Time Dependent Covariates[1]

STA312 Fall 2023

[1]See last slide for copyright information.

Background Reading

- "Using Time Dependent Covariates and Time Dependent Coefficients in the Cox Model" by Terry Therneau, Cynthia Crowson and Elizabeth Atkinson (2018):
  https://cran.r-project.org/web/packages/survival/vignettes/timedep.pdf

- Chapter 8 in *Applied Survival Analysis Using R* by Dirk Moore

Time Dependent Covariates: The Idea

- In predicting the next asthma attack, air quality is important. But air quality varies from day to day.
- In predicting when a couple will have a child, income could be important. But income can vary over time. .
- In predicting when a consumer will buy a new car, major repairs could matter. These happen from time to time.

Types of time-dependent covariate

- Internal: Variables that relate to the individuals, and can only be measured when an individual is alive. For example, blood glucose level, number of cigarettes, marital status.
- External: Variables that can be determined independently of the individual. For example, air quality, inflation rate, drug dose (if pre-determined).

Model

- For individual i, we have time to event, a failure indicator, and a set of covariate values over time.

$$(t_i, \delta_i, \{\mathbf{x}_i(t), t \in (0, t_i]\})$$

- Proportional hazards assumption:

$$h(t) = h_0(t) e^{\mathbf{x}(t)^\top \boldsymbol{\beta}},$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$, and we are assuming e^{β_0} is part of the hazard function.

Partial Likelihood

$$\mathrm{PL}(\boldsymbol{\beta}) = \prod_{i=1}^{D} \left(\frac{e^{\mathbf{x}(t_{(i)})^\top \boldsymbol{\beta}}}{\sum_{j \in R_{(i)}} e^{\mathbf{x}(t_{(i),j})^\top \boldsymbol{\beta}}} \right)$$

- The covariate values are those in force at time $t_{(i)}$.
- Some covariates (like type of disease) will not change over time.
- The individuals in the risk set don't depend on time, but the values of their covariates at time $t_{(i)}$ have to be available.
- It's mostly a matter of data format.

The start-stop data format[2]

Multiple lines of data per case

subject	time1	time2	status	age	creatinine	. . .
1	0	15	0	25	1.3	
1	15	46	0	25	1.5	
1	46	73	0	25	1.4	
1	73	100	1	25	1.6	
2	0	21	0	34	1.2	
2	21	50	0	34	1.4	
2	50	85	1	34	1.7	

Intervals (time1, time2] are closed on the right.

[2] Example adapted from Therneau et al. (2018)

Time-dependent covariates can help with a big problem

- It may seem obvious, but future values should not be used to predict something that happened in the past.
- Can having kids help a marriage last longer?
- You'd better watch how you analyze the data, because some couples get divorced too soon to have a child.
- Almost any event that can't happen if you're dead will be less likely to happen for individuals who fail early.
- So it may seem to help.
- For example, a heart transplant ...

The Stanford Heart Study

Annals of Internal Medicine

```
> # aim stands fort for Annals of Internal Medicine
> # Time to event (death) is futime, delta = fustat
> dim(aim); head(aim)

[1] 103   7
  patient fustat surgery      age futime wait.time transplant
1       1      1       0 30.84463     49        NA          0
2       2      1       0 51.83573      5        NA          0
3       3      1       0 54.29706     15         0          1
4       4      1       0 40.26283     38        35          1
5       5      1       0 20.78576     17        NA          0
6       6      1       0 54.59548      2        NA          0
```

Original analysis

The **surgery** variable is an indicator for prior bypass surgery

```
> summary( coxph(Surv(futime,fustat)~age+surgery+transplant,data=aim) )

Call:
coxph(formula = Surv(futime, fustat) ~ age + surgery + transplant,
    data = aim)

  n= 103, number of events= 75

               coef exp(coef) se(coef)      z Pr(>|z|)
age         0.05889   1.06065  0.01505  3.913 9.12e-05 ***
surgery    -0.41902   0.65769  0.37118 -1.129    0.259
transplant -1.71711   0.17958  0.27853 -6.165 7.05e-10 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

           exp(coef) exp(-coef) lower .95 upper .95
age           1.0607     0.9428    1.0298     1.092
surgery       0.6577     1.5205    0.3177     1.361
transplant    0.1796     5.5684    0.1040     0.310
```

Criticism

This was very embarrassing

- People who died on the wait list did not have a chance to get the surgery.
- Some of the “outcomes” were in the past.
- (Notice how much we want to say that the transplant *influenced* survival.)
- Solution: Treat transplant as a time-dependent covariate.

Re-format the data

```
> head(aim.ss2,40)

   id surgery      age tstart  tstop death transpl
1   1       0 30.84463      0   49.0     1       0
2   2       0 51.83573      0    5.0     1       0
3   3       0 54.29706      0   15.0     1       1
4   4       0 40.26283      0   35.0     0       0
5   4       0 40.26283     35   38.0     1       1
6   5       0 20.78576      0   17.0     1       0
7   6       0 54.59548      0    2.0     1       0
8   7       0 50.86927      0   50.0     0       0
9   7       0 50.86927     50  674.0     1       1

. . .

38 25       0 33.22382      0   24.0     0       0
39 25       0 33.22382     24 1799.0     0       1
40 26       0 30.53525      0 1400.0     0       0
```

Better Analysis

```
> betterheart = coxph(Surv(tstart,tstop,death) ~ age+surgery+transpl,
+ data=aim.ss2); summary(betterheart)

Call:
coxph(formula = Surv(tstart, tstop, death) ~ age + surgery +
    transpl, data = aim.ss2)

  n= 169, number of events= 75

            coef exp(coef) se(coef)      z Pr(>|z|)
age      0.03138   1.03187  0.01392  2.253   0.0242 *
surgery -0.77035   0.46285  0.35959 -2.142   0.0322 *
transpl -0.07894   0.92410  0.30608 -0.258   0.7965
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

        exp(coef) exp(-coef) lower .95 upper .95
age        1.0319     0.9691    1.0041    1.0604
surgery    0.4629     2.1605    0.2287    0.9365
transpl    0.9241     1.0821    0.5072    1.6836
```

Copyright Information