Time-dependent covariates

In many situations it is useful to consider covariates that change over time. These are called "time-dependent" covariates. Such are of two kinds:

1. Internal variables
   These are related to each patient and are measurable while the patient is under observation
2. External variables
   These are variables that do not depend on the physical observation of the patient such as
   (a) Variables such as age that are known once the birth date or age at enrollment to the study is known
   (b) Variables that are independent of any individual like levels of pollution or temperature

These time-updated or dependent variables can be entered into the Cox model in direct extension of the simpler non-time-updated case

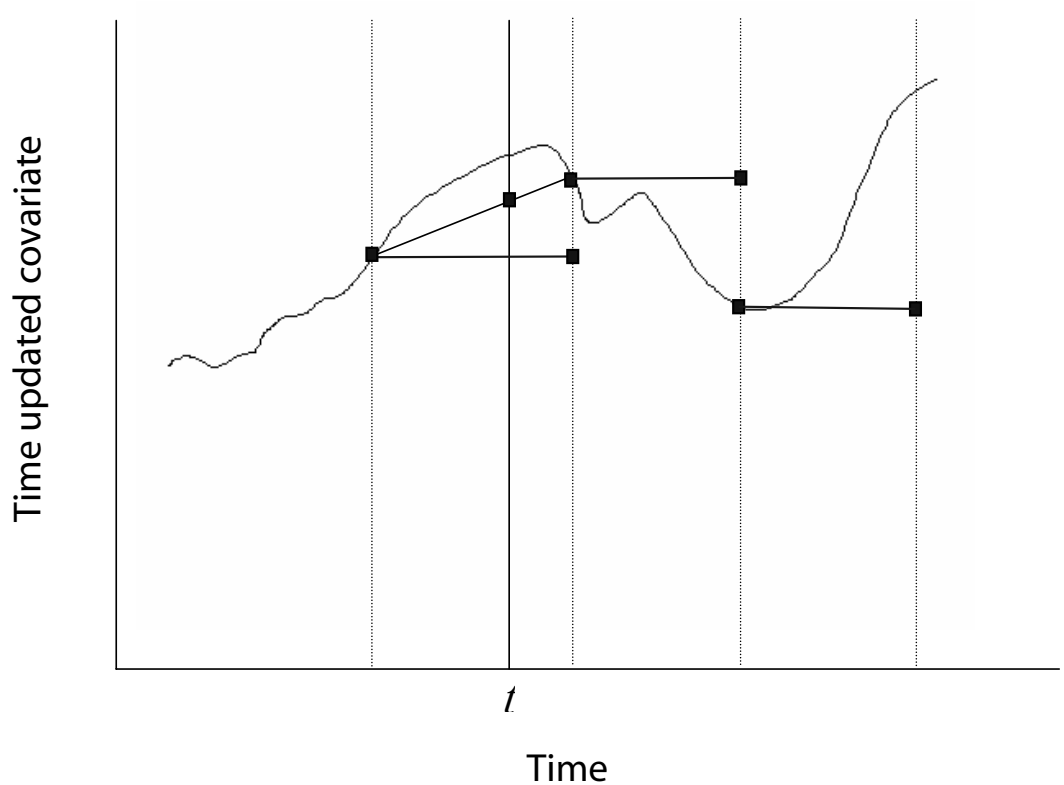
$$\lambda_i(t; \mathbf{Z}_i) = \lambda_0(t) \exp \sum_{j=1}^{n} \beta_j Z_{ij}(t)$$

where $\lambda_0(t)$ is the baseline hazard associated with all covariates being equal to zero *during all time points* t. So the Cox model is generalized as

$$\sum_{i=1}^{n} \delta_i \left\{ \sum_{j=1}^{p} \beta_j Z_{ij}(t_i) - \log \sum_{l \in R(t_i)} \exp \left(\sum_{j=1}^{p} \beta_j Z_{jl}(t_i) \right) \right\}$$

this means that we will need to have all the variable (especially internal ones) available at each event time. It is important to understand that this is no longer a proportional hazards model.

When the value of a time-updated covariate is not known during a failure time t we can use various methods to fill in a value for a particular time (see figure below). We can either extend the most recent value or, if two values are available on either side of the time point we can use interpolation.



The Stanford heart transplant data

We present here the famous Stanford heart transplant data set (Crowley & Hu, 1977). In this data set, 103 individuals waiting for a heart transplant were followed for survival. The problem that the study presented to the original investigators (and us) is that the effect of heart transplantation on survival is impossible to assess given the methods that we have been exposed to.

The reason is that the hazard of an individual is different before and after a transplantation and, for an individual to receive a transplant, they have to have survived up to the point that an organ is available. As Collett describes the situation (Section 7.3), the two groups are also not comparable at the time origin (entry into the study and time from transplantation).

Before considering the correct analysis, let's perform a naive analysis involving a conventional PH model

```
. stcox transplant
         failure _d:  fail
   analysis time _t:  survtime
Iteration 0:   log likelihood = -298.32561
Iteration 1:   log likelihood = -287.83084
Iteration 2:   log likelihood = -285.46286
Iteration 3:   log likelihood = -285.46262
Refining estimates:
Iteration 0:   log likelihood = -285.46262

Cox regression -- Breslow method for ties

No. of subjects =          103                     Number of obs   =       103
No. of failures =           75
Time at risk    =        31948
                                                   LR chi2(1)      =     25.73
Log likelihood  =   -285.46262                     Prob > chi2     =    0.0000
------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  transplant |   .2675782   .0652942    -5.40   0.000     .1658599    .4316781
------------------------------------------------------------------------------
```

This analysis indicates that transplantation is associated with one quarter of the hazard compared to no transplantation.
Given the misgivings about the appropriateness of the comparison, the solution is to introduce a time-updated covariate $Z(t)$ so that

$$Z(t) = \begin{cases} 1, \text{if } t > T_o \\ 0, \text{if } t \leq T_o \end{cases}$$

where T_o is the time of transplantation.

Crowley and Hu suggest that the hazard associated with this situation is

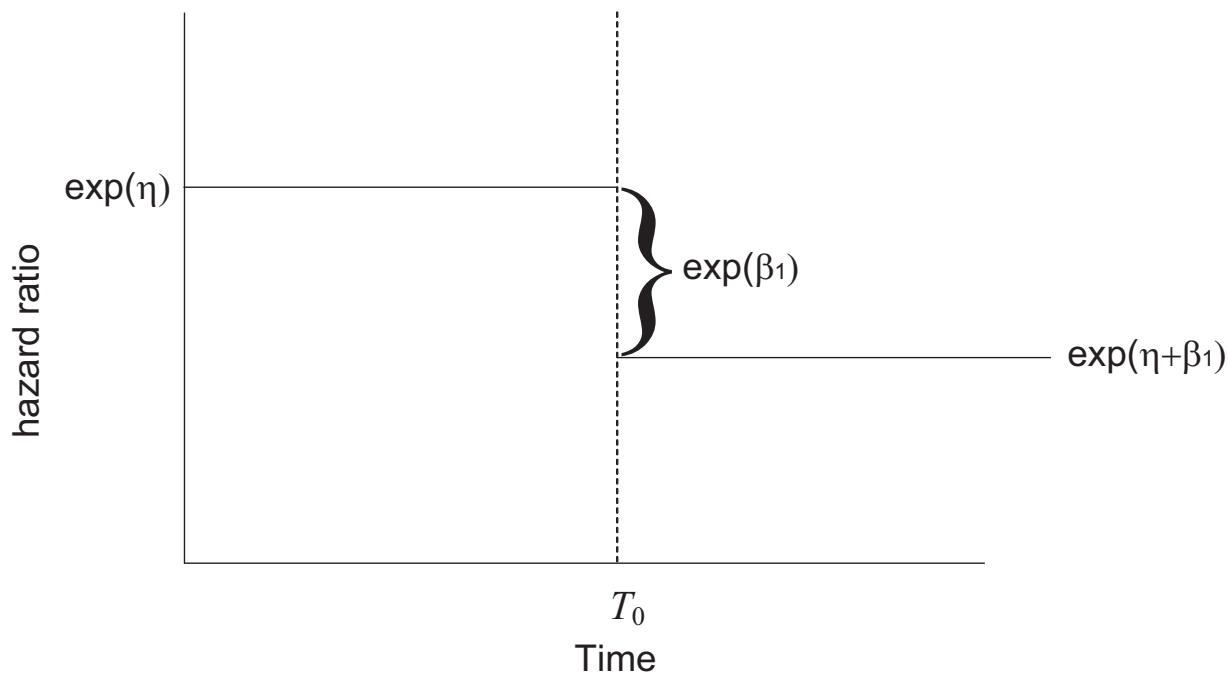
$$\lambda_i(t_i; \mathbf{Z}_i) = \lambda_0(t) \exp\{\eta_i + \beta_1 Z_{1i}(t)\}$$

where η_i is the summation of the products of all other covariates and their associated coefficients (excluding $Z_{1i}(t)$) measured on each individual i at each time t.

The hazard ratio (according to Crowley and Hu, 1977) is

$$\frac{\lambda(t_i; \mathbf{Z}_1(t))}{\lambda(t_i; \mathbf{Z}_0(t))} = \begin{cases} \exp\{\eta_i\}, \text{ before tranplantation} \\ \exp\{\eta_i + \beta_1\}, \text{ after tranplantation} \end{cases}$$

If $\beta_1 < 0$ then the hazard ratio of two individuals (one without a transplant and one with one) looks as follows (where T_0 is the time of transplantation:



In the original analysis, the effect of transplantation on the hazard is assessed by testing the significance of the coefficient β_1.

The null hypothesis $H_0 : \beta_1 = 0$ suggests that there is no effect on survival resulting from transplantation. On the other hand, the alternative hypothesis $H_A : \beta_1 < 0$ suggests a beneficial effect of the transplantation, while the alternative $H_A : \beta_1 > 0$ suggests a detrimental effect (increase in hazard of death) conferred by tranplantation.

Cox & Oakes' reanalysis of the heart transplant data

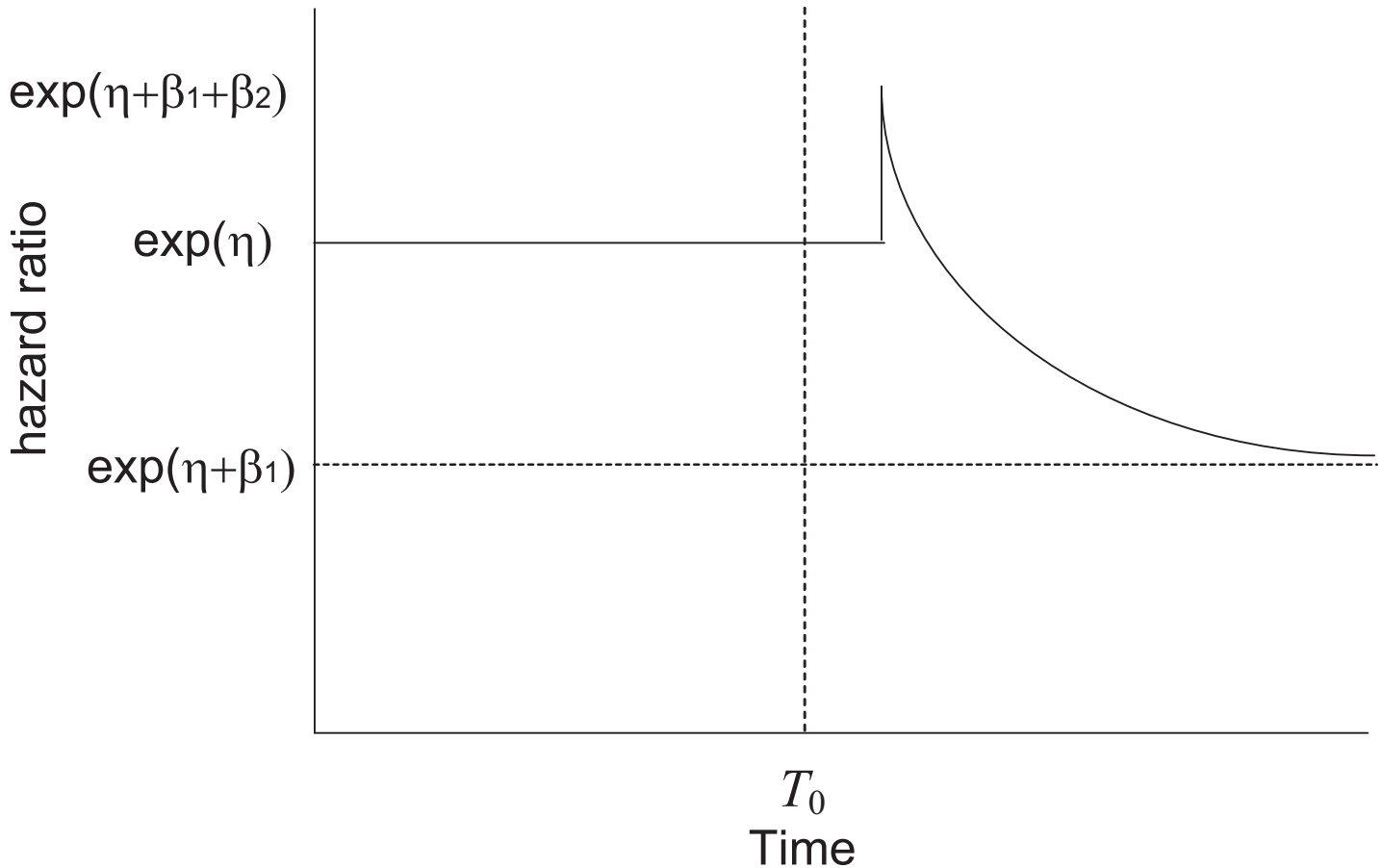
The previous model does not account for the fact that a heart tranplantation is a delicate and very dangerous operation. Thus, even if the hazard is ultimately reduced from the pre-transplant levels, a period of very high hazard is likely to follow the operation. Cox and Oakes (1984) improve on the analysis of Crowley and Hu by introducing factors β_2 and β_3 as follows:

$$\lambda_i(t_i; \mathbf{Z}_i) = \lambda_0(t) \exp\left\{\eta_i + \beta_1 + \beta_2 \exp[-\beta_3(t - T_0)]\right\}$$

The hazard ratio is

$$\frac{\lambda(t_i; \mathbf{Z}_1(t))}{\lambda(t_i; \mathbf{Z}_0(t))} = \begin{cases} \exp\{\eta_i\}, \text{ before tranplantation} \\ \exp\{\eta_i + \beta_1 + \beta_2\}, \text{ right at tranplantation} \\ \exp\{\eta_i + \beta_1\}, \text{ asymptotically (i.e., at } t \to \infty) \end{cases}$$

The Cox & Oakes reanalysis results in a hazard ratio that looks graphically as follows:



Notes

- In the reanalysis of Cox & Oakes, the effect of transplantation on the hazard is assessed by a more complex procedure.
- A large positive β_3 suggests a steep decrease of the hazard from an original level, just after transplantation, of $\exp(\eta_i + \beta_1 + \beta_2)$ to a level $\exp(\eta_i + \beta_1)$. A large positive value of β_2 suggests a large temporary increase of the hazard ratio post-transplantation. Conversely a smaller value of β_2 suggests small or negligible such increases.
- The latter asymptote ($\exp(\eta_i + \beta_1)$) depends on the magnitude and sign of β_1. The previous comments apply. That is, a large negative β_1 suggests a significant survival decrease (eventually) post tranplantation.
- Note that the Cox & Oakes model is equivalent to the Crowley & Hu model if $\beta_2 = 0$. The disadvantage of this model is that it requires specialized software to fit it.

To perform any analysis involving the time-updated transplant status, we need to create two lines (one pre-transplantation and one post-transplantation) for the patients that received a transplant. Thus, the line for patient 95 for example in the original data set is

patid	year	age	fail	survtime	priorsurg	transplant	waitime	missallele	missantigen	misscore
95	73	40	1	16	0	1	2	0	0	0.00

where waitime is the waiting time to transplantation. The data for this patient will be recoded as follows:

patid	year	age	fail	survtime	priorsurg	transplant	waitime	missallele	missantigen	misscore
95	73	40	0	2	0	0	2	0	0	0.00
95	73	40	1	16	0	1	2	0	0	0.00

in other words, we introduce a first (pre-transplantation) line with time going from $[0, 2)$ (i.e., just prior to transplantation.

During that time $\delta_i = 0$ since no failure has occurred, and $Z_1(t) = 0$ since a transplantation has not taken place. The second line (post-transplantation) is associated with the time interval $[2, 16)$ i.e, the 14 months of post-transplantation survival. During that time δ_i is set to whatever the failure status of the patient is (in this case the patient died under observation, so $\delta_i = 1$. Also $Z_i(t) = 1$ here. A situation arises with patient 38, who died on the same day of the transplantation (so waitime=survtime).

patid	year	age	fail	survtime	priorsurg	transplant	waitime	missallele	missantigen	misscore
38	70	41	1	5	0	1	5	3	0	0.87

Since this would cause most statistical software to exclude this case from consideration, we add a small fraction to the survival time (i.e., we assume that the patient lived a short time after receiving transplantation). This patient's data will look as follows:

patid	year	age	fail	survtime	priorsurg	transplant	waitime	missallele	missantigen	misscore
38	70	41	0	5	0	0	5	3	0	0.87
38	70	41	1	5.1	0	1	5	3	0	0.87

The STATA analysis of the Crowley & Hu model with only transplantation as the covariate is as follows:

```
. stcox tx, nohr
         failure _d:  censor
   analysis time _t:  time
                 id:  patid
Iteration 0:   log likelihood = -298.14428
Iteration 1:   log likelihood = -298.07974
Iteration 2:   log likelihood = -298.07974
Refining estimates:
Iteration 0:   log likelihood = -298.07974
Cox regression -- Breslow method for ties

No. of subjects =           99                     Number of obs   =       172
No. of failures =           75
Time at risk    =      31930.1
                                                   LR chi2(1)      =      0.13
Log likelihood  =   -298.07974                     Prob > chi2     =    0.7194

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          tx |   .1068495   .2984287     0.36   0.720      -.47806    .6917591
------------------------------------------------------------------------------
```

The estimate of β_1 is $\hat{\beta}_1 = 0.1068$ associated with a hazard ratio of of $e^(\beta_1) = 1.113$.

The interpretation is that transplantation increases slightly the hazard for death (by about 11%), an increase that is not statistically significant (p=0.720).

The log hazard ratio represented by β_1 concerns the comparison between a person that has undergone transplantation and one that has not .