PSTAT 175 Final Project

Yuxi Sun

12/18/2020

Abstract

In this project, we use the K-M estimation, cox ph model, and time dependent cox ph model to analysis the nwtco from the survival package. Our question is: is age, histology, and stage affect the survival rate of the patients? Based on our analysis: we conclude that the stage 2, 3, 4 patients have a higher hazard rate, and the baseline hazard function of UM histology is significant higher than the FM group. The age conclusion is tricky: though in histology FM group, age has a positive effect on the relapse (which is coincidence to the common sense), but in the UM group in the time 0-160, the age has a negative effect on the relapse, which is opposite to the common sense. Also we observe when time move over 160, the age effect in the UM group disappear.

Data source and research question

The data set was from survival package called nwtco, this data set is collected from the National Wilm's Tumor Study. It uses Tumor histology to predict survival rate, but prediction is stronger with central lab histology than with the local institution determination.

The variables include:

seqno: id number

instit: Histology from local institution or centre lab (1 or 2)

histol: Patients whose tumours are composed of one of the rare cell types knowncollectively as 'unfavourable histology' (UH) are much more likely to relapse and die thanare patients with tumours of 'favourable histology' (FH) (Beckwith and Palmer, 1978)

stage: Disease stage (1-4)

study: study (3 or 4) from the third and fourth clinical trials of the National Wilms Tumor Study Group (NWTSG) (D'Angio et al., 1989;Green et al., 1998)

rel: indicator for relapse (0 or 1)

edrel: time to relapse (day)

age: age in month

in.subcohort: Included in the subcohort for the example in the paper (T or F)

This data was first included in the paper:

NE Breslow and N Chatterjee (1999), Design and analysis of two-phase studies with binary outcome applied to Wilms tumour prognosis. Applied Statistics 48, 457–68.

In our analysis, we treat stage, age and histol as 3 important covariates for the survival probability. edrel is the response time. Instit 1 stands for local institution and 2 stands for center lab. The data from R is not well informed so we first need to prepare the data:

```
##   seqno instit histol stage study rel edrel      age in.subcohort
## 1     1 center     UH     I     3   0  6075 2.083333        FALSE
## 2     2  local     FH    II     3   0  4121 4.166667        FALSE
## 3     3 center     UH     I     3   0  6069 0.750000        FALSE
## 4     4 center     FH    IV     3   0  6200 2.333333         TRUE
## 5     5 center     UH    II     3   0  1244 4.583333        FALSE
## 6     6  local     FH    II     3   0  2932 2.666667        FALSE
```

```
##      seqno          instit     histol    stage         study
##  Min.   :   1   local :3622   FH:3569   I  :1572   Min.   :3.000
##  1st Qu.:1009   center: 406   UH: 459   II :1052   1st Qu.:3.000
##  Median :2022                           III: 944   Median :4.000
##  Mean   :2026                           IV : 460   Mean   :3.539
##  3rd Qu.:3039                                      3rd Qu.:4.000
##  Max.   :4088                                      Max.   :4.000
##       rel             edrel             age          in.subcohort
##  Min.   :0.0000   Min.   :   4.0   Min.   : 0.000   Mode :logical
##  1st Qu.:0.0000   1st Qu.: 856.8   1st Qu.: 1.583   FALSE:3360
##  Median :0.0000   Median :1939.0   Median : 3.083   TRUE :668
##  Mean   :0.1418   Mean   :2276.7   Mean   : 3.553
##  3rd Qu.:0.0000   3rd Qu.:3561.0   3rd Qu.: 4.833
##  Max.   :1.0000   Max.   :6209.0   Max.   :15.917
```

The paper has told us ‘unfavourable histology’ (UH) are much more likely to relapse and die thanare patients with tumours of ‘favourable histology’ (FH). And by common sense, the larger the stage is, the more likely to relapse and die. Also we would guess the age will have a positive influence on the relapse probability.

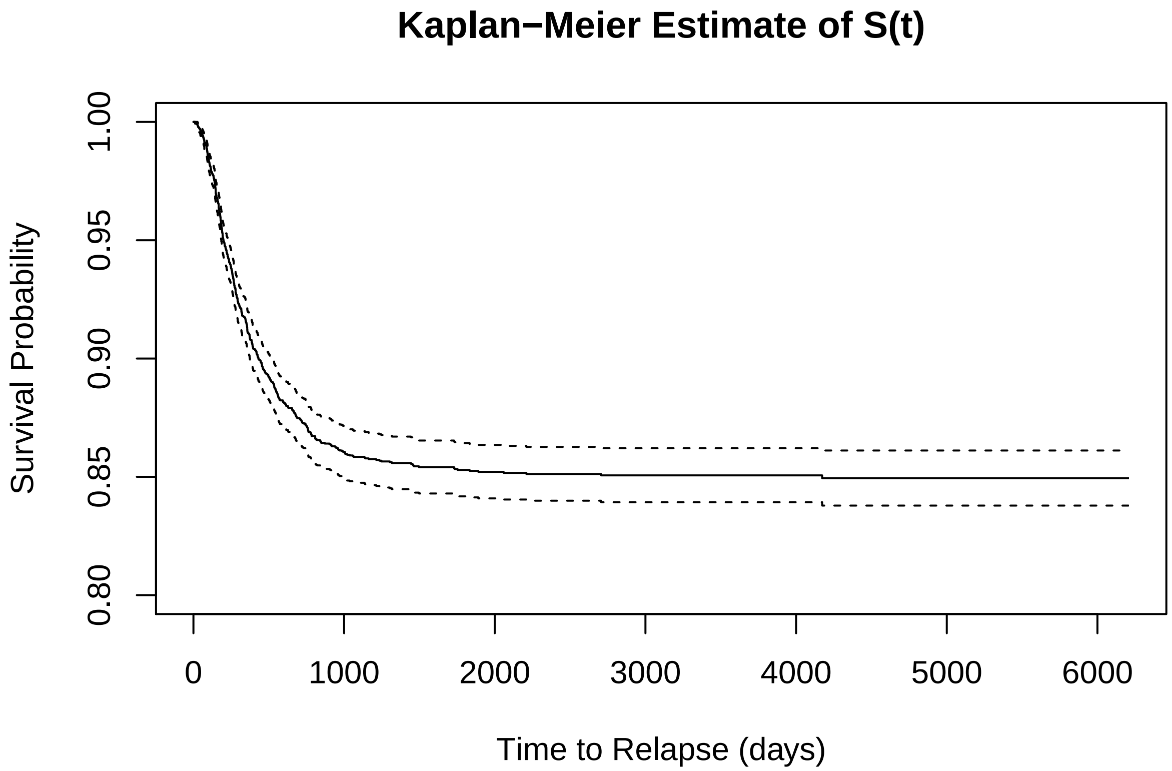
Kaplan-Meier estimation

To visualize our data, we use the original Kaplan-Meier estimator. It is a niave but useful approach to get the first hand impression of our dataset. For different category variables, we could calculate their K-M estimator separately to compare the survival function between different groups. The confidence interval is derived from the greenwood formula:

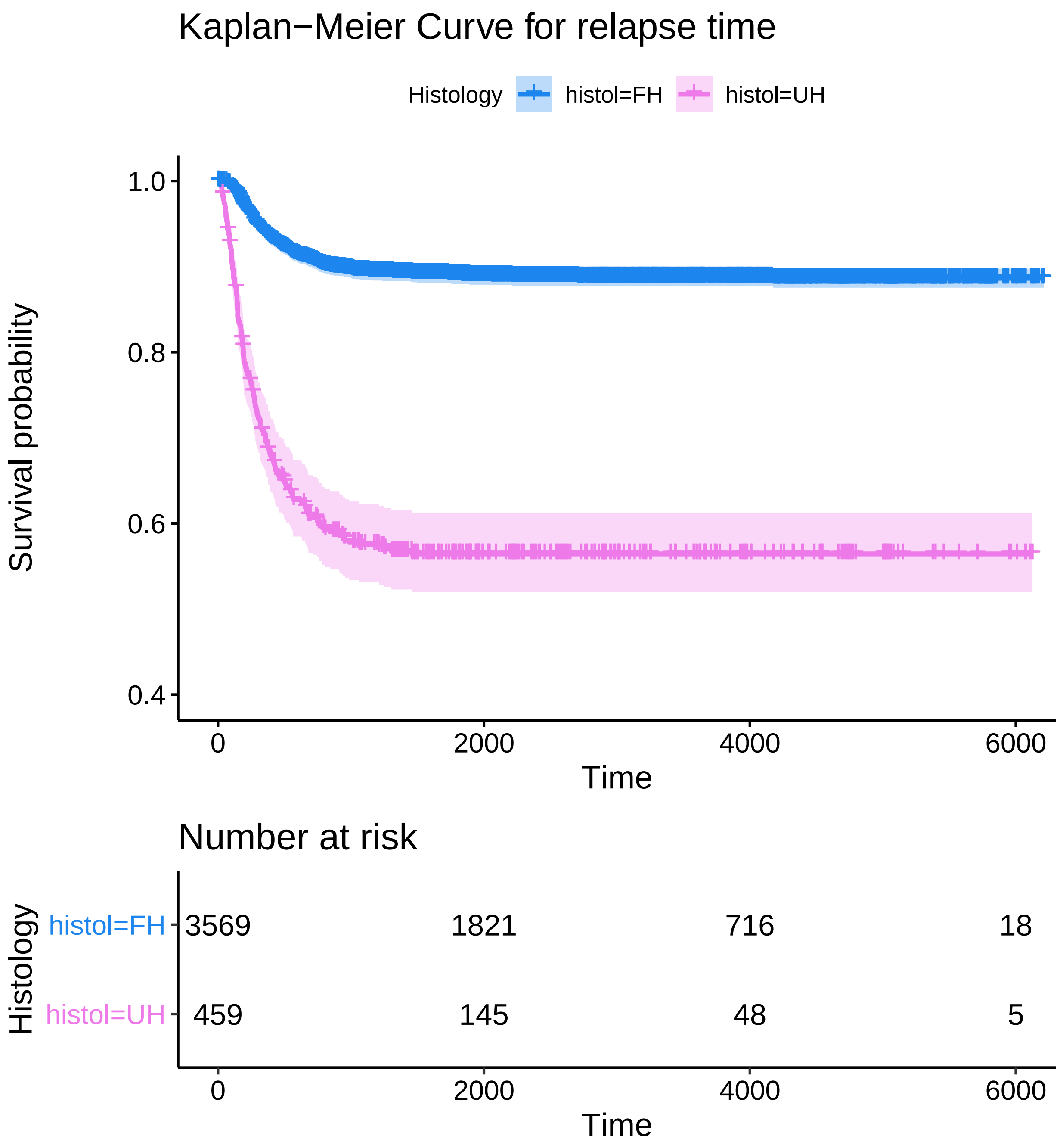
$$Var(\hat{S}_t) = \hat{S}_t^2 \sum_{t_i \leq t} \frac{d_i}{n_i(n_i - d_i)}$$

For each of the group, we will perform a log-rank test to get the p-value to test whether the two groups are different in KM-estimator.

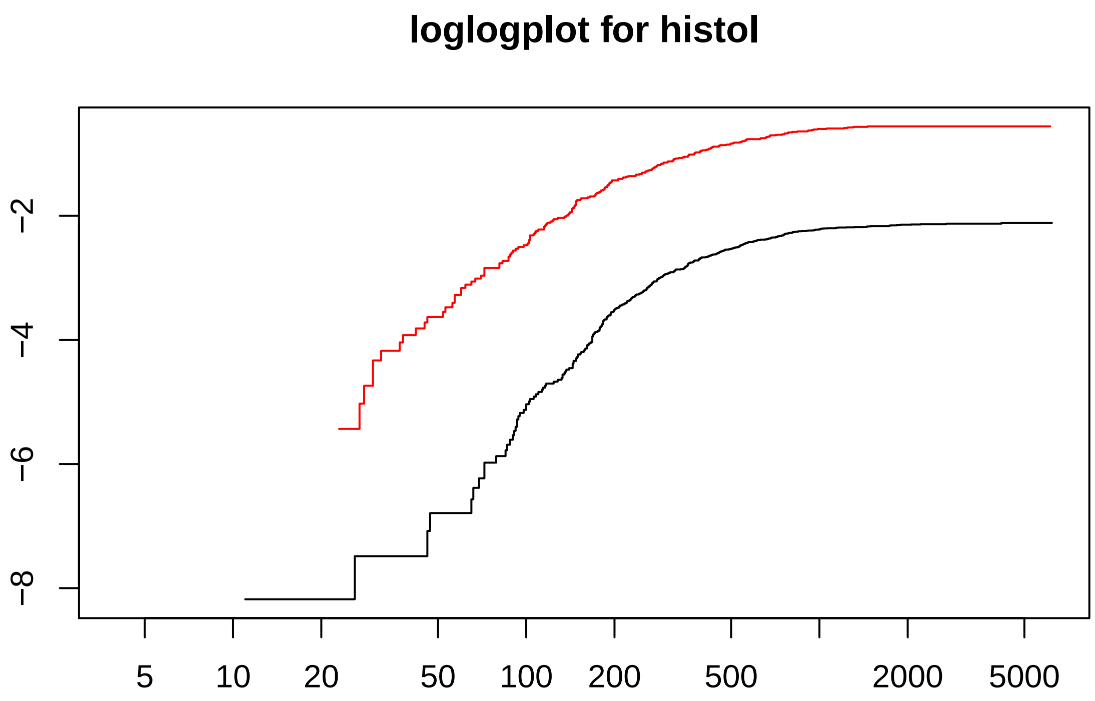
- K-M estimator and 95% confidence interval from greenwood formula. Total data:



- Data partitioned by histol



From Kaplan-Meier estimator we could see that there is a huge difference the histology group FH and UH. It's very obvious that the histology UH has a larger fatality rate. Then we use the log log plot to check the proportional hazard assumption used in Cox model:

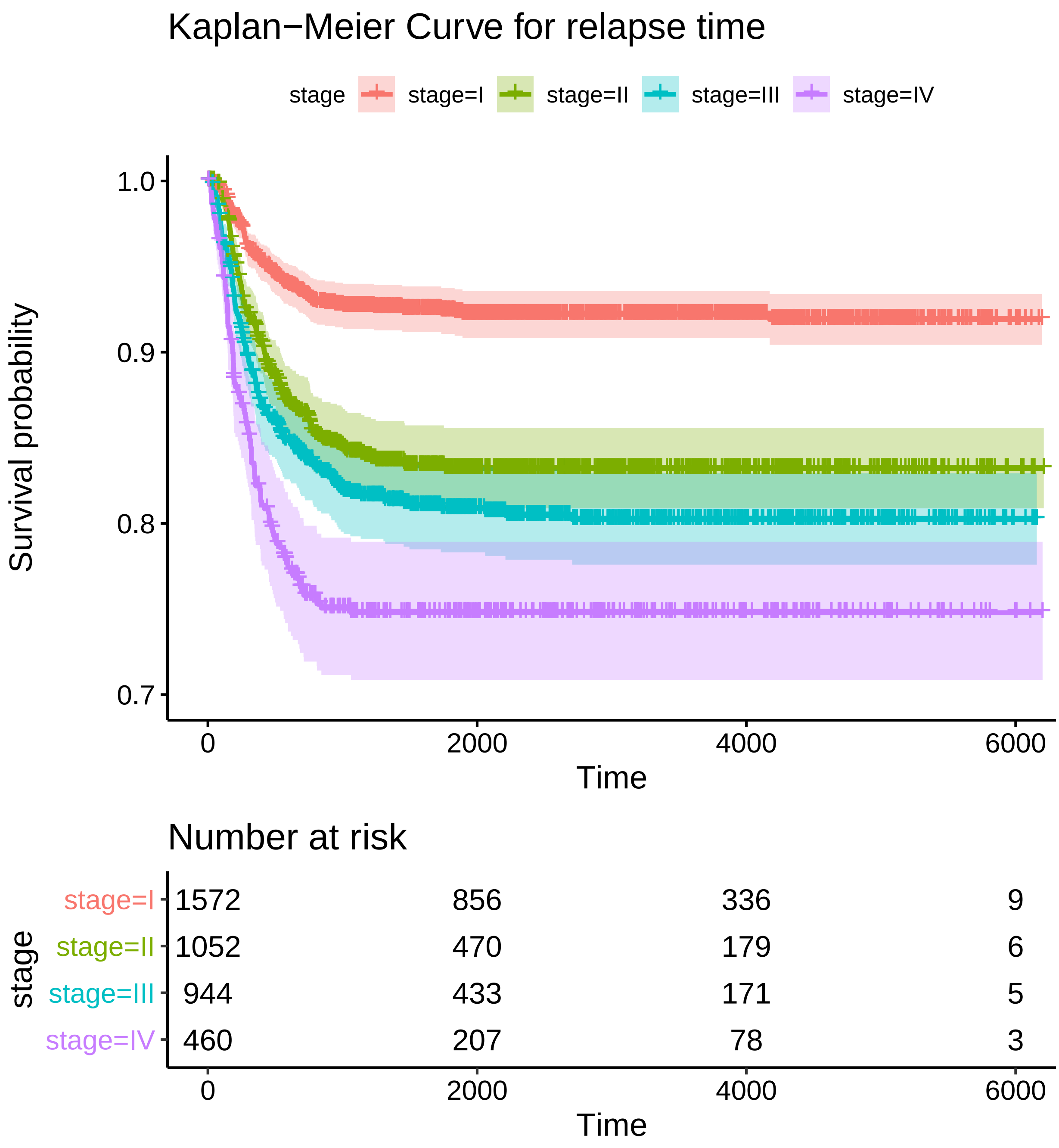


The log log plot shows the proportional assumption is not satisfied in the beginning, but after some time, it seems really proportional. We can also perform log rank test to test whether there is a significant difference between those 2 K-M curves:

```
## Call:
## survdiff(formula = Surv(edrel, rel) ~ histol, data = nwtco_prepared)
##
##                N Observed Expected (O-E)^2/E (O-E)^2/V
## histol=FH 3569      377    518.5      38.6       421
## histol=UH  459      194     52.5     382.0       421
##
##  Chisq= 421  on 1 degrees of freedom, p= <2e-16
```

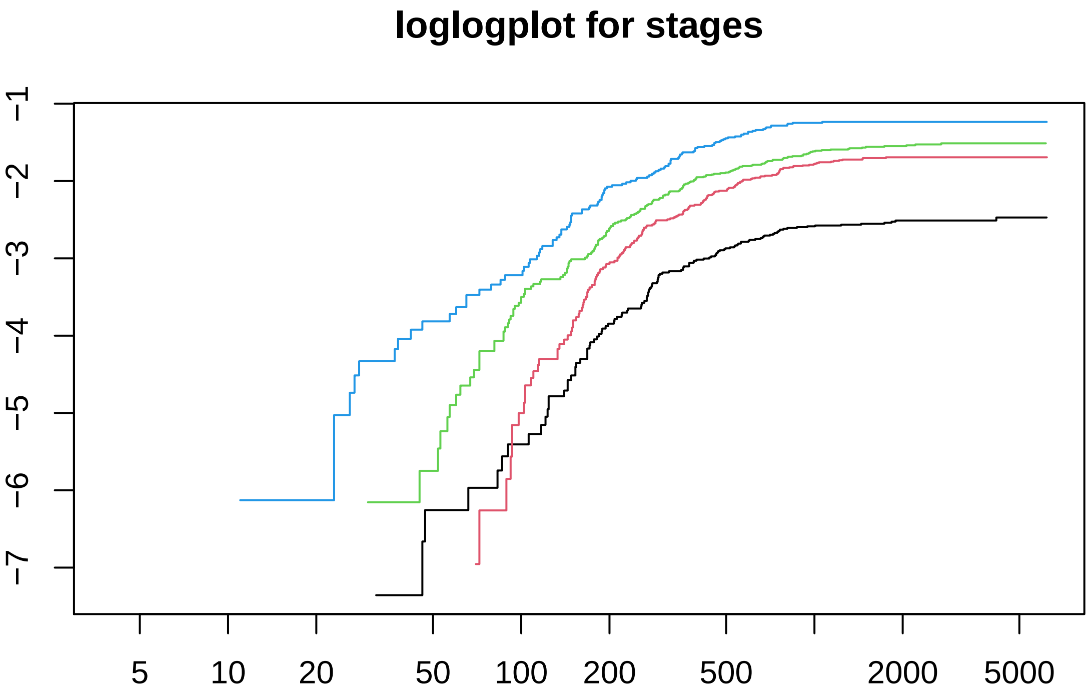
P-value is smaller than 0.05 so we can conclude that the histol variable is significant.

- Data partitioned by stage



stage	0	2000	4000	6000
stage=I	1572	856	336	9
stage=II	1052	470	179	6
stage=III	944	433	171	5
stage=IV	460	207	78	3

From Kaplan-Meier estimator we could see that there is a huge difference the histology group FH and UH. It's very obvious that the histology UH has a larger fatality rate. Then we use the log log plot to check the proportional hazard assumption used in Cox model:



loglogplot for stages

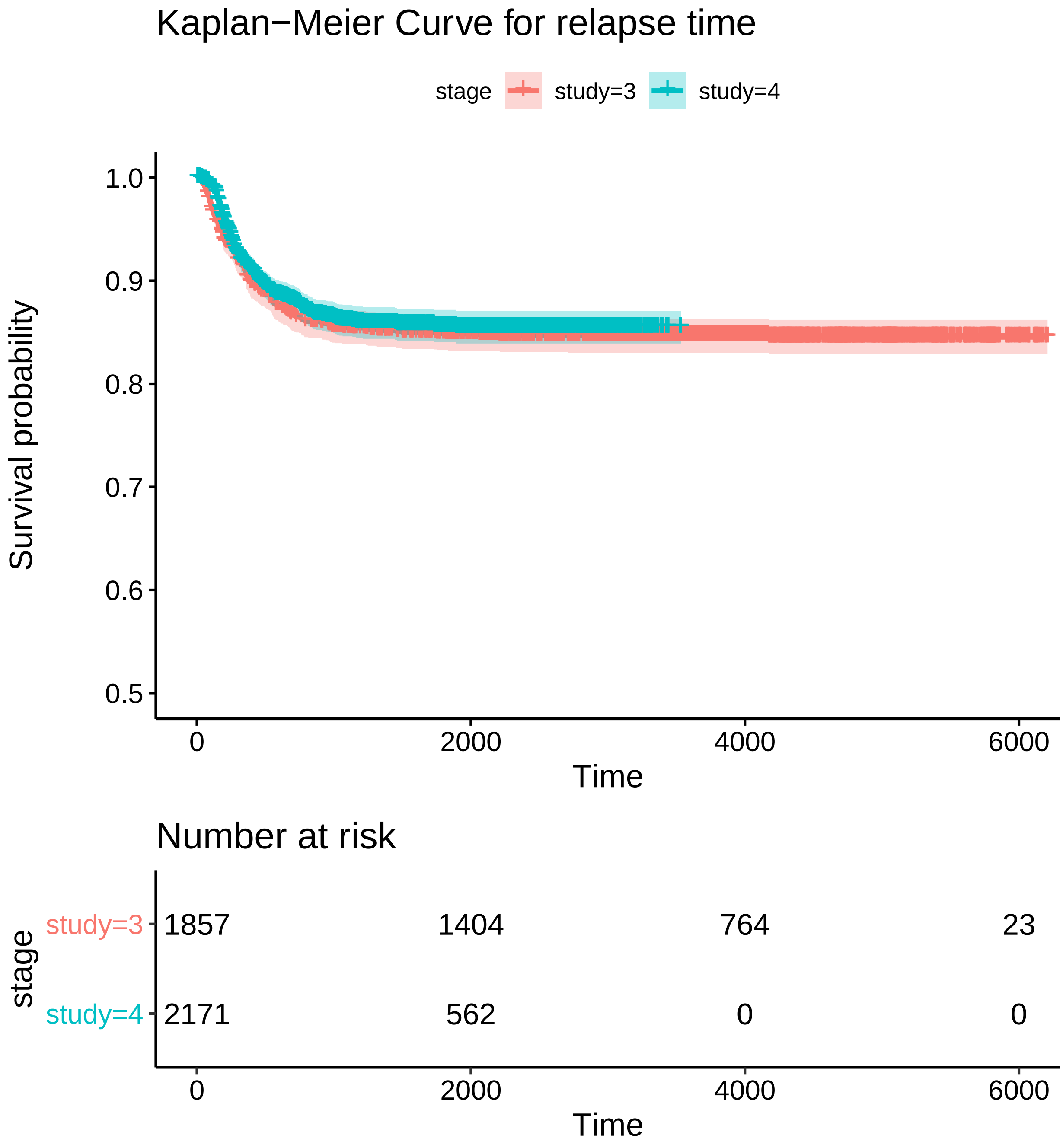
The log log plot shows the proportional assumption is violated by the stage 1 (black) and stage 2 (red) as there is an intersection. But after some time, the remaining events are proportional.

Then is the log rank test:

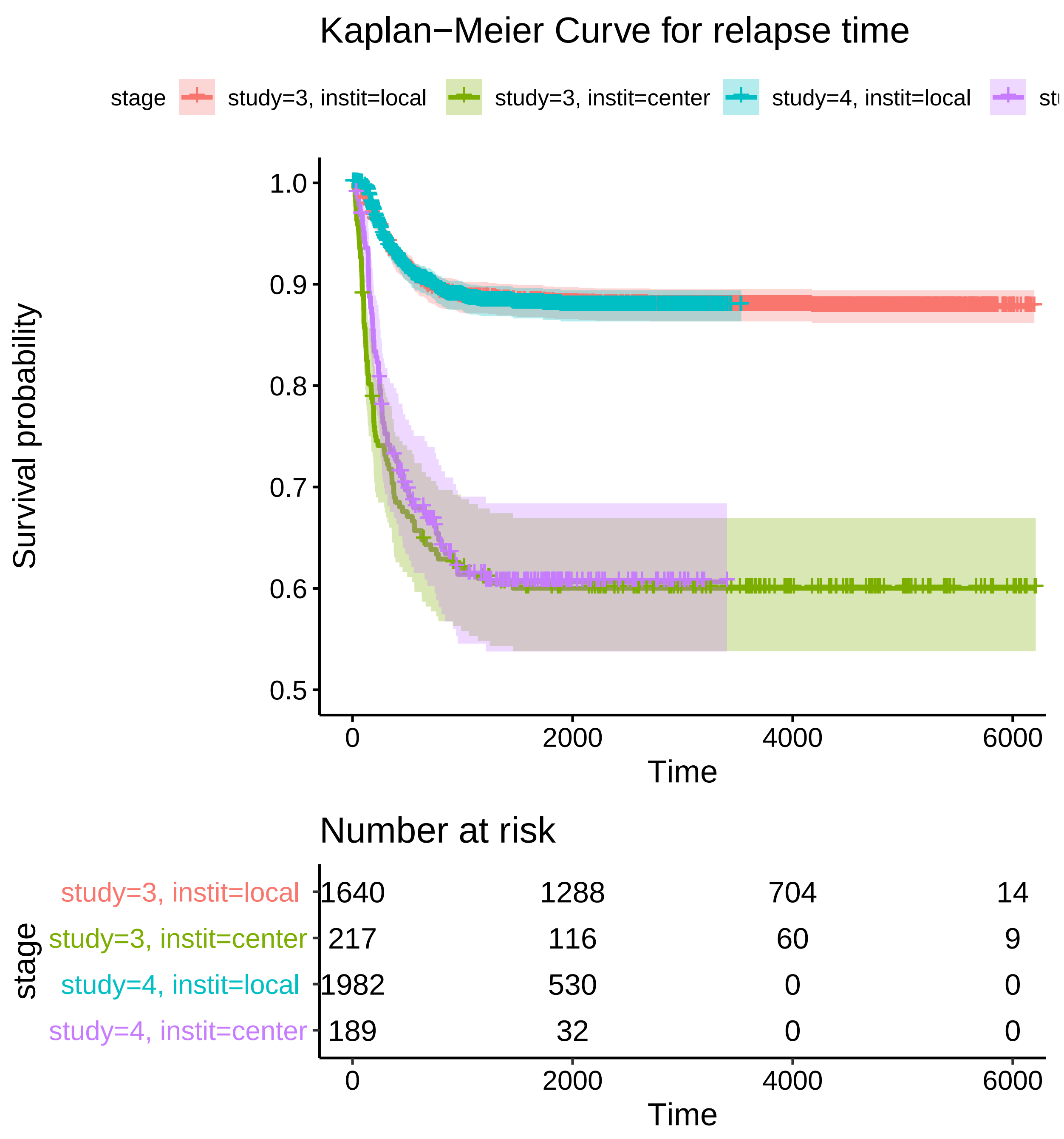
```
## Call:
## survdiff(formula = Surv(edrel, rel) ~ stage, data = nwtco_prepared)
##
##                N Observed Expected (O-E)^2/E (O-E)^2/V
## stage=I    1572      117    232.5     57.41     96.92
## stage=II   1052      166    148.0      2.19      2.96
## stage=III   944      175    130.2     15.43     19.99
## stage=IV    460      113     60.3     46.09     51.55
##
##  Chisq= 121  on 3 degrees of freedom, p= <2e-16
```

P-value is smaller than 0.05 so we can conclude that the stage variable is significant.

- Study



stage	0	2000	4000	6000
study=3	1857	1404	764	23
study=4	2171	562	0	0



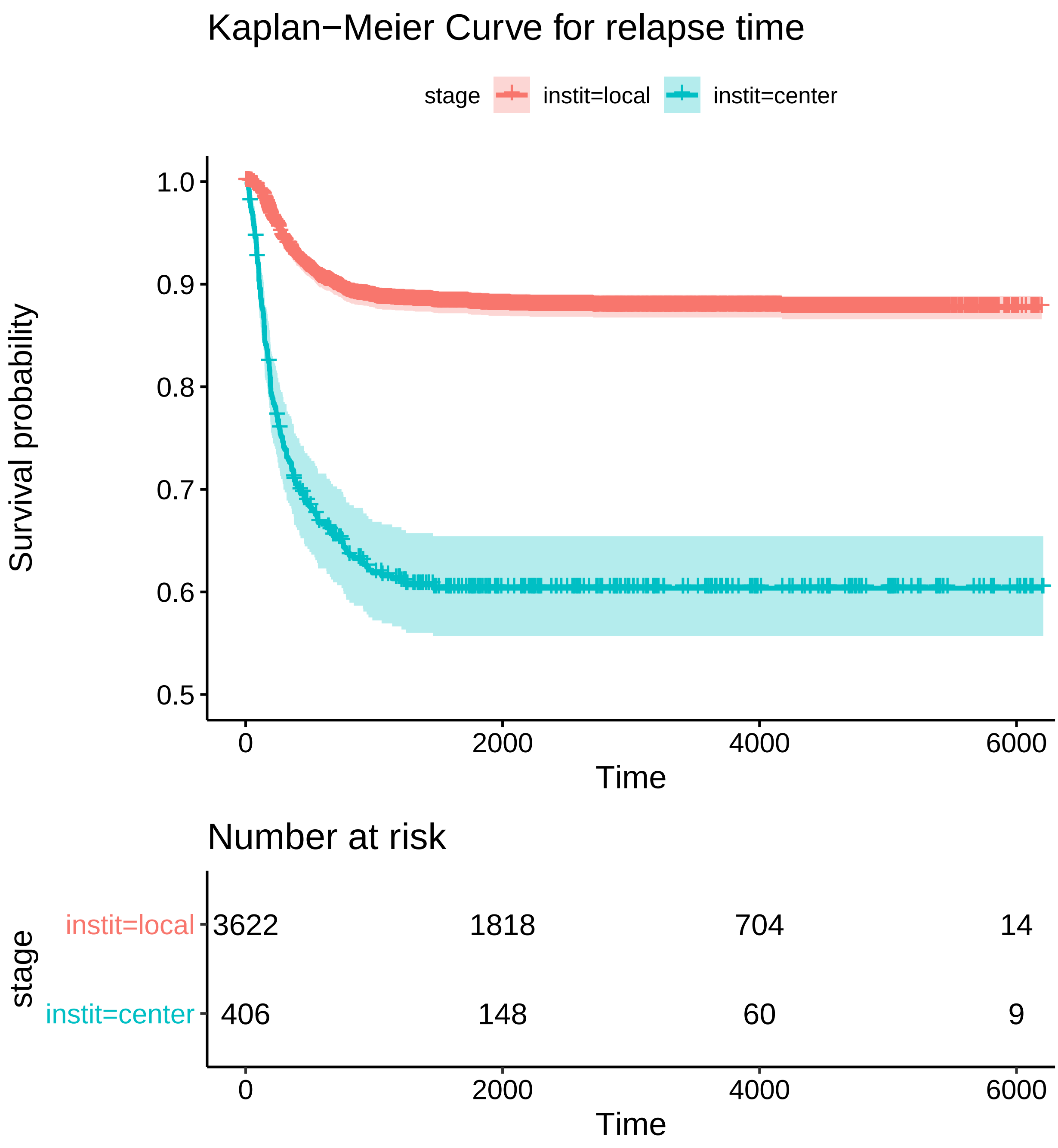
From the KM estimator, the study itself has nearly no effect, and still no effect even we consider the interaction with institution.

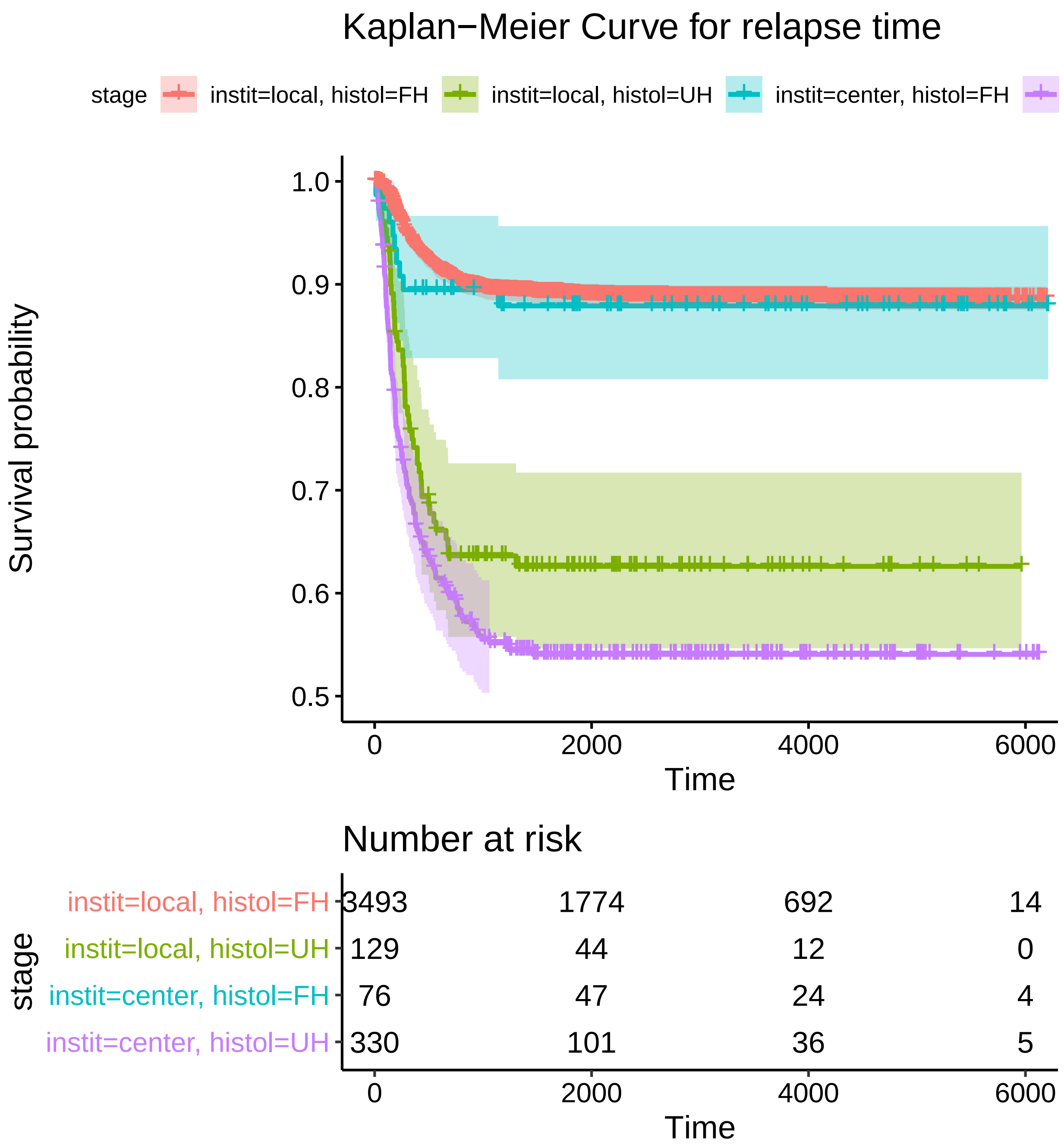
```
## Call:
## survdiff(formula = Surv(edrel, rel) ~ study, data = nwtco_prepared)
##
##             N Observed Expected (O-E)^2/E (O-E)^2/V
## study=3 1857      282      272     0.347     0.668
## study=4 2171      289      299     0.316     0.668
##
##  Chisq= 0.7  on 1 degrees of freedom, p= 0.4
```

The log rank P-value is 0.4 so the study effect is not significant.

So generally study should not be included in the variable.

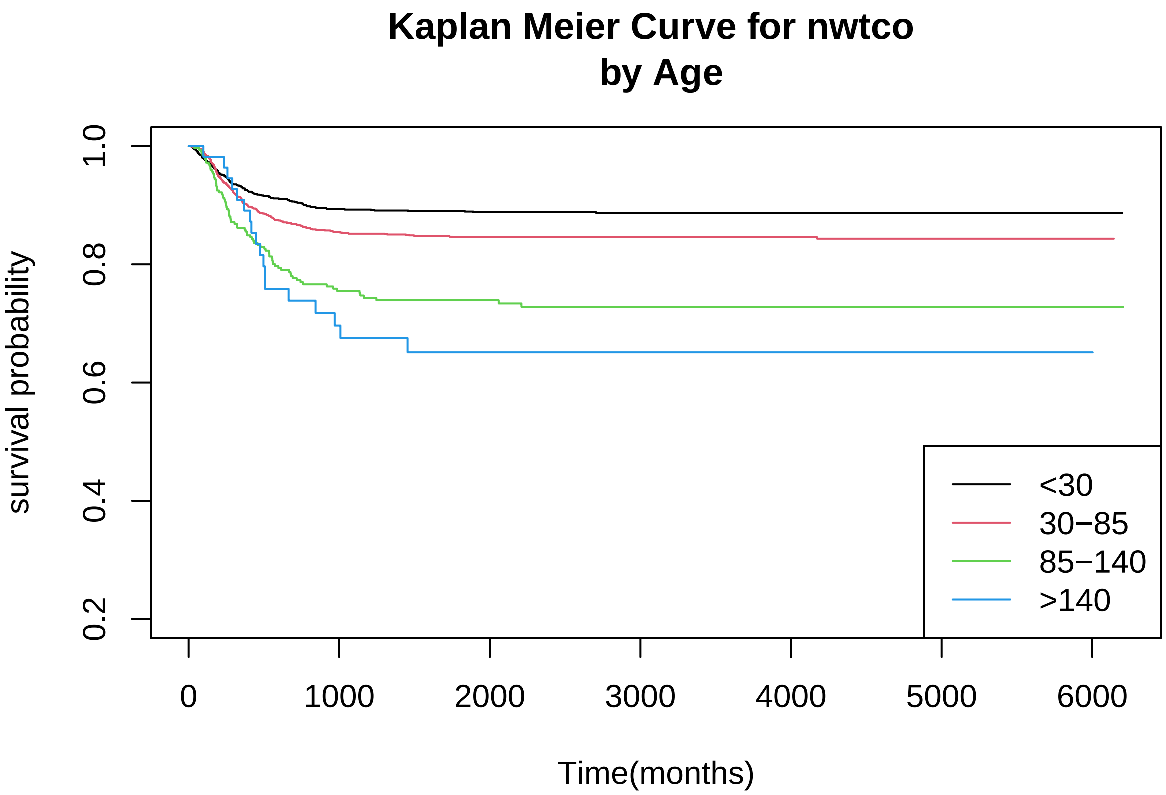
- institution and histology





From the KM plot of institution we can observe there is a significant effect of institution. But when we plot the KM plot of 4 different groups from institution and histology, we observe that the significance from institution was mainly resulted from the difference of proportion of the significant variable, histology. But we also observe a difference in the survival function of different institution when the histology is labeled UH.

- age for different groups



Apparently the age violates the proportional assumption.

So from the visualization and the preliminary log-rank test, we conclude that the two category variable are really important to include in the cox model. But for the stage category variable, the cox proportional hazard assumption is not satisfied so we may need to stratify them. We may need to consider the interaction between institution and histology.

The study category is not significant at all.

Cox proportional hazard model

To build the Cox proportional hazard model on the time-constant covariates, we need to first determine the variables. We already have the insights from the K-M estimation, and we could select the final model by the smallest AIC.

Use all study data:

Then we check the tests for proportional-hazards assumption. It could be obtained from *cox.zph()*, which computes a test for each covariates. The function tests proportional hazard assumption based on Schoenfeld residuals and test for the independence between residuals and time:

```
##        chisq df       p
## instit  17.7  1 2.6e-05
## histol  15.3  1 9.2e-05
## stage   10.1  3   0.018
## age     26.0  1 3.4e-07
## GLOBAL  58.3  6 9.9e-11
```

From the result we can see all of the variables are significantly dependent with time, thus violating the proportional hazard assumption. So actually this full data is not suitable for proportional model.

Use study 3 or study 4 data:

Study 3 data full model check:

```
##        chisq df       p
## instit 11.60  1 0.00066
## histol  9.86  1 0.00169
## stage   7.57  3 0.05568
## age    16.81  1 4.1e-05
## GLOBAL 38.30  6 9.8e-07
```

Study 4 data full model check:

```
##        chisq df      p
## instit  3.92  1 0.0477
## histol  5.21  1 0.0224
## stage   3.01  3 0.3901
## age     9.22  1 0.0024
## GLOBAL 19.33  6 0.0036
```

We observe that in the study 4, the variable stage does not seems to violate the cox proportional hazard assumption. So we could try the proportional model on the study 4 data only.

We first exclude age from the proportional part. Given that the age is significant, we should stratify the age as below:

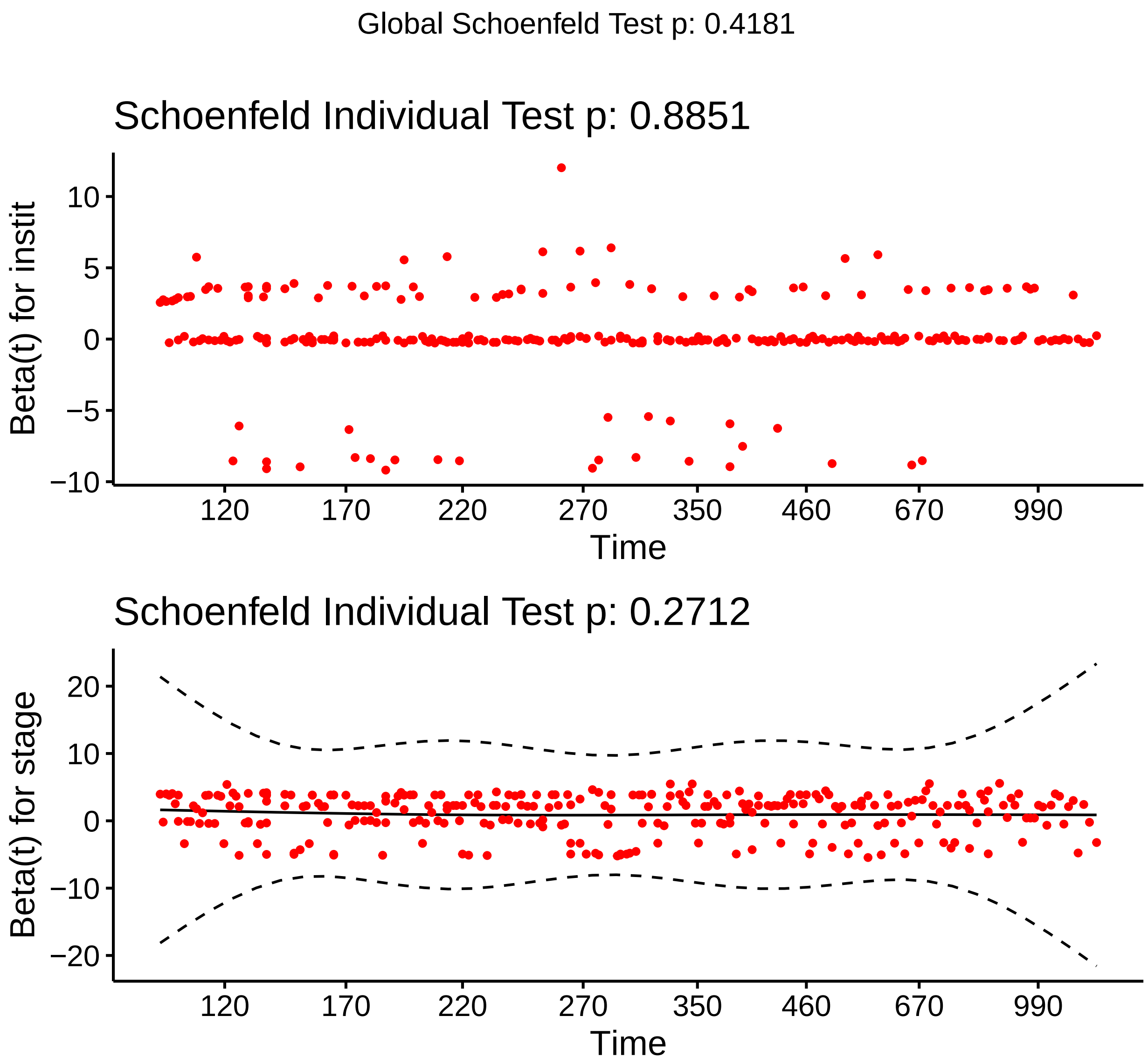
and we fit the cox model with stratified age and test the residual with the proportional assumption again:

```
##        chisq df     p
## instit  3.30  1 0.069
## histol  4.68  1 0.031
## stage   4.19  3 0.241
## GLOBAL  8.65  5 0.124
```

The remaining predictors has histol is still significant dependent, so we stratify the variable histol, fit the third model and test the assumption again:

```
##         chisq df    p
## instit 0.0209  1 0.89
## stage  3.9111  3 0.27
## GLOBAL 3.9119  4 0.42
```

We could also plot the Schoenfeld residuals here:



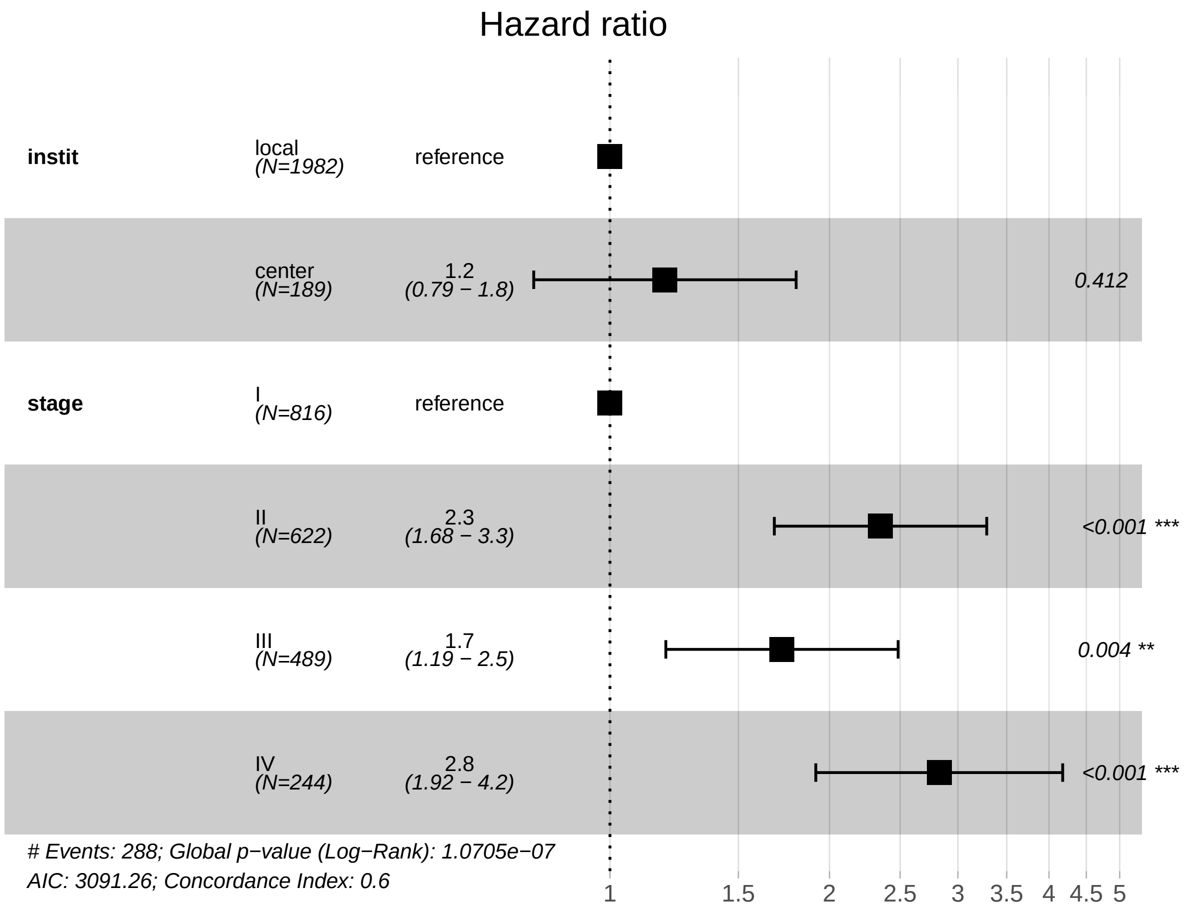
Now we observe all of the p values are really high, and the horizon line is within the 95% CI of the confidence interval, so there is no evidence the proportional assumption is violated. So our model is summary as below:

```
## Call:
## coxph(formula = Surv(edrel, rel) ~ instit + strata(histol) +
##     strata(age) + stage, data = nwtco_prepared4)
##
##   n= 2160, number of events= 288
##    (11 observations deleted due to missingness)
##
##                 coef exp(coef) se(coef)     z Pr(>|z|)
## institcenter 0.1736    1.1895   0.2114 0.821   0.4116
## stageII      0.8543    2.3496   0.1711 4.993 5.94e-07 ***
## stageIII     0.5431    1.7214   0.1871 2.903   0.0037 **
## stageIV      1.0397    2.8284   0.1989 5.228 1.72e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
##              exp(coef) exp(-coef) lower .95 upper .95
## institcenter     1.190     0.8407     0.786     1.800
## stageII          2.350     0.4256     1.680     3.286
## stageIII         1.721     0.5809     1.193     2.484
## stageIV          2.828     0.3536     1.915     4.177
##
## Concordance= 0.602  (se = 0.021 )
## Likelihood ratio test= 38.1  on 4 df,   p=1e-07
## Wald test            = 35.06  on 4 df,   p=5e-07
## Score (logrank) test = 36.76  on 4 df,   p=2e-07
```

Hazard ratios and C.I.

Now we use the *ggforest()* function to visualize the hazard ratio and their confidence interval for each covariate obtained from the fitted model summary:

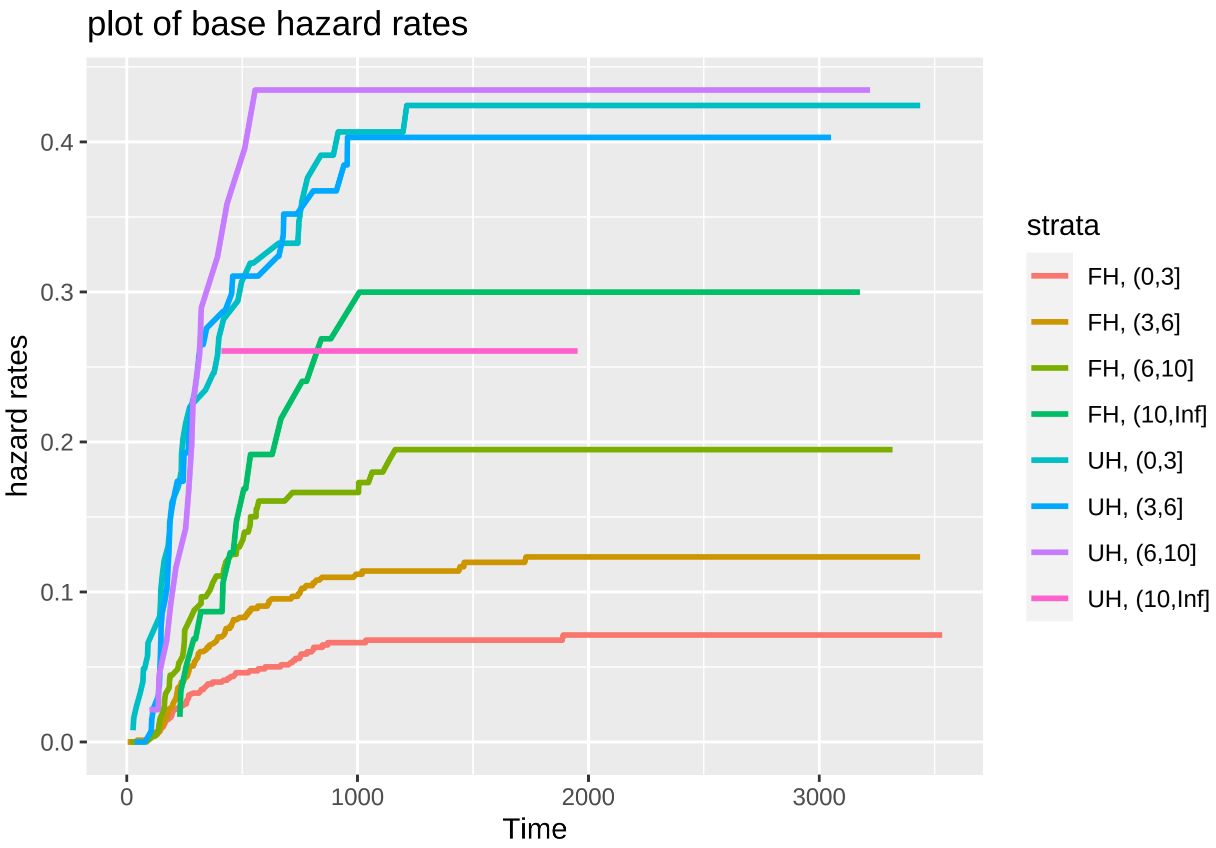


For the information in the hazard ratio chart above, we know that hazard ratio of all stages are significant. Stage 2 is centered at 2.3 with the 95% CI 1.68-3.3, stage 3 is centered at 1.7 with 95% CI 1.19-2.5, stage 4 is centered at 2.8 with 95% CI 1.92-4.2. It is interesting that the stage 2 has higher fatality rate then the stage 3. On the other hand, the institution is not significant, even the mean hazard ratio from center is higher.

So with all of the 3 stages p value < 0.05, we can conclude that the stage effect is significant for the relapse. Stage 2, 3, 4 patients are more likely to relapse.

Baseline Hazard Rates for strata histology

Though we don't have any coefficients for the covariates histology, we can plot the base hazard rate obtained by our cox ph model to illustrate the histology influence.



From the plot above, we clearly see that group with histology UH has higher baseline hazard rates than the FH group (most of the UH curves are lower than the FH curves. The UH, (10,Inf] is lower than FH, (10,Inf) may because of the small observations).

As a result, the UH histology group. will suffer a much higher fatality rate compared than the FH histology group.

We can't observe a clear age relationship between the UH group, but it seems clear that in FH group, as the age increase, the baseline hazard rate increase.

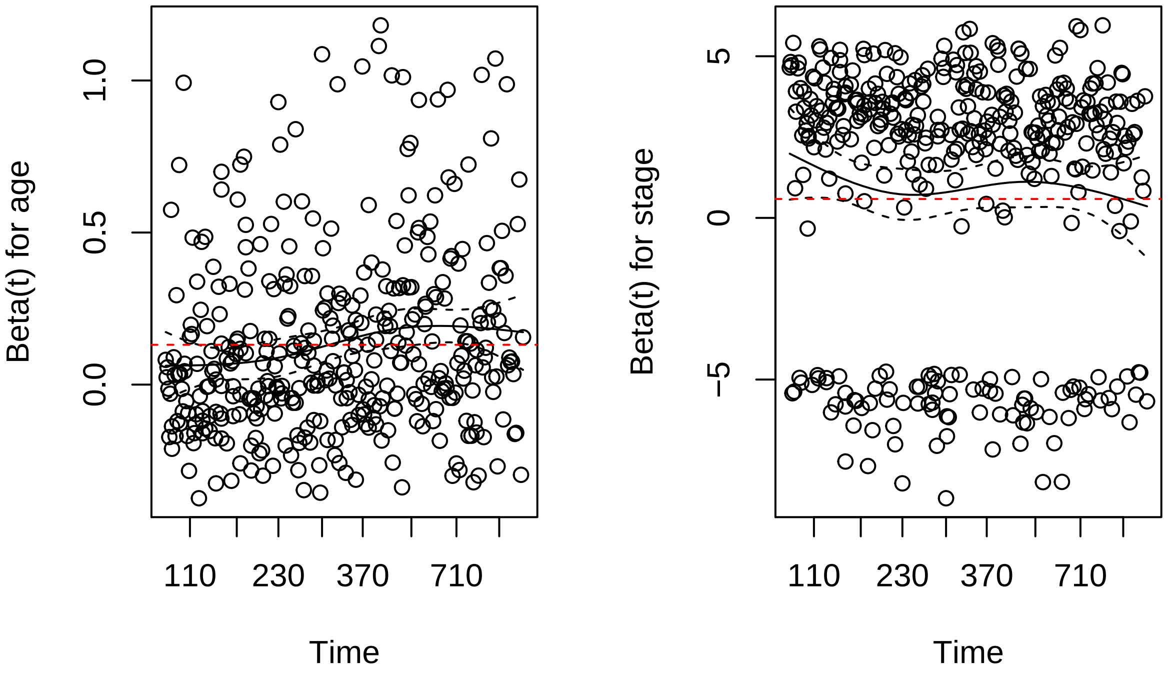
Though we are interested in the interaction of institution and histology, but as histology is stratified, we can't model the interaction from the cox ph model.

Step Functions for Time-Varying Coefficients in Cox model

The stratified Cox model can also be used in settings in which a continuous covariate does not satisfy the proportional hazards assumption, and we want to fit instead a Cox model with a time-varying coefficient.

As we analyzed in the last part, in the histology with FH group, the age seems to have a significant proportional effect. So now we only consider those from UH group.

```
##        chisq df      p
## age     8.48  1 0.0036
## stage   2.88  3 0.4099
## GLOBAL 13.00  4 0.0113
```



From the test and the plot we could observe that the covariate stage is time-dependent. But it seems the stage had a change in time 290. We make 2 different betas on those 2 intervals 0-290 and 290-inf:

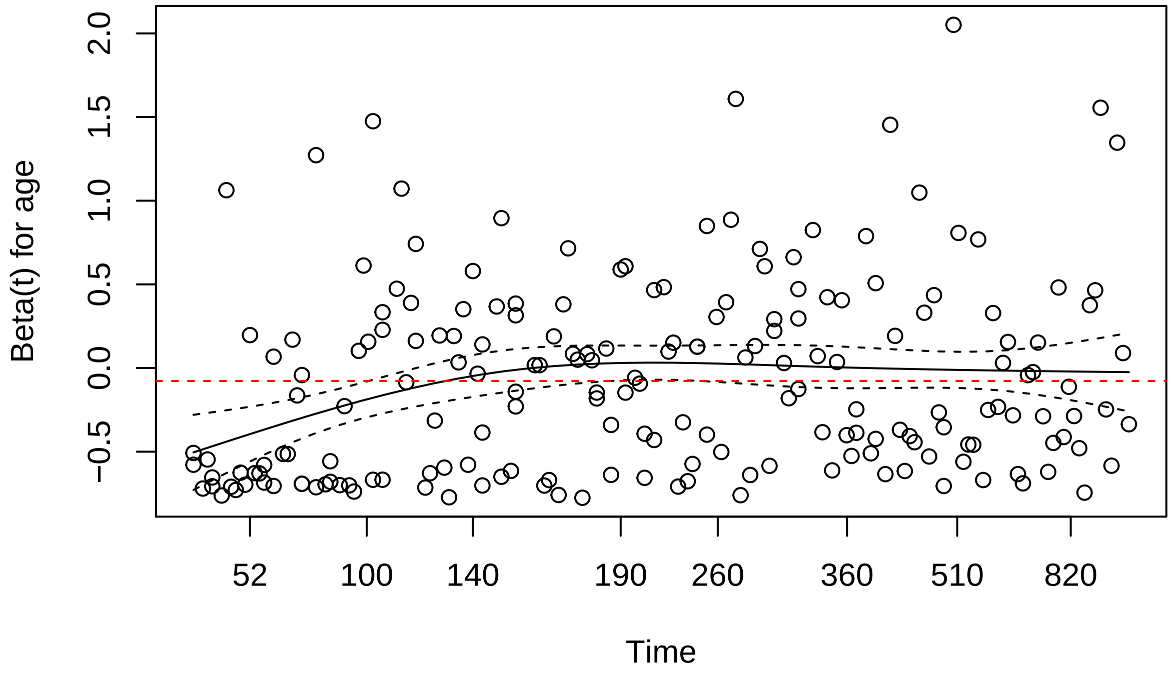
```
## Call:
## coxph(formula = Surv(tstart, edrel, rel) ~ age:strata(time_group) +
##     stage:strata(time_group), data = nwtco_prepared111)
##
##   n= 6889, number of events= 377
##
##                                            coef exp(coef) se(coef)     z
## age:strata(time_group)time_group=1      0.07936   1.08259  0.02592 3.062
## age:strata(time_group)time_group=2      0.17081   1.18627  0.02092 8.164
## strata(time_group)time_group=1:stageII  0.65841   1.93172  0.20177 3.263
## strata(time_group)time_group=2:stageII  0.52171   1.68490  0.19553 2.668
## strata(time_group)time_group=1:stageIII 0.45306   1.57312  0.21836 2.075
## strata(time_group)time_group=2:stageIII 0.59548   1.81389  0.19632 3.033
## strata(time_group)time_group=1:stageIV  0.70584   2.02555  0.24985 2.825
## strata(time_group)time_group=2:stageIV  0.65102   1.91749  0.22970 2.834
##                                         Pr(>|z|)
## age:strata(time_group)time_group=1       0.00220 **
## age:strata(time_group)time_group=2      3.25e-16 ***
## strata(time_group)time_group=1:stageII   0.00110 **
## strata(time_group)time_group=2:stageII   0.00763 **
## strata(time_group)time_group=1:stageIII  0.03800 *
## strata(time_group)time_group=2:stageIII  0.00242 **
## strata(time_group)time_group=1:stageIV   0.00473 **
## strata(time_group)time_group=2:stageIV   0.00459 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                                         exp(coef) exp(-coef) lower .95
## age:strata(time_group)time_group=1          1.083     0.9237     1.029
## age:strata(time_group)time_group=2          1.186     0.8430     1.139
## strata(time_group)time_group=1:stageII      1.932     0.5177     1.301
## strata(time_group)time_group=2:stageII      1.685     0.5935     1.149
## strata(time_group)time_group=1:stageIII     1.573     0.6357     1.025
## strata(time_group)time_group=2:stageIII     1.814     0.5513     1.235
## strata(time_group)time_group=1:stageIV      2.026     0.4937     1.241
## strata(time_group)time_group=2:stageIV      1.917     0.5215     1.222
##                                         upper .95
## age:strata(time_group)time_group=1          1.139
## age:strata(time_group)time_group=2          1.236
## strata(time_group)time_group=1:stageII      2.869
## strata(time_group)time_group=2:stageII      2.472
## strata(time_group)time_group=1:stageIII     2.413
## strata(time_group)time_group=2:stageIII     2.665
## strata(time_group)time_group=1:stageIV      3.305
## strata(time_group)time_group=2:stageIV      3.008
##
## Concordance= 0.649  (se = 0.014 )
## Likelihood ratio test= 116.8  on 8 df,   p=<2e-16
## Wald test            = 126.5  on 8 df,   p=<2e-16
## Score (logrank) test = 133.5  on 8 df,   p=<2e-16
```

We obtain 2coefficients for each previous estimate for the 2 different time stratas. From the result:

- The estimation of coefficient of age is clearly significant different in the two different time period. In interval 0-290 the estimate is 0.07936 and in interval 290-inf is 0.17081. The sd for both estimates are smaller than 0.03, so the difference between those 2 predictors are bigger than 2 times sd, which is very significant. So even though the age effect is time dependent, we can use this model to handle it.

Another interesting result comes from when we model the similar model on the age for the UH histology group:

```
##        chisq df      p
## age      9.9  1 0.0017
## GLOBAL   9.9  1 0.0017
```



As the residual had a change near time 160, we use the 2 different interval 0-160 and 160-inf and fit the time dependent model:

```
## Call:
## coxph(formula = Surv(tstart, edrel, rel) ~ age:strata(time_group),
##     data = nwtco_prepared112)
##
##   n= 838, number of events= 194
##
##                                         coef exp(coef)  se(coef)     z Pr(>|z|)
## age:strata(time_group)time_group=1 -0.208291  0.811971  0.060376 -3.45 0.000561
## age:strata(time_group)time_group=2 -0.009597  0.990449  0.038376 -0.25 0.802519
##
## age:strata(time_group)time_group=1 ***
## age:strata(time_group)time_group=2
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
##                                    exp(coef) exp(-coef) lower .95 upper .95
## age:strata(time_group)time_group=1    0.8120      1.232    0.7214     0.914
## age:strata(time_group)time_group=2    0.9904      1.010    0.9187     1.068
##
## Concordance= 0.575  (se = 0.024 )
## Likelihood ratio test= 14.03  on 2 df,   p=9e-04
## Wald test            = 11.96  on 2 df,   p=0.003
## Score (logrank) test = 11.92  on 2 df,   p=0.003
```

Now we can observe the effect of age in the interval 0-160 is significant negative for the relapse. It is really interesting that in the UM group and in the early time, the age has a totally different effect with the common sense.

Citation

1. NE Breslow and N Chatterjee (1999), Design and analysis of two-phase studies with binary outcome applied to Wilms tumour prognosis. Applied Statistics 48, 457–68.
2. Time dependent cox model

http://www.drizopoulos.com/courses/emc/ep03_%20survival%20analysis%20in%20r%20companion#step-functions-for-time-varying-coefficients

Appendix: Rcode

```
knitr::opts_chunk$set(echo=FALSE, message=FALSE, warning=FALSE, paged.print=FALSE)
library(KMsurv)
library(survival)
library(tidyverse)
library(survminer)
## Create factors
nwtco_prepared <- within(nwtco, {
    instit <- factor(instit, labels = c("local", "center"))
    histol <- factor(histol, labels = c("FH","UH"))
    stage  <- factor(stage,  labels = c("I","II","III","IV"))
    age    <- age / 12 # Age in years
 })

## Check
head(nwtco_prepared)
summary(nwtco_prepared)

KM_fit <- survfit(Surv(edrel, rel) ~ 1, data = nwtco_prepared)
plot(KM_fit, xlab = "Time to Relapse (days)", ylab = "Survival Probability", ylim = c(0.8,1),
    main = "Kaplan-Meier Estimate of S(t)")
ggsurvplot(survfit(Surv(edrel, rel) ~ histol,
                   data = nwtco_prepared),
           conf.int=TRUE,risk.table=TRUE, ylim = c(0.4,1),
           legend.title="Histology",
           palette=c("dodgerblue2", "orchid2"),
           title="Kaplan-Meier Curve for relapse time",
           risk.table.height=.30)
# *log log plot*
plot(survfit(Surv(edrel, rel) ~ histol,
                   data = nwtco_prepared),
     col=c("black", "red"), fun="cloglog")
title("loglogplot for histol")
survdiff(Surv(edrel, rel) ~ histol,
                   data = nwtco_prepared)
```

```r
ggsurvplot(survfit(Surv(edrel, rel) ~ stage,
                   data = nwtco_prepared),
           conf.int=TRUE,risk.table=TRUE, ylim = c(0.7,1),
           legend.title="stage",
           title="Kaplan-Meier Curve for relapse time",
           risk.table.height=.30)
# *log log plot*
plot(survfit(Surv(edrel, rel) ~ stage,
                   data = nwtco_prepared),
     col=c(1,2,3,4), fun="cloglog")
title("loglogplot for stages")
survdiff(Surv(edrel, rel) ~ stage,
                   data = nwtco_prepared)
ggsurvplot(survfit(Surv(edrel, rel) ~ study,
                   data = nwtco_prepared),
           conf.int=TRUE,risk.table=TRUE, ylim = c(0.5,1),
           legend.title="stage",
           title="Kaplan-Meier Curve for relapse time",
           risk.table.height=.30)

ggsurvplot(survfit(Surv(edrel, rel) ~ study+instit,
                   data = nwtco_prepared),
           conf.int=TRUE,risk.table=TRUE, ylim = c(0.5,1),
           legend.title="stage",
           title="Kaplan-Meier Curve for relapse time",
           risk.table.height=.30)
survdiff(Surv(edrel, rel) ~ study,
                   data = nwtco_prepared)
ggsurvplot(survfit(Surv(edrel, rel) ~ instit,
                   data = nwtco_prepared),
           conf.int=TRUE,risk.table=TRUE, ylim = c(0.5,1),
           legend.title="stage",
           title="Kaplan-Meier Curve for relapse time",
           risk.table.height=.30)

ggsurvplot(survfit(Surv(edrel, rel) ~ instit+histol,
                   data = nwtco_prepared),
           conf.int=TRUE,risk.table=TRUE, ylim = c(0.5,1),
           legend.title="stage",
           title="Kaplan-Meier Curve for relapse time",
           risk.table.height=.30)
nwtco1 = nwtco
nwtco1$age = as.factor(ceiling((nwtco$age+25)/55))
levels(nwtco1$age) = c("<30","30-85","85-140",">140")
nwtco_gfit = survfit(Surv(edrel,rel)~age,data = nwtco1)
plot(nwtco_gfit, col=c(1,2,3,4),
conf.int = F,
main="Kaplan Meier Curve for nwtco \n by Age",
ylab = "survival probability",ylim = c(0.2,1),
xlab = "Time(months)")
legend("bottomright",legend = c("<30","30-85","85-140",">140"),col = c(1,2,3,4), lty = 1)
nwtco_prepared3 = nwtco_prepared[nwtco_prepared$study == 3,]
nwtco_prepared4 = nwtco_prepared[nwtco_prepared$study == 4,]
```

```
Coxfull = coxph(formula = Surv(edrel, rel) ~
   instit + histol  + stage + age,
   data = nwtco_prepared)
check = cox.zph(Coxfull, transform = "km")
check
Coxfull = coxph(formula = Surv(edrel, rel) ~
   instit + histol  + stage + age,
   data = nwtco_prepared3)
check = cox.zph(Coxfull, transform = "km")
check
Coxfull = coxph(formula = Surv(edrel, rel) ~
   instit + histol  + stage + age,
   data = nwtco_prepared4)
check = cox.zph(Coxfull, transform = "km")
check
nwtco_prepared4$age = cut(nwtco_prepared4$age,c(0,3,6,10,Inf))
Coxfull = coxph(formula = Surv(edrel, rel) ~
   instit + histol + strata(age) + stage ,
   data = nwtco_prepared4)
check = cox.zph(Coxfull, transform = "km")
check
Coxfit = coxph(formula = Surv(edrel, rel) ~
   instit + strata(histol) + strata(age) + stage,
   data = nwtco_prepared4)
check = cox.zph(Coxfit, transform = "km")
check
ggcoxzph(check)
summary(Coxfit)
attr(Coxfit$terms,"dataClasses") = attr(Coxfit$terms,"dataClasses")[c(-3,-4)]
ggforest(Coxfit,data = nwtco_prepared4)
df = basehaz(Coxfit)
ggplot(df, aes(x = time, y = hazard, color = strata)) + geom_line(size = 1)+ labs(title = "plot of base
nwtco_prepared11 = nwtco_prepared[nwtco_prepared$histol == "FH",]
fit <- coxph(Surv(edrel, rel) ~
   age + stage,
   data = nwtco_prepared11)
cox.zph(fit)

par(mfrow = c(1,2))
plot(cox.zph(fit), var = 1)
abline(h = coef(fit)[1], lty = 2, col = "red")


plot(cox.zph(fit), var = 2)
abline(h = coef(fit)[2], lty = 2, col = "red")
nwtco_prepared111 <- survSplit(Surv(edrel, rel) ~
   age + stage, data = nwtco_prepared11,
                 cut = c(290), episode = "time_group")
fit_tv_coef <- coxph(Surv(tstart,edrel, rel) ~
                       age:strata(time_group) +
                       stage:strata(time_group),
                     data = nwtco_prepared111)
summary(fit_tv_coef)
```

```r
nwtco_prepared12 = nwtco_prepared[nwtco_prepared$histol == "UH",]
fit <- coxph(Surv(edrel, rel) ~
   age ,
   data = nwtco_prepared12)
cox.zph(fit)

plot(cox.zph(fit), var = 1)
abline(h = coef(fit)[1], lty = 2, col = "red")
nwtco_prepared112 <- survSplit(Surv(edrel, rel) ~
   age , data = nwtco_prepared12,
                   cut = c(160), episode = "time_group")
fit_tv_coef <- coxph(Surv(tstart,edrel, rel) ~
                       age:strata(time_group) ,
                     data = nwtco_prepared112)
summary(fit_tv_coef)
```