PhUSE EU Connect 2018

Paper AS05

A case study of fitting time dependent covariate in cox model using SAS.

Pankaj Kumar, GCE Solutions, Amsterdam, Netherlands

ABSTRACT

The collection and analysis of time-to-event data has been increasing in clinical trials. In many cases time-to-event data is modelled using COX regression. The SAS procedure PROC PHREG performs regression analysis based on the Cox proportional hazards model and is popular for fitting models on time to event data, such as models with time independent factors, time dependent factors, delayed entry and recurrent events.

The model can encompass covariates that change over time and these time dependent covariates play a key role in helping us understand the impact in exploratory analysis. There are diverse ways of coding a model with time dependent covariates in SAS. The paper will discuss a case study with two different ways of input.

1) Counting process approach
2) Programming statement.

It is good to understand both approaches as one has certain benefits over other. Brief algorithm, SAS code and results from both the approaches will be discussed.

INTRODUCTION

There are situations where we need to consider time varying covariate in out model, this seems more appropriate as this utilizes more data, which make our analysis more robust.

Sas codes, snapshot of datasets and results from COX proportional hazard model using time dependent covariate as predictor is presented with counting process approach and programming statement. Below analysis is done based on dummy data, in which 9000 subjects with event stroke are modeled with time-varying change from baseline in log(hsCRP), baseline log(hsCRP) and treatment as explanatory variable. For each subject hsCRP is collected over multiple visits. Snapshots of datasets are presented in Table 1, Table2 and Table 3.

SAMPLE DATA

SUBJID	Subject identifier
AVAL	Time to event
TRT01PN	Treatment
CNSR	Censoring information, 0: Event and 1: censored
ADY	Analysis relative day
HSCRP	log(hsCRP)
LBASE	log(hsCRP) at baseline
LCHG	change from baseline in log(hsCRP)
TMLCRP	time varying change from Baseline in log(hsCRP)

Table 1 Snapshot of input data ADEP

SUBJID	TRT01PN	CNSR	AVAL	ADY	LHSCRP	LBASE	LCHG
117008	1	1	190	0	2.821	2.821	0
117008	1	1	190	14	2.092	2.821	-0.73
117008	1	1	190	42	2.303	2.821	-0.519
117008	1	1	190	118	1.526	2.821	-1.295
141007	4	1	182	0	1.082	1.082	0
141007	4	1	182	15	1.03	1.082	-0.052
141007	4	1	182	46	1.946	1.082	0.864
141007	4	1	182	94	1.589	1.082	0.507
150017	2	0	73	0	2.447	2.447	0
150017	2	0	73	15	0.588	2.447	-1.859
150017	2	0	73	43	0.588	2.447	-1.859
150017	2	0	73	100	1.386	2.447	-1.06
150017	2	0	73	747	1.435	2.447	-1.012
150017	2	0	73	926	0.788	2.447	-1.658
150017	2	0	73	1111	1.335	2.447	-1.112
150017	2	0	73	1265	0.336	2.447	-2.11
150017	2	0	73	1458	1.808	2.447	-0.638
112012	3	1	180	0	3.25	3.25	0
112012	3	1	180	21	2.14	3.25	-1.11
112012	3	1	180	55	3.6	3.25	0.35
112012	3	1	180	91	2.976	3.25	-0.275

Cox proportional hazard model, using Treatment, LBASE and LCHG as explanatory variables is used. We can see that LCHG is changing over time. This data can be handled in two ways for fitting time dependent covariate model,

- counting process approach
- programming statement approach

Counting process approach

In counting process approach, multiple records are created for each subject. We divide time in such intervals that all covariates remain constant for each record. One record for each distinct pattern of the time-dependent measurements. Below is sample sas code to create data in this form. This data used in proc phreg step.

```
/*counting process approach*/

data adeff;
      set adep;
      by subjid ady;
/*    STCHG  :     log(hsCRP) Change from Baseline at Start Time  */
/*    ENCHG  :     log(hsCRP) Change from Baseline at End Time    */
      retain stval enval stchg enchg;

/*Assign Zero to the Start Time/Change for the First Interval*/
      if first.subjid then
            do;
                  stval=0; enval= 0; stchg = 0; enchg=0;
            end;
/*Carry Over Retained Start Time/Change from the Previous End Time/Change*/
/*Record the Current ADY and CHG as End Time/Change for Next Record      */
```

```
        if ady ne . and ady < aval then
                do;
                        stval=enval; enval=ady; stchg=enchg; enchg=lchg;
                        tmlcrp=stchg; enent= 0 ; output;
                end;

/*Final Event/Censored Record           */
        if last.subjid then
                do;
                        stval=enval; enval=aval; stchg=enchg; enchg= . ;
                        tmlcrp= stchg;
                        event=(cnsr=0);     output;
                end;
        keep subjid trt01p aval stval enval event lbase tmlcrp;
run;

/*Cox regression model*/
ods output ParameterEstimates = est1;
proc phreg data=all;
        class trt01pn (ref='4');
        model (stval,enval)*event(0) = trt01pn lbase tmlcrp / ties=exact risklimits;
        output out= dfb1 dfbeta = dt1-dt5; /* DFBETA estimates*/
run;
```

Table 2 Snapshot of counting process approach input data: Start, End value and Time varying change in log(hsCRP)

SUBJID	TRT01PN	Event	AVAL	LHSCRP	Start Value	End Value	TMLCRP
117008	1	0	190	2.821	0	14	0
117008	1	0	190	2.821	14	42	-0.73
117008	1	0	190	2.821	42	118	-0.519
117008	1	0	190	2.821	118	190	-1.295
141007	4	0	182	1.082	0	15	0
141007	4	0	182	1.082	15	46	-0.052
141007	4	0	182	1.082	46	94	0.864
141007	4	0	182	1.082	94	182	0.507
150017	2	0	73	2.447	0	15	0
150017	2	0	73	2.447	15	43	-1.859
150017	2	1	73	2.447	43	73	-1.859
112012	3	0	180	3.25	0	21	0
112012	3	0	180	3.25	21	55	-1.11
112012	3	0	180	3.25	55	91	0.35
112012	3	0	180	3.25	91	180	-0.275

As shown in above dataset, time duration of each subject is divided in interval such as TMLCRP is constant over interval (start, end].
EVENT indicator variable has EVENT=1 only corresponding to interval when there is an event.

Programming statement approach

In programming statement approach, data has one record for each subject. The time-varying covariates are defined in programming statements that are part of the PHREG step.

```
/*SAS code: Programming statement*/

/*below piece of SAS code changes data from vertical format to horizontal format i.e.
one observation per subject*/

/* Maximum number of surrogate measurements per patient */
proc sql;
create table a1 as select (count(subjid)) as m1 from adep group by subjid;
select max(m1) into: number from a1;
quit;

proc sort data=adep;
      by subjid trt01pn ady;
run;

/*create 3 datasets: CHG: change from baseline hsCRP, DAY: analysis relative day*/
/*BASE: baseline hsCRP information      */

proc transpose data=adep out=chg prefix=chg;
      by subjid trt01pn paramcd;
      var lchg;
run;

proc transpose data=adep out=day prefix=day;
      by subjid trt01pn paramcd;
      var ady;
run;

proc sql;
      create table base as select distinct subjid,trt01pn,aval,cnsr,paramcd,
armgr2,qmitgr4,lbase from adep order by subjid, trt01pn;
quit;

/*Merge CHG,DAY and BASE datasets*/

data all;
      merge chg day base;
      by subjid trt01pn ;
run;

ods output ParameterEstimates=est1;
proc phreg data=all;
      class trt01pn (ref='4');
      model aval*cnsr(1) = trt01pn lbase tmlcrp / ties=exact risklimits;
      tmlcrp= 0 ;
      array day(*) day1 - day%left(&number);
      array chg(*) chg1 - chg%left(&number);
      do j=1 to &number;
      if aval > day[j] and day[j] ne . then tmlcrp=chg[j];
       end;
   run;
```

Table 3 Snapshot of programing style input data

Subject	chg1	chg2	chg3	chg4	chg5	chg6	day1	day2	day3	day4	day5	day6	AVAL	Censor
117008	0	-0.73	-0.519	-1.295	.	.	0	14	42	118	.	.	190	1
141007	0	-0.052	0.864	0.507	.	.	0	15	46	94	.	.	182	1
150017	0	-1.859	-1.859	.	.	.	0	15	43	73	.	.	73	0
112012	0	-1.11	0.35	-0.275	.	.	0	21	55	91	.	.	180	1

Time varying hsCRP is defined in the proc PHREG step using array DAY and CHG. Using DO loop, values of change from baseline hsCRP are assigned in variable TMLCRP.

Below is result of parameter estimates which is identical for both the methods.

Parameter	ClassVal0	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Lower Confidence Limit	95% Upper Confidence Limit
TRT01PN	1	1	-0.14188	0.17991	0.6219	0.4303	0.868	0.61	1.235
TRT01PN	2	1	0.06002	0.16798	0.1277	0.7209	1.062	0.764	1.476
TRT01PN	3	1	0.07587	0.1711	0.1966	0.6574	1.079	0.771	1.509
lbase		1	0.2635	0.08807	8.9512	0.0028	1.301	1.095	1.547
tmlcrp		1	-0.13961	0.06598	4.4778	0.0343	0.87	0.764	0.99

CONCLUSION

When time-varying covariates are present, we should consider taking them into account in survival modeling to improve the estimation. In this paper an example of the two methods SAS uses to handle a time-varying covariate in Cox regression is shown. Counting process and programming statement, both gives same result for handling time-varying covariate in cox regression.

The programming statements are faster to code, but we should be careful while coding. Counting process is preferable as it is certain benefits over programming statement methods. Dataset is created, and we can debug errors easily in counting process approach. Using counting process approach, we can create DFBETA estimates, and residuals but using programing statement we can't get these. Same approach can be extended to multiple time-dependent covariates. Counting process type of input may be preferable in case of BUT need to ensure that time intervals do not overlap and there are no intervals of zero length.

CONTACT INFORMATION

Pankaj Kumar
GCE solutions
Kerkstraat 16A,
Eindhoven 5611GJ
Work Phone: +31626447862
Email: Pankaj.kumar@gcesolutions.com