

# Applied Survival Analysis Using R

## Chapter 8: Time Dependent Covariates

Qi Guo

Department of Mathematical Sciences  
The University of Texas at Dallas

April, 16 2019

- 1 Introduction
- 2 Landmark
- 3 Model as a time dependent variable
- 4 Predictable Time Dependent Variables

# Introduction

- The values of the covariates must be determined at time  $t = 0$ , when the patient enters the study, and *remain constant* thereafter.
- Such data *evolve over time*, and it would be improper to use the value a covariate to model survival information that is observed before the covariate's value is known.
- *An intervention that occurs* after the start of the trial, or *a covariate (such as air pollution exposure) that changes* values over the course of the study are two examples of time dependent variables.

# Example

## Transplant Program

The study of the survival of patients who had been enrolled into the *transplant program* appeared to show that patients who received heart transplants *lived significantly longer* than those who did not.

```

1 > result.heart <- coxph(Surv(futime, fustat) ~ transplant + age
2   + surgery, data=jasa)
3 > summary(result.heart)
4     n= 103, number of events= 75
5           coef      exp(coef)    se(coef)      z      Pr(>|z|)
6 transplant -1.71711      0.17958    0.27853    -6.165    7.05e-10 ***
7 age         0.05889      1.06065    0.01505     3.913    9.12e-05 ***
8 surgery     -0.41902      0.65769    0.37118    -1.129    0.259
9 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

```

# Problem

- The key covariate is “`transplant`”, which takes the value 1 for those patients who received a heart transplant and 0 for those who did not.
- The estimate of the transplant coefficient is -1.717, and the p-value is very small. This result may appear to indicate that transplants are extremely effective in increasing the lifespan of the recipients.
- The problem here is that receipt of a transplant is a *time dependent covariate*; patients who received a transplant had to live long enough to receive that transplant.

# Fix

- Define a “landmark” time to *divide* patients into two groups, patients who receive the intervention prior to the landmark go into the intervention group and those who did not are placed in the comparison group.
- Approach:
  - Only patients who survive up to the landmark are included in the study.
  - All patients remain in their originally assigned group regardless of what happens in the future.
- For the heart transplant data, we may set a landmark at *30 days*.

# Outcome

```

1 > ind30 <- jasa$futime >= 30
2 > transplant30 <- {{jasa$transplant == 1}&{jasa$wait.time < 30}}
3 > summary(coxph(Surv(futime, fustat)~transplant30 + age
4 + surgery, data=jasa, subset=ind30))
5   n= 79, number of events= 52
6
7      coef      exp(coef)    se(coef)      z    Pr(>|z|)
8 transplant30TRUE -0.04214    0.95874    0.28377   -0.148    0.8820
9 age              0.03720    1.03790    0.01714    2.170    0.0300 *
10 surgery         -0.81966    0.44058    0.41297   -1.985    0.0472 *
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

```

- Conclusion: The coefficient for transplant30 (a true/false indicator for transplant within the first 30 days) is -0.042, and the p-value is 0.88, there is *little or no difference* in survival between those who got a transplant and those who did not.

# Modify the partial likelihood function

- We have no guidance as to when to set the landmark.
- Better way is to directly model the variable “transplant” as a time dependent variable, but important *adjustments are required to obtain unbiased estimates*.
- Modify *the partial likelihood function* to accommodate these types of variables

$$L(\beta) = \prod_{i=1}^D \frac{\psi_{ii}}{\sum_{k \in R_i} \psi_{ki}} \quad (1)$$

where  $\psi_{ki} = e^{z_k(t_i)\beta}$ , in Chapter 5 we fixed at time 0 so  $z_k(t_i) = z_k$  for all failure time  $t_i$ .

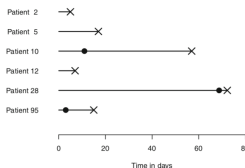


# Example

**Table 8.1** Sample of six patients from the Stanford heart transplant data set

id	wait.time	futime	fustat	transplant
2	—	5	1	0
5	—	17	1	0
10	11	57	1	1
12	—	7	1	0
28	70	71	1	1
95	1	15	1	1

**Fig. 8.1** Sample of six patients from the Stanford heart transplant data set. In this plot, death is denoted by an "X", and the time of transplant (for Patients 10, 28, and 95) by a *solid dot*. In the plot on the right, the timelines of patients who received a transplant are split into pre- and post-transplant components



- Patient #2 is the first to fail, at  $t = 5$ . At this time, all six patients are at risk, but only one, Patient #95, has had a transplant at this time, this is denominator, and numerator is 1 due to no-transplant died, so the partial likelihood function:

$$L(\beta) = \frac{1}{5 + e^{\beta}} \cdot \frac{1}{4 + e^{\beta}} \cdot \frac{e^{\beta}}{2 + 2e^{\beta}} \cdot \frac{1}{2 + e^{\beta}} \cdot \frac{e^{\beta}}{1 + e^{\beta}} \cdot \frac{e^{\beta}}{e^{\beta}} \quad (2)$$

# Transform the data

- If it has wait.time, divided into two parts, first 0-wait.time(tstart), second wait.time-futime(tstop)

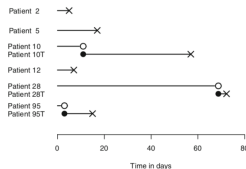
```

1 > sdata <- tmerge(heart.simple, heart.simple, id=id,
2 death=event(futime, fustat),transpl=tdc(wait.time))
3 > heart.simple.counting <- sdata[,-(2:5)]# drop columns 2 through 5
4 > heart.simple.counting
5   id   tstart   tstop   death   transpl
6 1  2      0      5      1      0
7 2  5      0     17      1      0
8 3 10      0     11      0      0
9 4 10     11     57      1      1
10 5 12      0      7      1      0
11 6 28      0     70      0      0
12 7 28     70     71      1      1
13 8 95      0      1      0      0
14 9 95      1     15      1      1

```

# Summary

**Fig. 8.2** Plot of sample of heart transplant patients in start-stop counting process format



- Use the `coxph` function as we did with *left-truncated* data:

```
1 > summary(coxph(Surv(tstart, tstop, death) ~ transpl,
2 data=heart.simple.counting))
3   n= 9, number of events= 6
4           coef    exp(coef)    se(coef)      z    Pr(>|z|)
5 transpl 0.2846    1.3292    0.9609    0.296    0.767
```

# Application

- Define “tdata” as a temporary data set, leaving off the dates and transplant-specific covariates. Also, we add 0.5 to the death time on day 0, and break a tied transplant time.

```

1 > tdata <- jasa[, -c(1:4, 11:14)]
2 > tdata$futime <- pmax(.5, tdata$futime)
3 > indx <- {{tdata$wait.time == tdata$futime} &
4 !is.na(tdata$wait.time)}
5 > tdata$wait.time[indx] <- tdata$wait.time[indx] - .5
6 > id <- 1:nrow(tdata)
7 > tdata$id <- id
8 > sdata <- tmerge(tdata, tdata, id=id,
9 death = event(futime, fustat), trans= tdc(wait, time))
10 > jasa.counting <- sdata[, c(7:11, 2:3)]
11 > head(jasa.counting)
12   id   tstart   tstop   death   trans   suregery   age
13 1 1      0      49      1      0      0      30.84463
14 2 2      0       5      1      0      0      51.83573
15 3 3      0      15      1      1      0      54.29706
16 4 4      0      35      0      0      0      40.26283
17 5 4     35      38      1      1      0      40.26283
18 6 5      0      17      1      0      0      20.78576

```

# Final Conclusion

```

1 > summary(coxph(Surv(tstart, tstop, death) ~ trans + surgery +
2 age, data=jasa.counting))
3     n= 170, number of events= 75
4           coef      exp(coef)      se(coef)      z      Pr(>|z|)
5 trans      0.01405      1.01415      0.30822      0.046      0.9636
6 surgery    -0.77326      0.46150      0.35966     -2.150      0.0316 *
7 age        0.03055      1.03103      0.01389      2.199      0.0279 *
8 ---
9 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

```

- Conclusion: As with the landmark analysis given earlier, that there is no evidence that receiving a heart transplant increases survival.
- This method is valid even though (unlike with the landmark method) no data are discarded.

# Recall

- Consider again the pancreatic data in Chapter 4, define a numerical (0/1) group variable, and fit the following model using the “pancreatic2” data in the “asaur” package:

```

1 > stage.n <- rep(0, nrow(pancreatic2))
2 > stage.n[pancreatic2$stage == "M"] <- 1
3 > result.panc <- coxph(Surv(pfs) ~ stage.n)
4 > result.panc
5           coef      exp(coef)    se(coef)      z      p
6 stage.n    0.593         1.81      0.401    1.48  0.14
7
8 Likelihood ratio test=2.43 on 1 df, p=0.119

```

- It shows little evidence of a group difference, and we have a plot of *Schoenfeld residuals* indicated that the hazard ratio appears not to be constant.

# Using the Time Transfer Function

- One way of dealing with this was to use *the Prentice modification* of the Wilcoxon test (using “rho = 1” in the “survdiffr” function)
- An alternative is to accommodate the changing hazard ratio by defining a time dependent covariate:

$$g(t) = z \cdot \log(t) \quad (3)$$

- Other time dependent function may not work, and the “time transfer” function “tt” can do this.

```

1 > result.panc2.tt <- coxph(Surv(pfs) ~ stage.n + tt(stage.n),
2 tt=function(x,t, ...) x*log(t))
3 > result.panc2.tt
4           coef      exp(coef)    se(coef)      z      p
5 stage.n      6.01      407.339      3.060     1.96  0.050
6 tt(stage.n)  -1.09       0.338       0.589    -1.84  0.065
7
8 Likelihood ratio test=6.33 on 2 df, p=0.0423

```

# Short introduction

- We know that the age of a patient increases in lock step with time itself, so can't we treat increasing age as a time dependent variable? The answer is **yes**, but doing so has no effect on the model.