Survival Analysis HW4

Faizan Khalid Mohsin

March 20, 2019

Contents

1 Question 1

Age could be added to the model also as a time-dependent covariate, to take into account that the patients are aging during the follow-up. Do this using the tt argument of the coxph function (please show your function call). What happened to the age effect estimate compared to the model with the fixed baseline age variable? Explain why.

First, we fit the time-independant age effect Cox model.

```
data(veteran)
veteran$prior = as.factor(veteran$prior/10)
veteran$trt = as.numeric(veteran$trt) - 1

model0 <- coxph(Surv(time, status) ~ trt + karno  + age  + prior + celltype + diagtime
                                  , data=veteran)
summary(model0)
```

```
## Call:
## coxph(formula = Surv(time, status) ~ trt + karno + age + prior +
##     celltype + diagtime, data = veteran)
##
##   n= 137, number of events= 128
##
##                          coef  exp(coef)   se(coef)      z Pr(>|z|)
## trt                 2.946e-01  1.343e+00  2.075e-01  1.419  0.15577
## karno              -3.282e-02  9.677e-01  5.508e-03 -5.958 2.55e-09 ***
## age                -8.706e-03  9.913e-01  9.300e-03 -0.936  0.34920
## prior1              7.159e-02  1.074e+00  2.323e-01  0.308  0.75794
## celltypesmallcell   8.616e-01  2.367e+00  2.753e-01  3.130  0.00175 **
## celltypeadeno       1.196e+00  3.307e+00  3.009e-01  3.975 7.05e-05 ***
## celltypelarge       4.013e-01  1.494e+00  2.827e-01  1.420  0.15574
## diagtime            8.132e-05  1.000e+00  9.136e-03  0.009  0.99290
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                    exp(coef) exp(-coef) lower .95 upper .95
## trt                   1.3426     0.7448    0.8939    2.0166
## karno                 0.9677     1.0334    0.9573    0.9782
## age                   0.9913     1.0087    0.9734    1.0096
## prior1                1.0742     0.9309    0.6813    1.6937
## celltypesmallcell     2.3669     0.4225    1.3799    4.0597
## celltypeadeno         3.3071     0.3024    1.8336    5.9647
## celltypelarge         1.4938     0.6695    0.8583    2.5996
## diagtime              1.0001     0.9999    0.9823    1.0182
##
## Concordance= 0.736  (se = 0.03 )
## Rsquare= 0.364   (max possible= 0.999 )
## Likelihood ratio test= 62.1  on 8 df,   p=1.799e-10
## Wald test            = 62.37  on 8 df,   p=1.596e-10
## Score (logrank) test = 66.74  on 8 df,   p=2.186e-11
```

We will now fit the time-dependant age effect Cox model.

```
model1 <- coxph(Surv(time, status) ~ trt + karno  + tt(age) + prior + celltype + diagtime
                                  , tt=function(x,t, ...) x + t/365.25, data=veteran)
summary(model1)
```

```
## Call:
## coxph(formula = Surv(time, status) ~ trt + karno + tt(age) +
##     prior + celltype + diagtime, data = veteran, tt = function(x,
##     t, ...) x + t/365.25)
##
##   n= 137, number of events= 128
##
##                          coef  exp(coef)   se(coef)      z Pr(>|z|)
## trt                 2.946e-01  1.343e+00  2.075e-01  1.419  0.15577
## karno              -3.282e-02  9.677e-01  5.508e-03 -5.958 2.55e-09 ***
## tt(age)            -8.706e-03  9.913e-01  9.300e-03 -0.936  0.34920
## prior1              7.159e-02  1.074e+00  2.323e-01  0.308  0.75794
## celltypesmallcell   8.616e-01  2.367e+00  2.753e-01  3.130  0.00175 **
## celltypeadeno       1.196e+00  3.307e+00  3.009e-01  3.975 7.05e-05 ***
## celltypelarge       4.013e-01  1.494e+00  2.827e-01  1.420  0.15574
## diagtime            8.132e-05  1.000e+00  9.136e-03  0.009  0.99290
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                    exp(coef) exp(-coef) lower .95 upper .95
## trt                   1.3426     0.7448    0.8939    2.0166
## karno                 0.9677     1.0334    0.9573    0.9782
## tt(age)               0.9913     1.0087    0.9734    1.0096
## prior1                1.0742     0.9309    0.6813    1.6937
## celltypesmallcell     2.3669     0.4225    1.3799    4.0597
## celltypeadeno         3.3071     0.3024    1.8336    5.9647
## celltypelarge         1.4938     0.6695    0.8583    2.5996
## diagtime              1.0001     0.9999    0.9823    1.0182
##
## Concordance= 0.736  (se = 0.234 )
## Rsquare= 0.364   (max possible= 0.999 )
## Likelihood ratio test= 62.1  on 8 df,   p=1.799e-10
## Wald test            = 62.37  on 8 df,   p=1.596e-10
## Score (logrank) test = 66.74  on 8 df,   p=2.186e-11
```

We can see that the estimate of the age coefficient is unchanged when we make it time-dependent. The estimate goes from $\hat{\beta_{age fixed}}$ = -0.0087065 (HZ = 0.9913313) to $\hat{\beta_{age-time-depend}}$ = -0.0087065 (HZ = 0.9913313). We also observe that the p-values of age in the two models are the same. Hence, the age effect estimate does not change when we make age time-dependant.

This can also be seen mathematically and we show this result in the attachment.

2 Question 2

2.1 Question 2 a

Going back to the model without age or other prognostic factors, one way to avoid assuming proportionality of the treatment effect over the entire follow-up period would be to estimate separate treatment effects for early and later parts of the follow-up period. Fit a Cox model that allows for separate treatment effects before and after 100 days of follow-up. Construct the required time-dependent covariates by using the tt argument of the coxph function (please show your coxph function call). Comment on the results compared to the Cox model that assumed a constant treatment effect. Do you see any problems in looking for different treatment effects over the follow-up period, especially if we chose the 100-day cutoff based on seeing the

non-parametric survival curves?

First, we fit the time-dependant treatment effect Cox model.

```
# Cox model that assumes a time dependant treatment effect

model3 <- coxph(Surv(time, status) ~ tt(trt)
                , tt=function(x,t, ...) cbind(I(t<=100)*x ,I(t>100)*x)
                , data=veteran)
summary(model3)
```

```
## Call:
## coxph(formula = Surv(time, status) ~ tt(trt), data = veteran,
##     tt = function(x, t, ...) cbind(I(t <= 100) * x, I(t > 100) *
##         x))
##
##   n= 137, number of events= 128
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## tt(trt)1  0.3988    1.4901   0.2277  1.751   0.0799 .
## tt(trt)2 -0.6799    0.5067   0.3163 -2.150   0.0316 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##          exp(coef) exp(-coef) lower .95 upper .95
## tt(trt)1    1.4901     0.6711    0.9536    2.3285
## tt(trt)2    0.5067     1.9737    0.2726    0.9418
##
## Concordance= 0.547  (se = 0.202 )
## Rsquare= 0.056   (max possible= 0.999 )
## Likelihood ratio test= 7.96  on 2 df,   p=0.01869
## Wald test            = 7.69  on 2 df,   p=0.02141
## Score (logrank) test = 7.89  on 2 df,   p=0.01938
```

We will now fit the time-independant treatment effect Cox model.

```
# Cox model that assumed a constant treatment effect

model4 <- coxph(Surv(time, status) ~ trt, data=veteran)
summary(model4)
```

```
## Call:
## coxph(formula = Surv(time, status) ~ trt, data = veteran)
##
##   n= 137, number of events= 128
##
##        coef exp(coef) se(coef)     z Pr(>|z|)
## trt 0.01774   1.01790  0.18066 0.098    0.922
##
##     exp(coef) exp(-coef) lower .95 upper .95
## trt     1.018     0.9824    0.7144      1.45
##
## Concordance= 0.525  (se = 0.026 )
## Rsquare= 0   (max possible= 0.999 )
## Likelihood ratio test= 0.01  on 1 df,   p=0.9218
## Wald test            = 0.01  on 1 df,   p=0.9218
## Score (logrank) test = 0.01  on 1 df,   p=0.9218
```

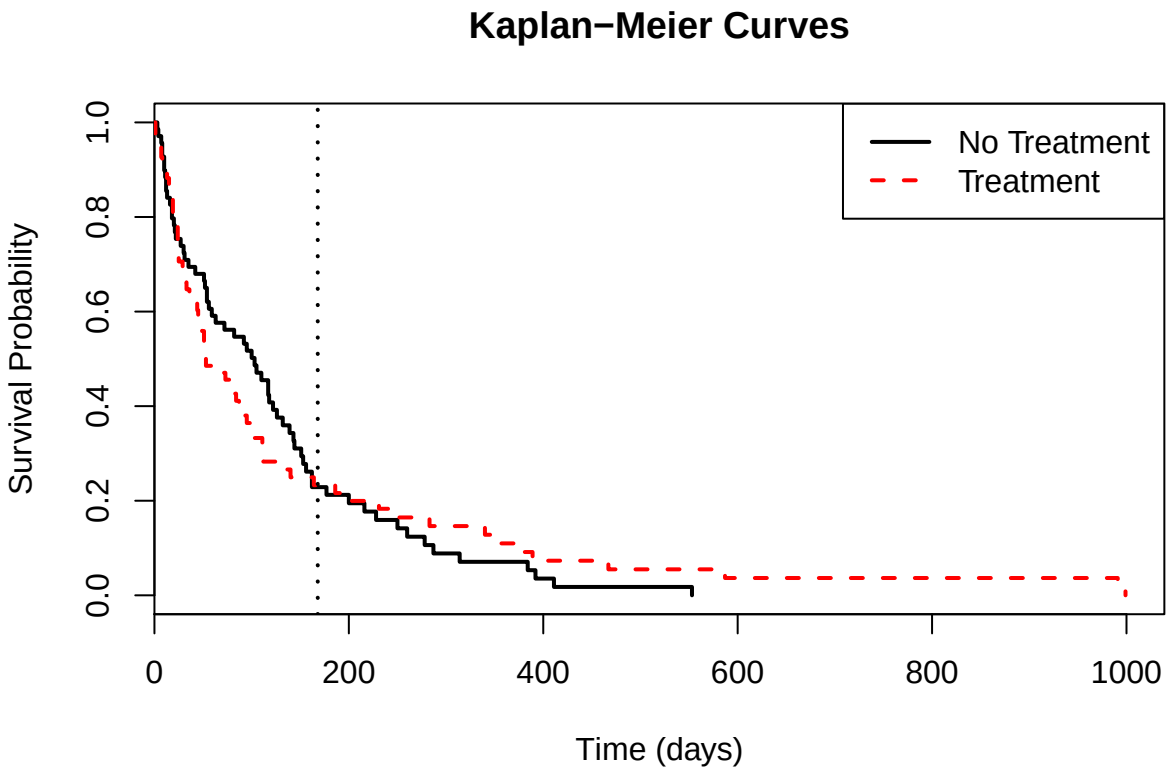
From the above output, for time-dependant treatment effect we see that the coefficient estimate of treatment effect changes from positive to negative when comparing before and after 100 days of follow-up.

In terms of comparing to the Cox model that assumed a constant treatment effect, it gives two very different treatment effects before and after 100 days of follow-up, whereas the constant treatment effect gives one "aggregate" effect estimate over all the follow-up period. Further, when we assume time-dependent treatment effect, the hazard ratios of the before and after 100 days (HR = 1.4901, p = 0.0799 and HR = 0.5067, p = 0.0316, respectively) are both more significant than the constant treatment effect hazard ratio (HR = 1.018, p = 0.922), with the hazard ratio for the follow-up period after the 100 days is statistically significant.

Further, when we divide the data based on cut-off times we will have multiple testing problem. Hence, our significance level should be $\alpha = 5\%/2 = 2.5\%$. In such a case, neither hazard ratio for before and after 100 days is statistically significant.

It is in fact difficult to choose the cut-off period when looking for if there are different treatment effects over the different follow-up time intervals. It would have perhaps been better to look at the survival curves of treatment and no-treatment, and based on that, chosen the cut-off point. Hence, we produce the survival curves below.

```
fit <- survfit(Surv(time, status) ~ trt , data=veteran)
plot(fit, ylab = "Survival Probability", xlab="Time (days)", lwd=2, lty=1:2, col=1:2 )
abline(v = 168, lty = "dotted", lwd = 2)
legend("topright", c("No Treatment", "Treatment"), lwd = 2,  col = 1:2, lty = 1:2)
title("Kaplan-Meier Curves")
```



The vertical dotted line is where the two curves intersect which is at about time 168 days. Hence, based on the K-M curves it would have been better to select the cut-off time to be 168 days instead of 100 days.

2.2 Question 2 b

Specify (algebraically) the hazard model you fitted in Q2(a)

Question done in the attachment.

3 Question 3

3.1 Question 3 i

From the transplant data, estimate (i) the probability of receiving transplant using the Kaplan-Meier method.

```
# Create time variables:

data(transplant)

# We will change the level withdraw to censored and create variable transplant$events_num
# that are coded 0, 1, 2. where 0 is censured, 1 is liver transplant, and 2 is death.
transplant$event = as.character(transplant$event)
transplant$events = ifelse(transplant$event == "withdraw", "censored", transplant$event)
transplant$events_num = ifelse(transplant$events == "censored", 0,
                               ifelse(transplant$events == "death", 2, 1))
table(transplant$events_num)
```

```
##
##   0   1   2
## 113 636  66
```

```
# We will change the follow-up time to years so it is in the same unit as age variable.

transplant$futime = transplant$futime / 365.25
str(transplant)
```

```
## 'data.frame':    815 obs. of  8 variables:
##  $ age       : num  47 55 52 40 70 66 41 55 50 61 ...
##  $ sex       : Factor w/ 2 levels "m","f": 1 1 1 2 1 2 2 1 1 1 ...
##  $ abo       : Factor w/ 4 levels "A","B","AB","O": 2 1 2 4 4 4 1 4 1 1 ...
##  $ year      : num  1994 1991 1996 1995 1996 ...
##  $ futime    : num  3.2772 0.0767 0.2327 0.6324 3.4798 ...
##  $ event     : chr  "death" "ltx" "ltx" "ltx" ...
##  $ events    : chr  "death" "ltx" "ltx" "ltx" ...
##  $ events_num: num  2 1 1 1 0 1 1 1 1 1 ...
```

```
# Age has missing data, with 18 people are missing age, which is a very
# small percentage of the data, hence, we will simply work with the complete cases.

transplant = transplant[complete.cases(transplant), ]

# Less than 80% actually received the transplant during the study:
sum(transplant$events_num==1)/nrow(transplant)
```

```
## [1] 0.7754078
```

```
# 8% actually died during the study:
sum(transplant$events_num==2)/nrow(transplant)
```

```
## [1] 0.08281054
```

From the above output it is important to note that 0.78 actually received the transplant during the stud and that 0.08 actually died during the study.

Below we plot the probability of receiving transplant over time in years.

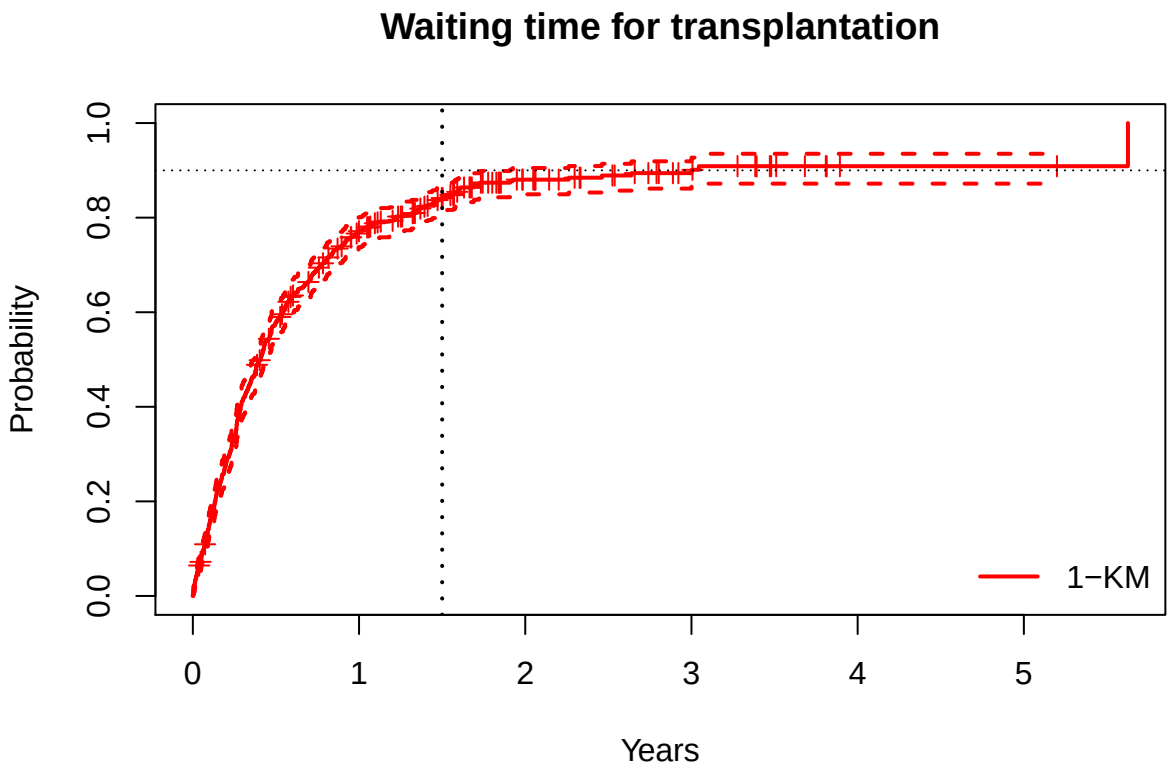
```
# Check that the event times are positive:

eps <- 0.00001
transplant$futime = ifelse(transplant$futime==0, eps, transplant$futime)

# K-M curve for probability of receiving transplantation:

fit <- survfit(Surv(futime,events_num==1)~1, data=transplant)

plot(fit, col=2, xlab="Years" #, ylim=c(0,1),xlim=c(0,1)
     , ylab="Probability", lwd=2, conf.int=T, mark.time=TRUE,
     main="Waiting time for transplantation", fun="event")
legend("bottomright", col=2, legend=c("1-KM"), lwd=2, bty='n')
abline(h = 0.90, lty = "dotted", lwd = 1)
abline( v = 1.5, lty = "dotted", lwd = 2)
```



From the above plot we can see that the longer we wait the higher the probability becomes for receiving a transplant. And the probability is dramatically increasing uptill 1.5 years (dotted vertical line) of getting a transplant and plateaus and remains at 90% (the dotted horizontal line) after 2 years.

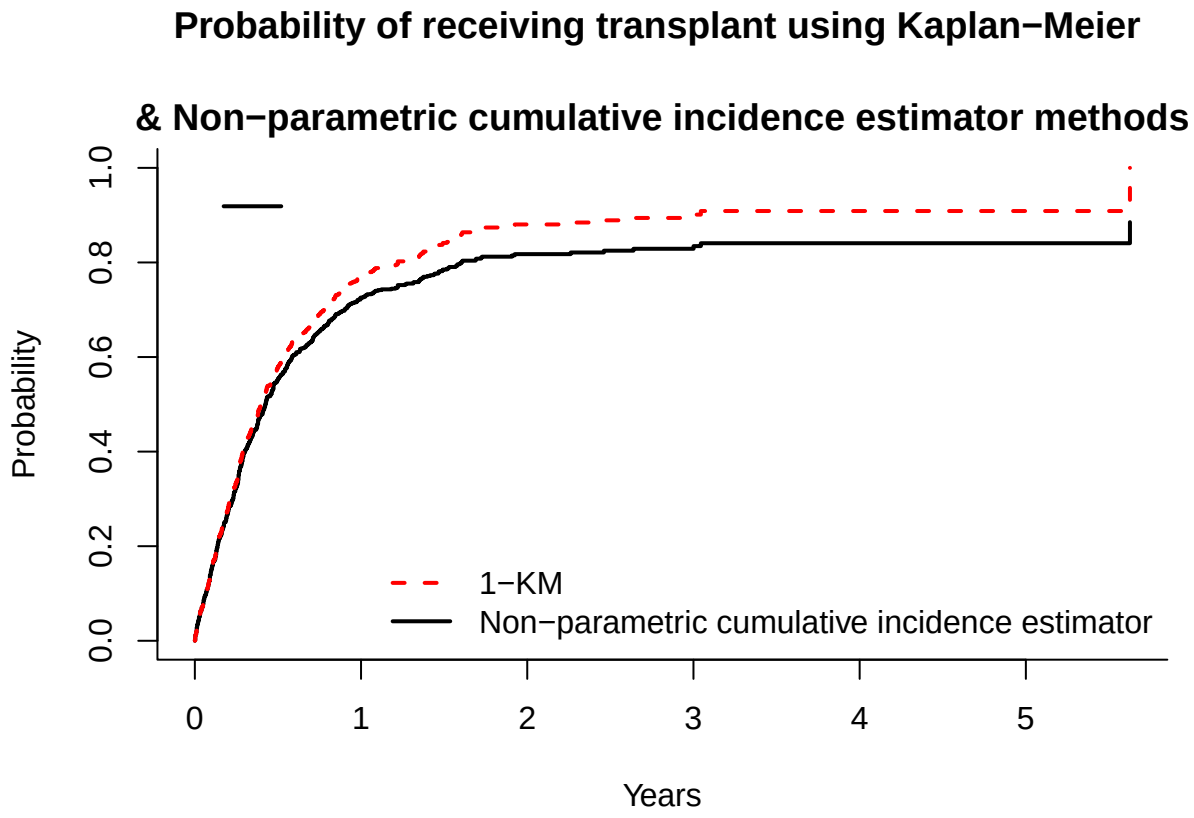
3.2 Question 3 ii

From the transplant data, estimate the cumulative incidence of receiving transplant using the non-parametric cumulative incidence estimator. Present the results as curves over time.

```
# Non-parametric cumulative incidence analysis:

library(cmprsk)
cifit <- cuminc(transplant$futime, transplant$events_num, cencode=0)

plot(cifit[1], curvlab = NA,  xlab="Years", lwd=2)
lines(fit, lwd=2, col=2, conf.int=F, fun="event", lty = 2)
legend("bottomright", col=2:1, legend=c("1-KM", 'Non-parametric cumulative incidence estimator')
       , lwd=2, bty='n', lty = c(2,1))
title("Probability of receiving transplant using Kaplan-Meier
      \n& Non-parametric cumulative incidence estimator methods")
```



3.3 Question 3 iii

Which method would you prefer and why?

We would choose the cumulative incidence method because it takes into account the competing risks (that there are more than one event of interest). The Kaplan Meier estimate does not take into account the competing risk. If there was only one event of interest (death or transplant) then one minus the Kaplan-Meier curve would give us the cumulative incidence curve. However, since we have two events of interest (death and transplant) then one minus the Kaplan Meier curve does not give us exactly the cumulative incidence curve.

Hence, it is better to actually calculate the cumulative incidence curves directly using the above method.

From the above plot it can be seen that the Kaplan-Meier method over estimates the probability of receiving a transplant compared to the non-parametric cumulative incidence estimator.

4 Question 4

Fit appropriate Cox and Fine & Gray models adjusted for age, sex, ABO blood group and year, and use these to calculate the cumulative incidence curves for receiving transplant and death at the average covariate values.

```
# Cox models for transplantation and death before transplantation:
transplant$year <- transplant$year - mean(transplant$year)

model1 <- coxph(Surv(transplant$futime,transplant$events_num==1)~age + sex + abo + year,
                data=transplant)
summary(model1)
```

```
## Call:
## coxph(formula = Surv(transplant$futime, transplant$events_num ==
##     1) ~ age + sex + abo + year, data = transplant)
##
##   n= 797, number of events= 618
##
##            coef exp(coef)  se(coef)       z Pr(>|z|)
## age   -0.003038  0.996967  0.004029  -0.754   0.4509
## sexf  -0.059804  0.941950  0.081380  -0.735   0.4624
## aboB  -0.319808  0.726288  0.131239  -2.437   0.0148 *
## aboAB  0.088977  1.093056  0.187909   0.474   0.6358
## aboO  -0.586539  0.556249  0.090500  -6.481 9.11e-11 ***
## year  -0.304127  0.737767  0.016314 -18.642  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##       exp(coef) exp(-coef) lower .95 upper .95
## age      0.9970     1.0030    0.9891    1.0049
## sexf     0.9419     1.0616    0.8031    1.1048
## aboB     0.7263     1.3769    0.5616    0.9393
## aboAB    1.0931     0.9149    0.7563    1.5798
## aboO     0.5562     1.7978    0.4658    0.6642
## year     0.7378     1.3554    0.7145    0.7617
##
## Concordance= 0.755  (se = 0.013 )
## Rsquare= 0.368   (max possible= 1 )
## Likelihood ratio test= 365.2  on 6 df,   p=0
## Wald test            = 419.7  on 6 df,   p=0
## Score (logrank) test = 431  on 6 df,   p=0
```

```
model2 <- coxph(Surv(transplant$futime,transplant$events_num==2)~age + sex + abo + year,
                data=transplant)
summary(model2)
```

```
## Call:
## coxph(formula = Surv(transplant$futime, transplant$events_num ==
```

```
##     2) ~ age + sex + abo + year, data = transplant)
##
##   n= 797, number of events= 66
##
##           coef exp(coef) se(coef)      z Pr(>|z|)
## age    0.01987   1.02007  0.01294  1.536 0.124595
## sexf  -0.45355   0.63537  0.25778 -1.759 0.078498 .
## aboB   0.16608   1.18066  0.38716  0.429 0.667954
## aboAB  0.24458   1.27708  0.61886  0.395 0.692691
## aboO   0.01123   1.01130  0.28617  0.039 0.968691
## year  -0.18311   0.83268  0.05246 -3.490 0.000482 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##       exp(coef) exp(-coef) lower .95 upper .95
## age      1.0201     0.9803    0.9945    1.0463
## sexf     0.6354     1.5739    0.3834    1.0530
## aboB     1.1807     0.8470    0.5528    2.5216
## aboAB    1.2771     0.7830    0.3797    4.2953
## aboO     1.0113     0.9888    0.5772    1.7720
## year     0.8327     1.2009    0.7513    0.9229
##
## Concordance= 0.647  (se = 0.04 )
## Rsquare= 0.02   (max possible= 0.623 )
## Likelihood ratio test= 15.79  on 6 df,   p=0.01494
## Wald test            = 16.48  on 6 df,   p=0.01139
## Score (logrank) test = 16.74  on 6 df,   p=0.01029
```

```
# Calculate the cumulative incidence of receiving transplant at average covariate values:

L1 <- basehaz(model1, centered=FALSE)
L1$hazard <- L1$hazard * as.numeric(exp(t(colMeans(model.matrix(model1))) %*% coef(model1)))
names(L1) <- c('ch1','time')
L2 <- basehaz(model2, centered=FALSE)
L2$hazard <- L2$hazard * as.numeric(exp(t(colMeans(model.matrix(model2))) %*% coef(model2)))
names(L2) <- c('ch2','time')
intersect(names(L1),names(L2))
```

```
## [1] "time"
```

```
ci <- merge(L1, L2)

# Cumulative hazard increments:

ci$h1 <- c(ci$ch1[1],ci$ch1[2:nrow(ci)]-ci$ch1[1:(nrow(ci)-1)])
ci$h2 <- c(ci$ch2[1],ci$ch2[2:nrow(ci)]-ci$ch2[1:(nrow(ci)-1)])
#head(ci)
#tail(ci)

# Overall survival probability (two alternative estimators):

ci$surv <- exp(-(c(0.0,ci$ch1[1:(nrow(ci)-1)]) + c(0.0,ci$ch2[1:(nrow(ci)-1)])))
ci$kmsurv <- cumprod(1.0 - (c(0.0,ci$h1[1:(nrow(ci)-1)]) + c(0.0,ci$h2[1:(nrow(ci)-1)])))
#head(ci)
#tail(ci)
```

```r
# plot(ci$time, ci$surv, type='s')
# lines(ci$time, ci$kmsurv, type='s', col='red')

# Calculate cumulative incidence:

ci$ci1 <- cumsum(ci$h1 * ci$kmsurv)
ci$ci2 <- cumsum(ci$h2 * ci$kmsurv)
#head(ci)
#tail(ci)

# Plot:

# plot(ci$time, ci$ci1, type='s', lwd=2, lty=c('solid'), #ylim=c(0,1), xlim=c(0,1),
#      xlab="Years since entering the waiting list",
#     ylab="Cumulative incidence", main='Cumulative incidence functions')
# lines(ci$time, ci$ci2, type='s', lwd=2, lty=c('dashed'))
# legend('right', legend=c('Transplant','Death'), lwd=2, lty=c('solid','dashed'), bty='n')

# Fine & Gray models:

mm <- cbind(transplant$age, transplant$sex, transplant$abo, transplant$year)
colnames(mm) <- c('age','sex','abo', 'year')

fgmodel1 <- crr(transplant$futime, transplant$events_num, cov1=mm, failcode=1, cencode=0)
summary(fgmodel1)
```

```
## Competing Risks Regression
##
## Call:
## crr(ftime = transplant$futime, fstatus = transplant$events_num,
##     cov1 = mm, failcode = 1, cencode = 0)
##
##          coef exp(coef) se(coef)      z p-value
## age  -0.00636     0.994  0.00454 -1.403 1.6e-01
## sex   0.09175     1.096  0.09509  0.965 3.3e-01
## abo  -0.16281     0.850  0.03416 -4.766 1.9e-06
## year -0.23315     0.792  0.02398 -9.721 0.0e+00
##
##      exp(coef) exp(-coef)  2.5% 97.5%
## age      0.994      1.006 0.985 1.003
## sex      1.096      0.912 0.910 1.321
## abo      0.850      1.177 0.795 0.909
## year     0.792      1.263 0.756 0.830
##
## Num. cases = 797
## Pseudo Log-likelihood = -3596
## Pseudo likelihood ratio test = 241  on 4 df,
```

```r
fgmodel2 <- crr(transplant$futime, transplant$events_num, cov1=mm, failcode=2, cencode=0)
summary(fgmodel2)
```

```
## Competing Risks Regression
##
## Call:
```

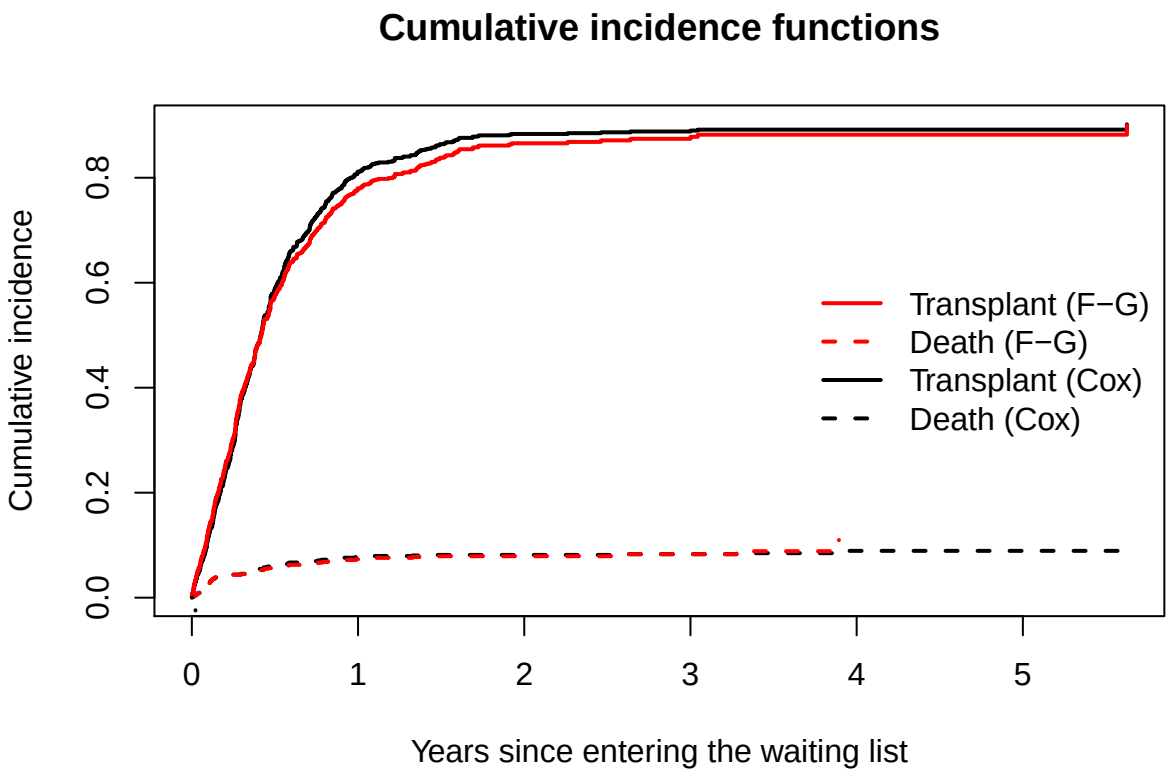
```
## crr(ftime = transplant$futime, fstatus = transplant$events_num,
##     cov1 = mm, failcode = 2, cencode = 0)
##
##          coef exp(coef) se(coef)      z p-value
## age   0.01944     1.020   0.0114  1.699   0.089
## sex  -0.41564     0.660   0.2610 -1.592   0.110
## abo   0.11498     1.122   0.0889  1.294   0.200
## year -0.00838     0.992   0.0472 -0.178   0.860
##
##      exp(coef) exp(-coef)  2.5% 97.5%
## age      1.020      0.981 0.997  1.04
## sex      0.660      1.515 0.396  1.10
## abo      1.122      0.891 0.942  1.34
## year     0.992      1.008 0.904  1.09
##
## Num. cases = 797
## Pseudo Log-likelihood = -427
## Pseudo likelihood ratio test = 6.4  on 4 df,
```

```
# Cumulative incidence at average covariate values:

fgci1 <- predict(fgmodel1, cov1=t(colMeans(mm)))
fgci2 <- predict(fgmodel2, cov1=t(colMeans(mm)))
```

Below are the cumulative incidence curves for receiving transplant and death at the average covariate values.

```
plot(ci$time, ci$ci1, type='s', lwd=2, lty=c('solid'), # ylim=c(0,1), xlim=c(0,1),
     xlab="Years since entering the waiting list", ylab="Cumulative incidence",
     main='Cumulative incidence functions')
lines(ci$time, ci$ci2, type='s', lwd=2, lty=c('dashed'))
lines(fgci1[,1], fgci1[,2], type='s', lwd=2, lty=c('solid'), col='red')
lines(fgci2[,1], fgci2[,2], type='s', lwd=2, lty=c('dashed'), col='red')
lines(ci$time, ci$ci2 - ci$ci1, type='s', lwd=2, lty=c('dotted'))
#abline(a = 0.8, b = 0, type='s', lwd=1, lty=c('dotted'))
legend('right', legend=c('Transplant (F-G)','Death (F-G)','Transplant (Cox)','Death (Cox)'), lwd=2,
       lty=c('solid','dashed','solid','dashed'), col=c('red','red','black','black'), bty='n')
```



How do the results compare between the models?

From the above plot we see that for transplant, the Fine and Gray model predicts higher cumulative incidence probability compared to the Cox model at the average covariate values.

For death, they are the same as the curves overlap one another.

5 Question 5

Use the fitted Cox and Fine & Gray models adjusted to calculate individual-level oneyear cumulative incidences of receiving the transplant. Present these in a scatterplot.

```
# Comparison of predictions: calculate 1-year cumulative incidences of receiving transplant
# for everyone in the dataset:

ciall <- rep(NA, nrow(transplant))
fgciall <- rep(NA, nrow(transplant))
s <- 1 # 1-year cumulative incidence

for (i in 1:nrow(transplant)) {
    L1 <- basehaz(model1, centered=FALSE)
    L1$hazard <- L1$hazard * as.numeric(exp(t(model.matrix(model1)[i,]) %*% coef(model1)))
    names(L1) <- c('ch1','time')
    L2 <- basehaz(model2, centered=FALSE)
    L2$hazard <- L2$hazard * as.numeric(exp(t(model.matrix(model2)[i,]) %*% coef(model2)))
    names(L2) <- c('ch2','time')
```

```
    ci <- merge(L1, L2)

    ci$h1 <- c(ci$ch1[1],ci$ch1[2:nrow(ci)]-ci$ch1[1:(nrow(ci)-1)])
    ci$h2 <- c(ci$ch2[1],ci$ch2[2:nrow(ci)]-ci$ch2[1:(nrow(ci)-1)])

    ci$surv <- exp(-(c(0.0,ci$ch1[1:(nrow(ci)-1)]) + c(0.0,ci$ch2[1:(nrow(ci)-1)])))
    ci$kmsurv <- cumprod(1.0 - (c(0.0,ci$h1[1:(nrow(ci)-1)]) + c(0.0,ci$h2[1:(nrow(ci)-1)])))

    ci$ci1 <- cumsum(ci$h1 * ci$kmsurv)
    ci$ci2 <- cumsum(ci$h2 * ci$kmsurv)
    ciall[i] <- ci$ci1[findInterval(s, ci$time)]
    if (i %% 100 == 0)
        print(i)
}
```

```
## [1] 100
## [1] 200
## [1] 300
## [1] 400
## [1] 500
## [1] 600
## [1] 700
```

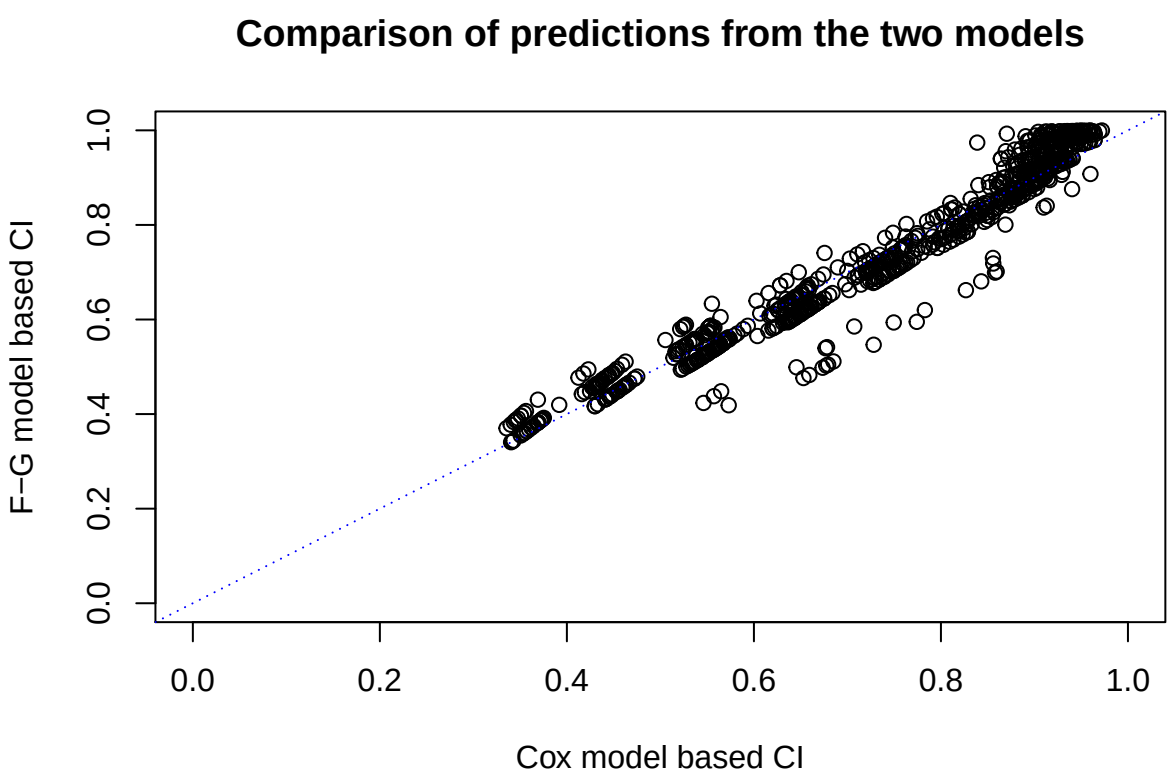
```
#calculate individual-level one-year cumulative incidences of receiving the transplant
sum(ciall)
```

```
## [1] 589.4142
```

```
# Fine Gray #calculate individual-level one-year cumulative incidences of receiving the transplant
fgci <- predict(fgmodel1, cov1=mm)
fgciall <- fgci[findInterval(s, fgci[,1]),2:ncol(fgci)]
sum(fgciall)
```

```
## [1] 590.848
```

```
plot(ciall, fgciall, xlab='Cox model based CI', ylab='F-G model based CI',
     main='Comparison of predictions from the two models',
     xlim = c(0,1), ylim = c(0,1))
abline(a=0, b=1, lty='dotted', col='blue')
```



The points that are above the blue line are the points which have a higher probability predicted by the Fine and Gray model compared to the Cox model. And the points that are below the blue line have higher probabilities predicted by the Cox model compared to the Fine and Gray Model.

How do the results compare between the models?

Firstly, there is a clear positively correlated linear trend. Further, we can see a pattern that near the edge of the graph (between probabilities 0.4-0.5 and 0.9-1) for the majority of the observations the Fine & Gray model over predicts the probability compared to the Cox model and that between probabilities 0.5-0.9 the Cox model over predicts higher probabilites compared to the Fine & Gray model.

Furuther, we find that for the Cox model the individual-level one year cumulative incidences of receiving the transplant is 589.4141597.

We find that for the Fine & Gray model the individual-level one year cumulative incidences of receiving the transplant is 590.8479813.

The two estimates are approximately the same with a difference of only 1.4338216.