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

Time-dependent covariate was used to test the assumption of proportional hazards (section 4.4, page 139). We also fit the piece-wise Cox model (non-proportional hazards model) by introducing time-dependent covariate (recall: The two "catheter placements" for kidney dialysis patients).

1. **Fixed-time (time-independent) covariates and Time-dependent covariates**.

   Fixed-time covariates are those whose values are fixed throughout the course of the study. For example, sex, race, and many others in the previous examples.

   It is typical in many survival studies that individuals are monitored for the duration of the study and some explanatory variables are recorded whose values may change during the course of the study. For example, status of neutrophil recovery of patients with leukemia after transplantation (*discrete-time*), number of CD4 T-cells of HIV/AIDS patients measured at (ir)regular intervals (*continuous*).

2. **Types of time-dependent variable**

   (a) Internal variable: relates to a particular "alive" individual. Examples: lung function, number of CD4 T-cells, etc.

   (b) External variable: its value will be known in advance at any future time. Examples: patient's age, predetermined dose of a drug; enviromental factors.

3. **Cox model with time-dependent covariates**

   (a) The hazard function is

   $$h_i(t) = exp\{\sum_{j=1}^{p} \beta_j x_{ji}(t)\} h_0(t),$$

   where the first part is function of time. The hazard ratio is also time-dependent. Consider the hazard ratio for two individuals:

   $$\frac{h_r(t)}{h_s(t)} = exp\{\beta_1[x_{r1}(t) - x_{s1}(t)] + \ldots + \beta_p[x_{rp}(t) - x_{sp}(t)]\}.$$

(b) The interpretation of β-parameters: β_j is log-hazard ratio for two individuals whose value of x_j at t differs by one, with the same values of all other covariates **at time t**.

(c) The survival function is

$$S_i(t) = exp\{-\int_0^t exp(\sum_{j=1}^{p} \beta_j x_{ji}(u))h_0(u)du\}.$$

The survival function $S_i(t)$, which depends on the values of covariates over time, can not be expressed as a power of $S_0(t)$.

(d) Fitting the Cox model: The partial log-likelihood can be generalized to

$$\sum_{i=1}^{n} \delta_i \{\sum_{j=1}^{p} \beta_j x_{ji}(t_i) - log \sum_{l \in R(t_i)} exp(\sum_{j=1}^{p} \beta_j x_{jl}(t_i))\},$$

The Newton-Raphson procedure can be used to get estimate of β and its variance. For the time-dependent covariates, we have to know their values at each death time. It's very likely that the covariates were not measured at those death times.

(e) The partial likelihood depends the rank of survival times and values of covariates **only at event times**.

(f) In real life, since covariates may not be measured at the event times, thus, approximation to time-dependent covariates is needed. Options are

   i. The last recorded value

   ii. The value clsoest to that time

   iii. Linear interpolation between consecutive values

(g) Illustration: page 254.

4. **Programming**

   (a) In SAS PHREG procedure, in order to incorporate time-dependent covariates, programming statements are needed (*data step*) inside the procedure, which are used to create or modify the values of the explanatory variables in the **MODEL** statement.

(b) A dynamic assignment (on-the-fly) at each event time; can not be performed by a simple modification of the data.

(c) Programming statements can be avoid by using *counting process style of input*. (require program skills in data step).

(d) In Splus (**coxph()**), you have to use counting process style of input if there are time-dependent covariates.

5. **Some applications: summary**

   (a) Evaluate the assumption of proportional hazards.

   (b) Take 'Real' time-dependent covariates into account: Examples

   Myocardial infarct: waiting time from the infarct to ICU;

   The Stanford heart transplant program: Time-dependent variables was used to model the effects of subjects transferring from one treatment group to another (used many many times in the literature, e.g. JASA, 1974, 74-80; JASA, 1977, 27-36).

   i. Patients suitable for transplant vs patients not suitable for transplant.

   ii. For those patients suitable for transplant, those survived the waiting time vs those did not.

   iii. The solution to this problem is to introduce a time-dependent varible $X_1(t)$.

   $$h_i(t) = exp\{\eta_i + \beta_1 x_{1i}(t)\}h_0(t),$$

   where

   $$X_1(t) = \begin{cases} 0 & \text{if } t \leq \text{transplant time} \\ 1 & \text{if } t > \text{transplant time} \end{cases} \tag{1}$$

   iv. The interpretation of the coefficient of a time-dependent covariate: see the description in the book (page 305) and also the example below.

   UCB transplant: One measure of performance for patients with Umbilical Cord Blood transplant is the time to neutrophil recovery ($ANC > 500$ for 3 consecutive days) after transplantation.

   (c) Piece-wise Cox model: a special case of time-dependent variable.

6. **Example**: *Cord blood transplantation in adult recipients* (ref: Laughlin, et al, NEJM, 2001).

   (a) The data:

Obs	EFS	EFSCEN	DTANC5
1	332	1	30
2	39	1	.
3	768	0	36
4	64	1	32
5	679	0	21
6	14	1	.
7	100	1	28
8	518	0	28
9	35	1	.
20	25	1	.
21	79	1	14
22	1554	0	32
23	28	1	.
24	1349	0	26
25	190	1	14
26	284	1	30
27	156	1	22
28	51	1	35
29	1288	0	24
...			
58	14	1	.
59	482	0	43
60	47	1	31
61	538	0	16
62	679	0	16
63	2	1	.
64	95	1	19
65	21	1	21
66	664	0	31
67	671	0	26
68	53	1	34

(b) The SAS output:

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	365.891	361.486
AIC	365.891	363.486
SBC	365.891	365.398

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	4.4052	1	0.0358
Score	4.6070	1	0.0318
Wald	4.4842	1	0.0342

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
anc5cat	1	-1.24299	0.58698	4.4842	0.0342

Analysis of Maximum Likelihood Estimates

Variable	Hazard Ratio	95% Hazard Ratio Confidence Limits	
anc5cat	0.289	0.091	0.912

(c) The definition of time-dependent variable:

$$anc5cat(t) = \begin{cases} 0 & \text{if } t < \text{DAY TO ANC 500} \\ 1 & \text{if } t \geq \text{DAY TO ANC 500} \end{cases} \quad (2)$$

(d) The interpretation:

i. Likelihood ratio: Difference of log-likelihood with and without time-dependent covariate(s): The significant value means that neutrophil recovery has an effect on event free survival.
ii. The Chi-Square test on the coefficient of anc5cat also means neutrophil recovery has an effect on EFS.
iii. The negative coefficient indicates the earlier the neutrophil recovery patients had, the longer the patients survived after transplantation.

(e) SAS program:

```
options ls = 80;
libname fu '/meta/d6/pxf/laughlin/moddata';
data w1;
        set fu.graft;
proc print;
        var efs efscen dtanc5;

/*time-dependent covariate of absolute neutrophil recovery (DTANC500)*
proc phreg;
        model efs * efscen(0) = anc5cat /rl;
                if dtanc5 = . or efs < dtanc5 then anc5cat = 0;
                else anc5cat = 1;
run;
```

7. **Comparison of treatment**

(a) If no treatment effect after taking account of time-dependent covariate(s), one explanation could be that time-dependent variable(s) has masked the treatment difference.

(b) Biological marker/surrogate: CD4 T-cells etc for HIV/AIDS patients.

(c) Comparison of the results of analysis that incorporates time-dependent variables with an analysis that uses baseline values alone.