk individuals, the results from KM model account of delayed entry seemed to be not

sensible compared to the KM model not account of delayed entry. Based on my opinion, the cause of concern of why the KM model account of the delayed entry is not sensible is because the independence assumption required for valid delayed entry adjustment is not plausible at this example. When we try to do the valid delayed entry adjustment, we want the age at recruitment into the study and age at the blindness due to diabetic retinopathy are independent in the target population (independent truncation). In other words, we want the individuals in the risk set at time t are representative of individuals in the target population who have survived until time t. However, in this study, the age at recruitment into the study is the age of diagnosis. I think the age of diagnosis is not independent of the age at the blindness due to diabetic retinopathy. Although the study only recruit people visual acuity of at least 20/100, people still varies in visual acuity. People who has worse visual acuity may have higher probability to find out they have diabetes and be dianosised by going to hospital researching why they have such symptoms, and also become more possible to become blind due to the diabetic retinopathy. Therefore, age of the diagnosis and age of the blindness due to the diabetic retinopathy are positively associated instead of being independent. Then the independence assumption required for valid delayed entry adjustment is not plausible in this study, leading to the KM model account of the delayed entry is not sensible.