Survival Analysis, Lecture 8 Alternative models for survival data

Ruth Keogh

Department of Medical Statistics
London School of Hygiene and Tropical Medicine

Overview

- So far we have focused on proportional hazards models
  - exponential and Weibull
  - Cox proportional hazards model
- This model is not always appropriate
- Sometimes direct extensions are possible
- Sometimes it is useful to model the impact of explanatory variables on survival in another way
  - accelerated failure time models
  - Aalen's additive hazards model

Overview

Part 1

- Reminder of some direct extensions of the proportional hazards model
- Introduction to the accelerated failure time model

Part 2

- More on the accelerated failure time model
- Parametric accelerated failure time models

Part 3

- Aalen's additive hazards model

Quick revision

Reminder: Cox proportional hazards model

$$h(t|x) = h_0(t)e^{\beta x}$$

- X could be a vector
- Two ways of directly extending this model to accommodate non-proportional hazards
  - the stratified Cox model
  - extended Cox model incorporating time-interactions

The stratified Cox model

The stratified Cox proportional hazards model

- Sometimes the proportional hazards assumption will hold for some explanatory variables but not others
- Instead of assuming the proportional hazards assumption holds overall for all explanatory variables, we can assume that it holds for some variables within groups or strata defined by other variables

Standard Cox proportional hazards model

$$h(t|SEX, TRT) = h_0(t) \exp\{\beta_{\text{sex}} SEX + \beta_{\text{trt}} TRT\}$$

Cox proportional hazards model stratified by sex

$$h(t|SEX, TRT) = \begin{cases} h_{0M}(t) \exp\{\beta_{\text{trt}} TRT\} & \text{for men} \\ h_{0F}(t) \exp\{\beta_{\text{trt}} TRT\} & \text{for women} \end{cases}$$

Stratified Cox model - more generally

Stratified Cox proportional hazards model: definition

$$h(t|s,x) = h_{0s}(t)e^{\beta x}$$

Cox model with time-interactions

Standard Cox proportional hazards model

$$h(t|x) = h_0(t)e^{\beta x}$$

Cox model with time-varying effects

$$h(t|x) = h_0(t)e^{\beta(t)x}$$

...for example

$$h(t|x) = h_0(t)e^{\beta_0 x + \beta_1 tx}$$

Comments

Stratified Cox model

- We cannot estimate the effect on survival of the variables represented by the strata.
- But we can estimate interactions between main exposures and the stratifying variable(s).

Cox model with time-interactions

- Often a good option
- But we have to specify the functional form of the time-varying parameters
- It is more tricky to obtain estimates of survivor curves

Accelerated failure time models

Introduction

Reminder: proportional hazards assumption

$$h_1(t) = \psi_{PH} h_0(t)$$

- There is a different way of modelling the effect of explanatory variables on survival which is based on modelling the impact of explanatory variables on the event time itself

Accelerated failure time model

T_0 : event time in the control group

T_1 : event time in the treatment group

$$T_1 = \psi_{AFT} T_0$$

Introduction

Accelerated failure time model

T_0 : event time in the control group

T_1 : event time in the treatment group

$$T_1 = \psi_{AFT} T_0$$

Accelerated failure time model

$$S_0(t) = \Pr(T_0 > t)$$

$$S_1(t) = \Pr(T_1 > t) = S_0\left(\frac{t}{\psi_{AFT}}\right)$$

Example: Accelerated failure time model

Accelerated failure time model

$$T_1 = \psi_{AFT}\, T_0$$

$$S_1(t) = \Pr(T_1 > t) = S_0\left(\frac{t}{\psi_{AFT}}\right)$$

Example: $\psi_{AFT} = 2$

- event times in the treatment group tend to be two times the event times in the control group.
- So the event tends to happen later for the individuals in the treatment group.
- The survival probability at time t is the treatment group is what it was at time $t/2$ is the control group

Example: Accelerated failure time model

Contrast this with the interpretation of ψ_{PH}.

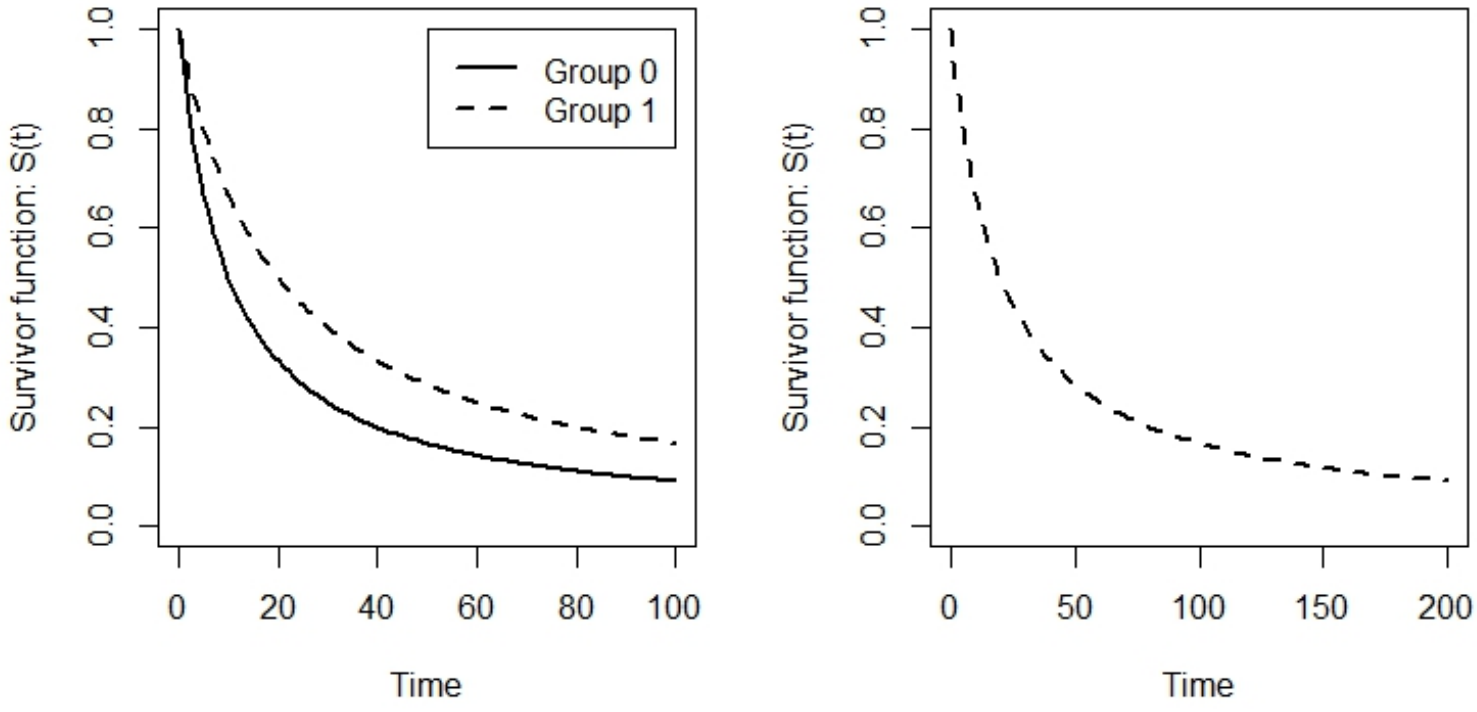
Reminder(again): proportional hazards assumption

$$h_1(t) = \psi_{PH} h_0(t)$$

Example: $\psi_{PH} = 2$

- The hazard at time t in the treatment group is two times the hazard at time t in the control group
- This is true at any time t

Example: Accelerated failure time model



Under accelerated life model, the time-to-event for individuals in the treatment group is speeded up or slowed down compared to individuals in the control group.

Why are we interested in AFT models?

- The AFT model can be useful when the proportional hazards model is not appropriate
- The AFT model provides a natural way of describing the effects of variables on disease progression
  - some people move faster along the progression path than others.

More on the accelerated failure time model

Proportional hazards model

In a proportional hazards model we usually write

$$h_1(t) = e^{\beta_{PH}} h_0(t)$$

instead of

$$h_1(t) = \psi_{PH} h_0(t)$$

Accelerated failure time model

In an accelerated failure time model it is convenient to do something similar and to write

$$T_1 = e^{-\beta_{AFT}} T_0$$

instead of

$$T_1 = \psi_{AFT} T_0$$

AFT model: the survivor function

Accelerated failure time model

$$T_1 = e^{-\beta_{AFT}} T_0$$

Survivor function in the control group

$$S_0(t) = \Pr(T_0 > t)$$

Survivor function in the treatment group

$$\begin{aligned} S_1(t) &= \Pr(T_1 > t) \\ &= \Pr(e^{-\beta_{AFT}} T_0 > t) \\ &= \Pr(T_0 > e^{\beta_{AFT}} t) \\ &= S_0(e^{\beta_{AFT}} t) \end{aligned}$$

AFT model: the survivor function

Survivor functions

Survivor function in the control group

$$S_0(t) = \Pr(T_0 > t)$$

Survivor function in the treatment group

$$S_1(t) = S_0(e^{\beta_{AFT}} t)$$

- If $e^{\beta_{AFT}} > 1$: survival times in the treatment group will tend to be earlier than survival times in the control group.
- If $e^{\beta_{AFT}} < 1$: survival times in the treatment group will tend to be later than survival times in the control group.
- $e^{\beta_{AFT}}$ has an interpretation as an acceleration factor.

AFT model: general formulation

Relation between survival times

$$T_x = T_0 e^{-\beta_{AFT} x}$$

Survivor function:

$$S(t|x) = S_0(te^{\beta_{AFT} x})$$

Hazard function:

$$h(t|x) = h_0(te^{\beta_{AFT} x})e^{\beta_{AFT} x}$$

Probability density function:

$$f(t|x) = f_0(te^{\beta_{AFT} x})e^{\beta_{AFT} x}$$

Fully parametric accelerated failure time models

Introduction

- So far we have discussed the AFT model in a general way
- In practice we need to have an underlying model and then incorporate the effects of exposures into that model
- Just like for PH models there are models for survival data which can be classified as AFT models
  - Weibull
  - Log-logistic
  - Others: Log-normal, Gamma, Inverse gamma

Oseltamivir treatment for influenza in adults: a meta-analysis of randomised controlled trials

Joanna Dobson, Richard J Whitley, Stuart Pocock, Arnold S Monto

Summary

Background Despite widespread use, questions remain about the efficacy of oseltamivir in the treatment of influenza. We aimed to do an individual patient data meta-analysis for all clinical trials comparing oseltamivir with placebo for treatment of seasonal influenza in adults regarding symptom alleviation, complications, and safety.

Methods We included all published and unpublished Roche-sponsored randomised placebo-controlled, double-blind trials of 75 mg twice a day oseltamivir in adults. Trials of oseltamivir for treatment of naturally occurring influenza-like illness in adults reporting at least one of the study outcomes were eligible. We also searched Medline, PubMed, Embase, the Cochrane Central Register of Controlled Trials, and the ClinicalTrials.gov trials register for other relevant trials published before Jan 1, 2014 (search last updated on Nov 27, 2014). We analysed intention-to-treat infected, intention-to-treat, and safety populations. The primary outcome was time to alleviation of all symptoms analysed with accelerated failure time methods. We used risk ratios and Mantel-Haenszel methods to work out complications, admittances to hospital, and safety outcomes.

Oseltamivir treatment for influenza in adults: a meta-analysis of randomised controlled trials

Joanna Dobson, Richard J Whitley, Stuart Pocock, Arnold S Monto

"We included data from nine trials including 4328 patients. In the intention-to-treat infected population, we noted a 21% shorter time to alleviation of all symptoms for oseltamivir versus placebo recipients (time ratio 0.79, 95% CI 0.74-0.85; $p < 0.0001$)."

T_0 : Time to alleviation of symptoms in placebo group

T_1 : Time to alleviation of symptoms in treatment group

$$T_1 = 0.79\,T_0$$

The AFT family of models

- What do we mean by saying a model is in the accelerated failure time family of models?
- We mean that after incorporating explanatory variables into a given model, the model with covariates is from the same distribution as the baseline model only with different parameters

Simple example: the exponential model is a PH model

$$h(t|x) = \lambda e^{\beta x}$$

$$h_0(t) = \lambda$$

The Weibull model is a PH model

$$h_0(t) = \kappa\lambda t^{\kappa-1}$$

$$h(t|x) = \kappa\lambda t^{\kappa-1} e^{\beta x}$$

Log-logistic distribution: not a PH model

$$h_0(t) = \frac{e^{\theta}\kappa t^{\kappa-1}}{1 + e^{\theta} t^{\kappa}}$$

$$h(t|x) = \frac{e^{\theta}\kappa t^{\kappa-1}}{1 + e^{\theta} t^{\kappa}} e^{\beta x}$$

Weibull model

Baseline survivor function under a Weibull model:

$$S_0(t) = exp\{-\lambda t^{\kappa}\}$$

Incorporating explanatory variables x using the acceleration factor $e^{\beta_{AFT}}$

$$S(t|x) = S_0(e^{\beta_{AFT} x} t) = \exp\left\{-\lambda e^{\kappa \beta_{AFT} x} t^{\kappa}\right\}$$

- $S(t|x)$ has the same form as $S_0(t)$ except that λ has been replaced by $\lambda e^{\kappa \beta_{AFT} x}$
- So $S(t|x)$ also has the form of a survivor function from a Weibull model

The Weibull model is therefore an accelerated failure time model

“But I thought the Weibull model was a proportional hazards model?”

- The Weibull model is both a proportional hazards model and an accelerated failure time model
- It is the only distribution for survival data with this property

Weibull model: Comparing PH and AFT parameterizations

Weibull model: PH

$$S(t|x) = \exp\left\{-\lambda e^{\beta_{PH} x} t^{\kappa}\right\}$$

Weibull model: AFT

$$S(t|x) = \exp\left\{-\lambda e^{\kappa \beta_{AFT} x} t^{\kappa}\right\}$$

Relationship between the hazard ratio and the acceleration factor under a Weibull model

$$\beta_{PH} = \kappa \beta_{AFT}$$

Example: Leukaemia patients data: Comparison of estimates from Weibull PH and Weibull AFT

- 42 individuals with leukaemia were followed up after diagnosis
- 21 individuals in a treatment group and 21 in a control group
- The event of interest was death

Example: Leukaemia patients data: Comparison of estimates from Weibull PH and Weibull AFT

```
 . streg i.group, distribution(weibull) nohr

Weibull PH regression

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       group |
  Treatment  |  -1.730872   .4130819    -4.19   0.000    -2.540497   -.9212463
       _cons |  -3.070704   .5580701    -5.50   0.000    -4.164501   -1.976907
-------------+----------------------------------------------------------------
      /ln_p |   .3117092   .1472919     2.12   0.034     .0230224     .600396
-------------+----------------------------------------------------------------
           p |   1.365757    .201165                      1.02329     1.82284
         1/p |   .7321944   .1078463                     .5485944    .9772406
------------------------------------------------------------------------------
```

Example: Leukaemia patients data: Comparison of estimates from Weibull PH and Weibull AFT

```
. streg i.group, distribution(weibull) time

Weibull AFT regression

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       group |
  Treatment  |   1.267335   .3106399     4.08   0.000     .6584916    1.876178
       _cons |   2.248352   .1659718    13.55   0.000     1.923054    2.573651
-------------+----------------------------------------------------------------
      /ln_p |   .3117092   .1472919     2.12   0.034     .0230224     .600396
-------------+----------------------------------------------------------------
           p |   1.365757    .201165                      1.02329     1.82284
         1/p |   .7321944   .1078463                     .5485944    .9772406
------------------------------------------------------------------------------
```

Example: Leukaemia patients data: Comparison of estimates from Weibull PH and Weibull AFT

Estimates of the effect of treatment on time to death

	Estimate	SE	95% CI
PH model			
β_{PH}	-1.731	0.413	(-2.54,-0.921)
κ	1.366	0.201	(1.023,1.823)
$\log \lambda$	-3.071	0.558	(-4.165,-1.977)
AFT model			
$-\beta_{AFT}$	1.267	0.311	(0.658,1.876)
κ	1.366	0.201	(1.023,1.823)
$\kappa^{-1} \log \lambda$	2.248	0.166	(1.923,2.574)

- What are the interpretations of $e^{\hat{\beta}_{PH}} = 0.18$ and $e^{-\hat{\beta}_{AFT}} = 3.55$?

Example: Leukaemia patients data: Comparison of estimates from Weibull PH and Weibull AFT

What are the interpretations of $e^{\hat{\beta}_{PH}} = 0.18$ and $e^{-\hat{\beta}_{AFT}} = 3.55$?

Proportional hazards model: $h_1(t) = e^{\beta_{PH}} h_0(t)$

- $e^{\hat{\beta}_{PH}} = 0.18$: The hazard in group 1 is 0.18 times the hazard in group 0, i.e. it is 82% lower.

AFT model: $T_1 = e^{-\beta_{AFT}} T_0$

AFT model: $S_1(t) = S_0(e^{\beta_{AFT}} t)$

- $e^{-\hat{\beta}_{AFT}} = 3.55$: The time to event in group 1 is expected to be 3.55 times the time to event in group 0.
- $e^{\hat{\beta}_{AFT}} = 0.28$: The probability of survival to time t in group 1 is equal to the probability of survival to time $0.28t$ in group 0.

Example: Leukaemia patients data: Comparison of estimates from Weibull PH and Weibull AFT

Estimates of the effect of treatment on time to death

	Estimate	SE	95% CI
PH model			
β_{PH}	-1.731	0.413	(-2.54,-0.921)
κ	1.366	0.201	(1.023,1.823)
$\log \lambda$	-3.071	0.558	(-4.165,-1.977)
AFT model			
$-\beta_{AFT}$	1.267	0.311	(0.658,1.876)
κ	1.366	0.201	(1.023,1.823)
$\kappa^{-1} \log \lambda$	2.248	0.166	(1.923,2.574)

- What is the relationship between the parameters β_{PH} and β_{AFT}?
- $\beta_{PH} = \kappa \beta_{AFT}$

The log-logistic model: an AFT model

Baseline survivor function:

$$S_0(t) = \frac{1}{1 + e^{\theta} t^{\kappa}}$$

Incorporating explanatory variables

$$S(t|x) = \frac{1}{1 + e^{\theta} e^{\kappa \beta_{AFT} x} t^{\kappa}}$$

- $S(t|x)$ has the same form as $S_0(t)$ except that e^{θ} has been replaced by $e^{\theta} e^{\kappa \beta_{AFT} x}$
- So $S(t|x)$ also has the form of a survivor function from a log-logistic model

The log-logistic model: an AFT model

```
. streg i.group, distribution(loglogistic)

Loglogistic AFT regression

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       group |
  Treatment  |   1.265463   .3256613     3.89   0.000       .627179    1.903748
       _cons |   1.892691   .2076196     9.12   0.000      1.485764    2.299618
-------------+----------------------------------------------------------------
    /lngamma |  -.6041234   .1501221    -4.02   0.000    -.8983572   -.3098896
-------------+----------------------------------------------------------------
       gamma |   .5465533   .0820497                      .4072381     .733528
------------------------------------------------------------------------------
```

Model comparison

- If one model is nested within another then we can compare the fit of the two models using a likelihood ratio test
- If two models are non-nested than we can compare the fit of the two models using the Akaike Information Criterion

$$AIC = -2 \times \log \text{likelihood} + 2p$$

A lower AIC indicates a better fitting model

Model comparison

```
streg i.group, distribution(weibull) time
estimates store A

streg i.group, distribution(loglogistic)
estimates store B

estimates table _all, stats(aic)
-------------------------------------
    Variable |     A             B
-------------+-----------------------
_t           |
       group |
  Treatment  |  1.2673347    1.2654634
       _cons |  2.2483522    1.8926906
-------------+-----------------------
       /ln_p |  .31170922
    /lngamma |             -.60412339
-------------+-----------------------
Statistics   |
         aic |   100.1282    102.29205
```

Aalen's additive hazards model

Aalen's additive hazards model

Definition for a single covariate x

$$h(t|x) = \beta_0(t) + \beta_1(t)x$$

- $\beta_0(t)$ is the baseline hazard (for people with $x = 0$)
- $\beta_1(t)$ is the excess hazard for each unit increase in x
- Covariates act on the hazard in an additive way

Aalen's additive hazards model

Odd Aalen. 1980. A model for non-parametric regression of life times. In W. Klonecki, A. Kozek, J Rosinski Eds). *Mathematical Statistics and Probability Theory*, Volume 2 of *Lecture Notes in Statistics*, pp 1-25. New York: Springer-Verlag.

Odd Aalen. A linear regression model for the analysis of life times. *Statistics in Medicine*, 1989; 8:907-925.

Aalen's additive hazards model

More general definition

$$h(t|x) = \beta_0(t) + \beta_1(t)x_1 + \beta_2(t)x_2 + \cdots + \beta_p(t)x_p$$

..and including time-dependent covariates

$$h(t|x) = \beta_0(t) + \beta_1(t)x_1(t) + \beta_2(t)x_2(t) + \cdots + \beta_p(t)x_p(t)$$

- A key feature of this model is that it is non-parametric
- We do not specify how the β coefficients depend on t
- In this way the model naturally allows for time-varying covariate effects
- i.e. the impact of x_1 on the hazard at time $t = 1$ can be different from the impact of x_1 on the hazard at time $t = 2$, for example

Aalen's additive hazards model: a few comments

- It has not been as commonly used as the Cox model
- A criticism is that it does not constrain the hazard to be non-negative.

Some advantages of the additive hazards model

- It handles time-varying effects of covariates more easily than the extended Cox model, in which we have to specify the functional form of time-varying parameters
- It may sometimes better describe the underlying relationships
- Additive hazards models are collapsible, unlike multiplicative hazard models such as the Cox model
- It has properties that make it attractive for use in causal inference and in studies of direct and indirect effects (mediation analysis).

Fitting the additive hazards model

More general definition

$$h(t|x) = \beta_0(t) + \beta_1(t)x_1 + \beta_2(t)x_2 + \cdots + \beta_p(t)x_p$$

- The additive hazards model is a linear model
- To fit the model we actually use a series of linear regressions
- ....one regression at each time that an event occurs
- The justification for this relies on counting processes

Fitting the additive hazards model

Model with covariate x

$$h(t|x) = \beta_0(t) + \beta_1(t)x$$

Cumulative regression coefficients

$$B_0(t) = \int_0^t \beta_0(u)du, \quad B_1(t) = \int_0^t \beta_1(u)du, \quad \text{etc}$$

- The results of fitting theadditive hazards model are presented graphically
- Using plots of the estimated cumulative regression coefficients against time
- This provides information about how the covariate is related to the hazard over time

Example: treatment programs for drug abuse

David Hosmer, Patrick Royston. Using Aalen's linear hazards model to investigate time-varying effects in the proportional hazards regression model. *The Stata Journal* 2002; 2: 331-350.

- Randomized trial of residential treatment programs for drug abuse: n=575
- The two trial arms are programs of different lengths (0: short, 1: long)
- Outcome: self-reported return to drug abuse
- Covariates: Treatment arm, Age, Beck depression score

Example: treatment programs for drug abuse

```
. stlh age_c beck_c treat, xlabel(0,90,180,270,377) l1title("Hazard") /*
*/ testwt(1 2 3 4) b1title(" ") b2title("Time") gen(uis)
```

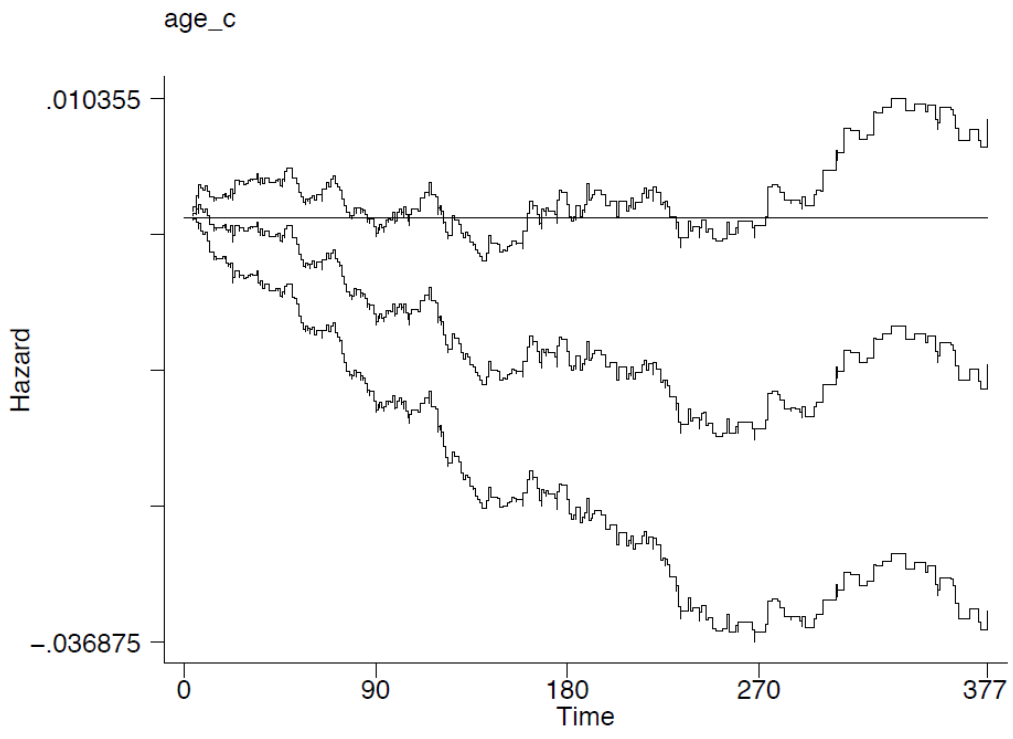


Figure 4: Plot of the estimated cumulative regression coefficient for centered age and for the pointwise 95 percent confidence bands.

Example: treatment programs for drug abuse

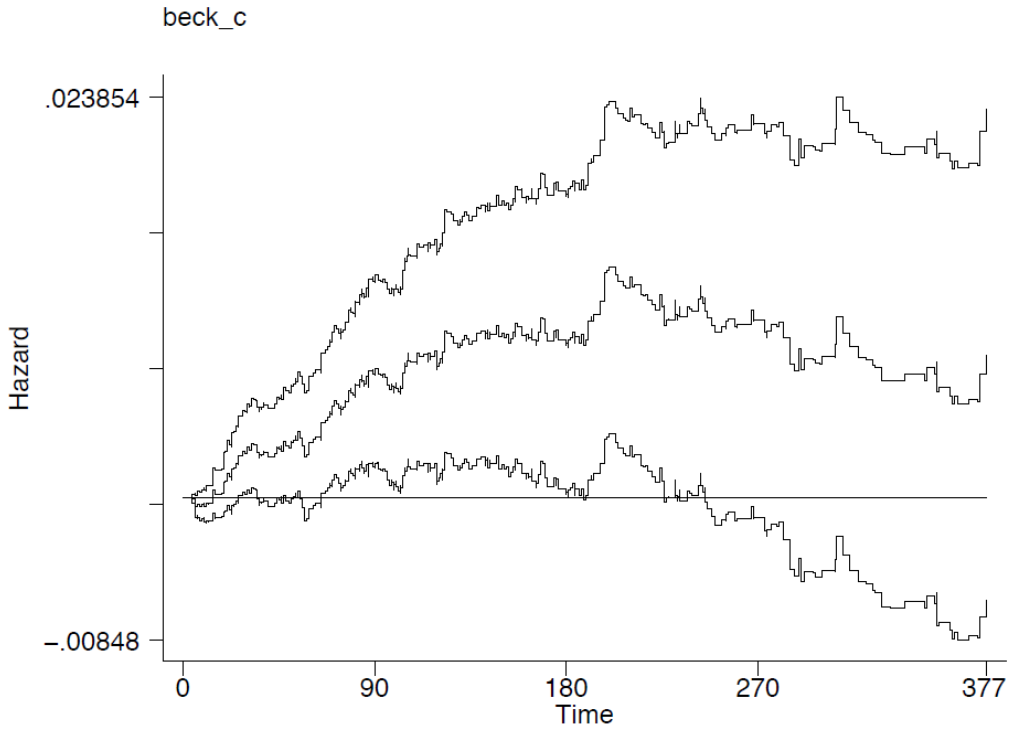


Figure 5: Plot of the estimated cumulative regression coefficient for centered Beck score and the pointwise 95 percent confidence bands.

Example: treatment programs for drug abuse

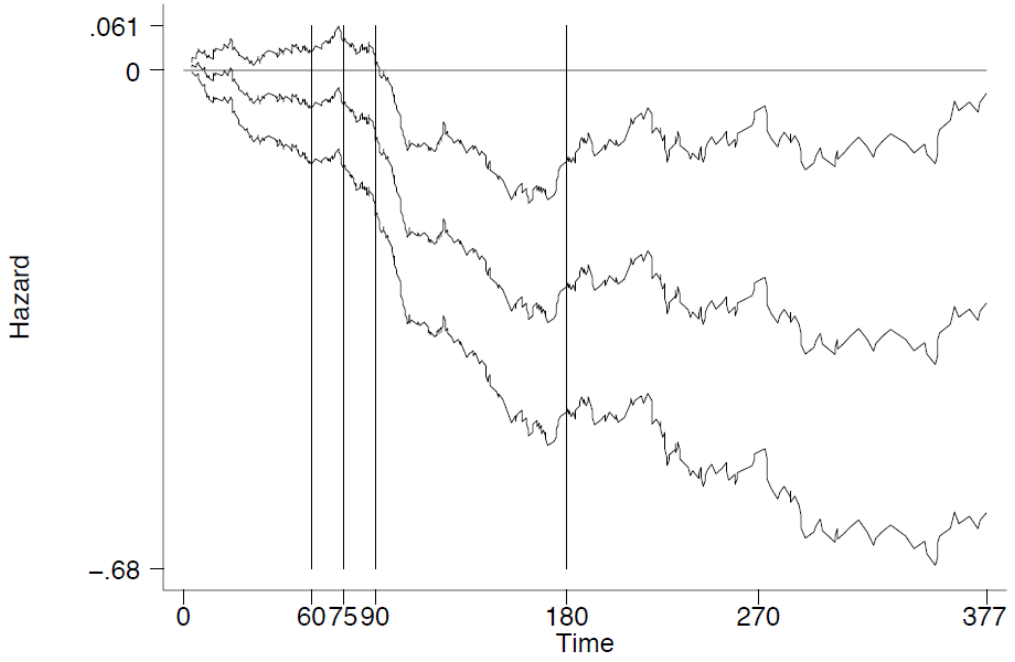


Figure 6: Plot of the estimated cumulative regression coefficient for treatment and the pointwise 95 percent confidence bands.

Concluding comments

- The proportional hazards model is not the only one available to us for survival data
- Sometimes an AFT model will provide a better fit and/or provide the most convenient way of explaining how an explanatory variables relates to time-to-event
- Sometimes an additive hazard model will provide a better fit
- The additive hazards model makes fewer assumptions than PH and AFT models because it naturally accommodates time-varying effects
- We should assess the fit of any model we use
- Sometimes the choice of model is based on what we're doing and how we want to communicate the results