Lecture twenty-one: Cox Model with Time-dependent Covariates (II)

Time-dependent covariate was used for testing PH assumption. We also illustrated the usage of discrete time-dependent covariate (eg. UCB: neutrophil recovery for leukemia patients after UCB transplant). The difficulty is to generate the right format of input data for the SAS PHREG or Splus function coxph() when there are more than one time-dependent covariates measured at irregular time points during a study.

1. **Example 8.1: BMT in the treatment of leukaemia**

   The data (Table 8.1, page 308) contain following variables:

```
ST: survival time in days,          CENSOR: censoring indicator
PAGE: age of patient,               DAGE: age of donor
GROUP: disease group (1 = ALL, 2 = low-risk AML, 3 = high-risk AML)
P: Platelet recovery indicator (0 = no, 1 = yes)
PTIME: time in days to return of platelets to normal level (if P = 1).
```

   The intermediate event after transplantation, *PTIME*, is a time-dependent variable. Define

$$tplate(t) = \begin{cases} 0 & \text{if } t < \text{PTIME} \\ 1 & \text{if } t \geq \text{PTIME} \end{cases} \qquad (1)$$

   (a) Cox model with **PAGE** and **DAGE**

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
AGE	1	-0.01395	0.03941	0.1253	0.7234
DAGE	1	0.04640	0.04275	1.1782	0.2777

(b) Cox model with platelet recovery (**TPLATE**)

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	67.131	62.207

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	4.9242	1	0.0265
Score	8.4539	1	0.0036
Wald	4.8110	1	0.0283

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
tplate	1	-2.69640	1.22932	4.8110	0.0283

(c) Cox model with disease group and **TPLATE**

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
X1	1	2.07607	1.15104	3.2532	0.0713
X3	1	2.46550	1.20765	4.1680	0.0412
tplate	1	-2.31956	1.22938	3.5599	0.0592

Variable	Hazard Ratio	95% Hazard Ratio Confidence Limits	
X1	7.973	0.835	76.101
X3	11.769	1.104	125.519
tplate	0.098	0.009	1.094

(d) SAS programs: ex81.sas and ex81a.sas (counting process format - see Example 8.4 and Table 8.9 as well) were posted at the course website.

2. **Example 8.2: Chemotherapy in ovarian cancer patients (cont.)**

The data (Table 5.6) were analyzed using Weibull PH model (page 213), and the final model includes **age** and **treat**. We check PH assumption for age.

(a) Cox model without interaction between age and survival time

$$\hat{h}_i(t) = exp\{0.147Age_i - 0.796Treat_i\}h_0(t).$$

The value of $-2log\,\hat{L}$ for this model is 54.148.

(b) Test PH assumption for age: Cox model with interaction between age and survival time

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	69.970	53.613
AIC	69.970	59.613
SBC	69.970	61.068

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
AGE	1	0.21565	0.11261	3.6673	0.0555
aget	1	-0.0002031	0.0002832	0.5141	0.4734
c	1	-0.66359	0.66955	0.9823	0.3216

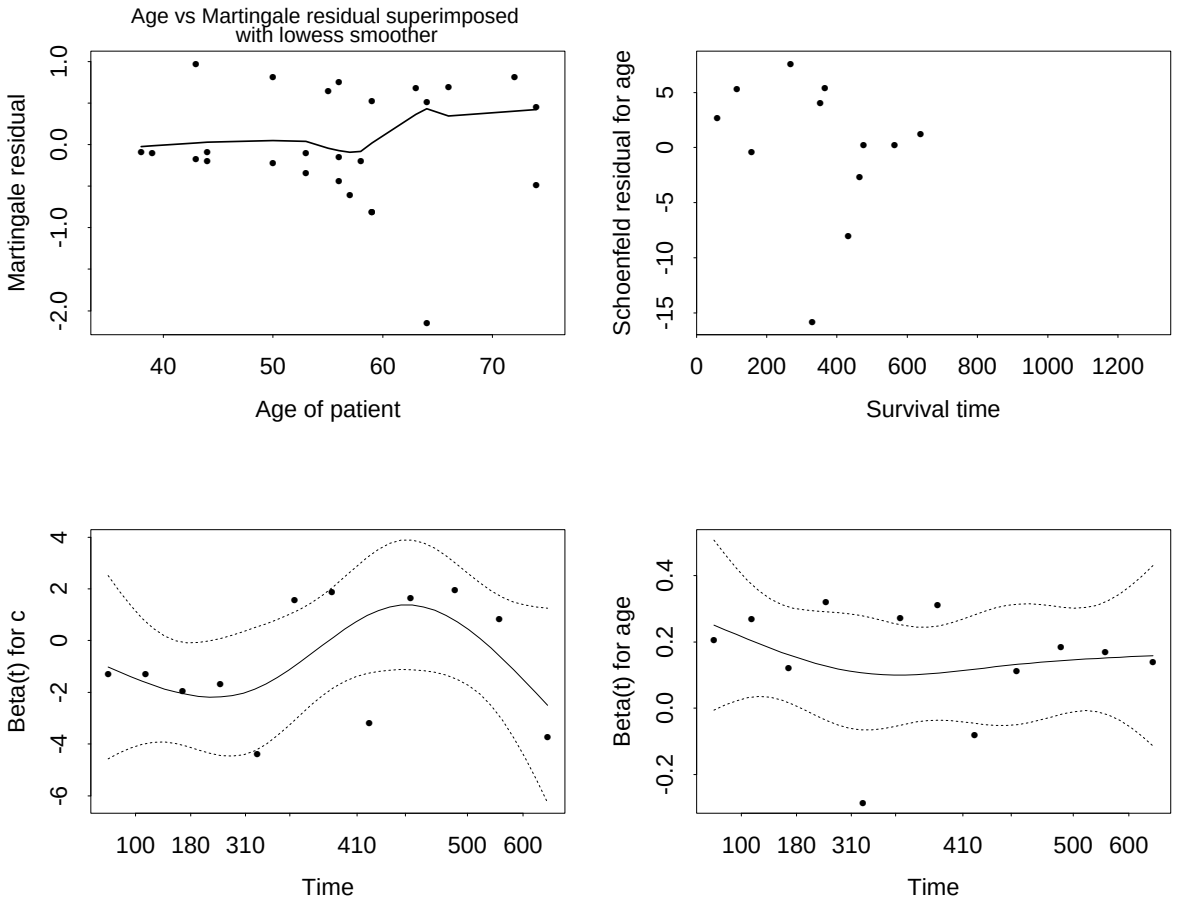
(c) SAS program

```
options ls=80;
libname fu '../sdata';
data work;
        set fu.ovarian;
        c = treat - 1;
proc phreg;
```

```
        model survt*cens(0) = c age /risklimits;
proc phreg;
        model survt*cens(0) =  age aget c  /risklimits;
                aget = age*survt;
run;
```

(d) Figures: Residual plots and plots from cox.zph

Figure 1:



(e) Another type of test of PH assumption for age: The results from cox.zph()

```
          rho chisq     p
     c  0.203 0.501 0.479
   age -0.105 0.145 0.704
GLOBAL     NA 0.758 0.685
```

(f) Splus program

```
ex82a.s<-function(){
        tmpdf <- importData("../../sdata/rsas7bdat")
        motif()
        par(mfrow=c(2,2))
        attach(tmpdf)
        treat <- c + 1
        plot(age, mart, xlab="Age of patient", xlim=c(35,75),
                ylab="Martingale residual", ylim=c(-2.16, 1))
        lines(lowess(age, mart), lty=1, lwd=2)
        title("Age vs Martingale residual superimposed
                        with lowess smoother",cex=0.6)
        plot(survt, schage, xlab="Survival time",
            ylab="Schoenfeld residual for age", xlim=c(50,1300),
                ylim=c(-16,7.8))
        detach()
        coxfit <- coxph(Surv(survt, cens)~c+age, data=resovardf)
        zphout<-cox.zph(coxfit)
        plot(zphout, var = 1)
        plot(zphout, var = 2)
        zphout
}
```

3. **Example 8.3: Data from a hypothetical cirrhosis study**

(a) The data, which are based on Table 8.3 and Table 8.4, are available at the course website (cirrhos.dat). The variables measured are

Time: Survival time of patient in days
Cen: censoring indicator (0 = censored, 1 = event)
Treat: Treatment group (placebo = 0, liverol = 1)
Age: age of patients in years
lbr: logarithm of bilirubin level which measured many times during follow-up in addition to its baseline measurement

1	281	1	0	46	3.2	47	3.8	184	4.9	251	5.0	.	.	.	.
2	604	0	0	57	3.1	94	2.9	187	3.1	321	3.2	.	.	.	.
3	457	1	0	56	2.2	61	2.8	97	2.9	142	3.2	359	3.4	440	3.8
4	384	1	0	65	3.9	92	4.7	194	4.9	372	5.4	.	.	.	.
5	341	0	0	73	2.8	87	2.6	192	2.9	341	3.4	.	.	.	.
6	842	1	0	64	2.4	94	2.3	197	2.8	384	3.5	795	3.9	.	.
7	1514	1	1	69	2.4	74	2.9	202	3.0	346	3.0	917	3.9	1411	5.1
8	182	0	1	62	2.4	90	2.5	182	2.9	.	.	.	.	.	.
9	1121	1	1	71	2.5	101	2.5	410	2.7	774	2.8	1043	3.4	.	.
10	1411	0	1	69	2.3	182	2.2	847	2.8	1051	3.3	1347	4.9	.	.
11	814	1	1	77	3.8	167	3.9	498	4.3	.	.	.	.	.	.
12	1071	1	1	58	3.1	108	2.8	187	3.4	362	3.9	694	3.8	.	.

(b) For time-dependent variable *lbr*, the last recorded value was used (assume constant between any two adjacent time points) (recall: There were three options mentioned at last lecture).

(c) Model selection with a time-dependent variable

Table 1: Values of $-2log\,\hat{L}$ for models with a time-dependent variable

Terms in model	$-2log\,L$
null model	25.121
age	22.135
lbrt	12.053
age, lbrt	11.147
treat, lbrt	10.678

The estimated hazard function for the final model (with time-dependent variable lbrt and treatment) is

$$\hat{h}_i(t) = exp\{3.614 lbr_i(t) - 1.48 Treat_i\} h_0(t),$$

The output from SAS program for this model is

```
                   Model Fit Statistics

                            Without          With
           Criterion     Covariates     Covariates
```

-2 LOG L	25.121	10.678
AIC	25.121	14.678
SBC	25.121	14.837

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
TREAT	1	-1.48048	1.34050	1.2198	0.2694
lbrt	1	3.61362	2.23295	2.6190	0.1056

(d) Model selection without a time-dependent variable

Table 2: Values of $-2log\,\hat{L}$ for models without a time-dependent variable

Terms in model	$-2log\,L$
null model	25.121
age	22.135
lbr	21.662
age, lbr	18.475
age, treat, lbr	13.293

The estimated hazard function for the final model (with age, lbr and treatment) is

$$\hat{h}_i(t) = exp\{-0.085Age_i + 2.561lbr_i - 3.052Treat_i\}h_0(t),$$

The output from SAS program for this model is

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	25.121	13.293
AIC	25.121	19.293

```
                 SBC                  25.121          19.531

                 Analysis of Maximum Likelihood Estimates

                         Parameter    Standard
Variable      DF      Estimate       Error     Chi-Square    Pr > ChiSq

AGE            1      -0.08503     0.07728         1.2108        0.2712
LBR            1       2.56080     1.27211         4.0523        0.0441
TREAT          1      -3.05192     1.56150         3.8200        0.0506
```

(e) Two models with different conclusion about the effect of treatment (drug liverol). One possible explanation for the difference is that the effect of the treatment is to change the values of the bilirubin level, so the time-dependent variable has masked the treatment effect. If there is no clear treatment groups, it's important to find a biological marker or surrogate (bilirubin? CD4 T-cell?).

(f) SAS program: electronic copy is available at the course website (cirrhos1.sas).

```
options ls=80;
libname fu '../../sdata';
data fu.cirrhos;
infile '../../data/cirrhos.dat';
input id survt cen treat age lbr ft1 lbr1 ft2 lbr2 ft3 lbr3 ft4
         lbr4 ft5 lbr5;
proc phreg;
        model survt*cen(0) = age / covb risklimits;
proc phreg;
        model survt*cen(0)=lbrt /covb risklimits;
        if (survt < ft1 and ft1 ne .) then lbrt = lbr;
        if (ft1 <= survt and ft1 ne .) then lbrt = lbr1;
        if (ft2 <= survt and ft2 ne .) then lbrt = lbr2;
        if (ft3 <= survt and ft3 ne .) then lbrt = lbr3;
        if (ft4 <= survt and ft4 ne .) then lbrt = lbr4;
        if (ft5 <= survt and ft5 ne .) then lbrt = lbr5;
proc phreg;
        model survt*cen(0) = age lbrt /covb risklimits;
        if (survt < ft1 and ft1 ne .) then lbrt = lbr;
```

```
        if (ft1 <= survt and ft1 ne .) then lbrt = lbr1;
        if (ft2 <= survt and ft2 ne .) then lbrt = lbr2;
        if (ft3 <= survt and ft3 ne .) then lbrt = lbr3;
        if (ft4 <= survt and ft4 ne .) then lbrt = lbr4;
        if (ft5 <= survt and ft5 ne .) then lbrt = lbr5;
proc phreg;
        model survt*cen(0) = treat lbrt /covb risklimits;
        if (survt < ft1 and ft1 ne .) then lbrt = lbr;
        if (ft1 <= survt and ft1 ne .) then lbrt = lbr1;
        if (ft2 <= survt and ft2 ne .) then lbrt = lbr2;
        if (ft3 <= survt and ft3 ne .) then lbrt = lbr3;
        if (ft4 <= survt and ft4 ne .) then lbrt = lbr4;
        if (ft5 <= survt and ft5 ne .) then lbrt = lbr5;
proc phreg;
        model survt*cen(0)= lbr /covb risklimits;
proc phreg;
        model survt*cen(0) = age lbr / covb risklimits;
proc phreg;
        model survt*cen(0) = age lbr treat / covb risklimits;
run;
```