

Department of Natural Sciences

<http://www.dina.kvl.dk/~torbenm/DINA/survival>

# Cox-regression part III and more specialised topics

Torben Martinussen

[torbenm@dina.kvl.dk](mailto:torbenm@dina.kvl.dk)

# Outline

- Cox-regression
- Delayed entry
- Time-dependent covariates
- Time-dependent regression effects
- Data with cluster effects

# Predictions based on Cox-model

Let  $X^0$  denote the covariates for a given subject.

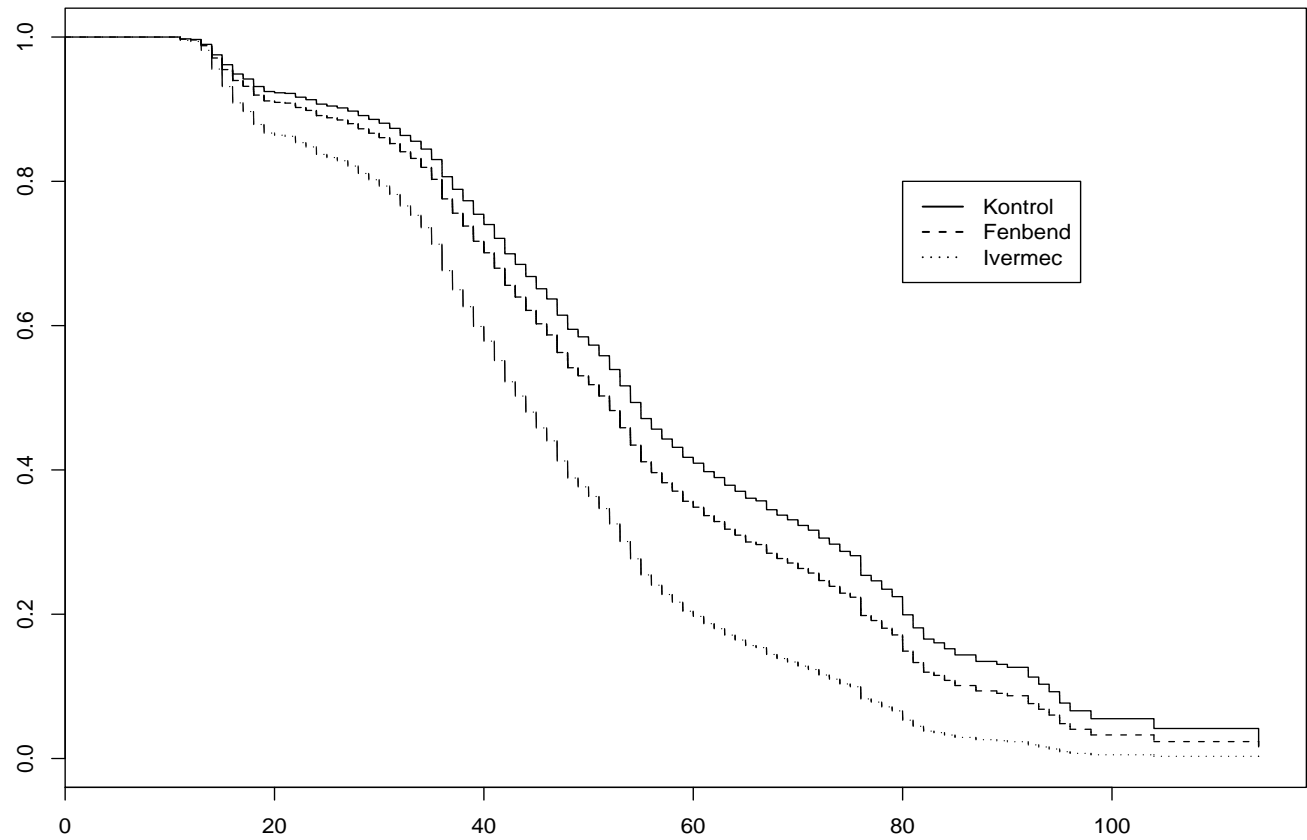
The survival function is given by

$$P(T^0 > t) = \exp\left(-\int_0^t \lambda(s, X^0) ds\right) = \exp(-\Lambda_0(t)e^{PI(\beta)}),$$

where  $PI(\beta) = \beta_1 X_1^0 + \dots + \beta_p X_p^0$  is called the prognostic index.

The survival function may thus be estimated by insertion of  $\hat{\beta}$ .

```
fit<-coxph(Surv(time,status==1)~treatF+treatI+cluster(pair))
temp<-data.frame(treatF=0,treatI=0)
plot(survfit(fit,newdata=temp),conf.int=F)
temp<-data.frame(treatF=1,treatI=0)
lines(survfit(fit,newdata=temp),conf.int=F,lty=2)
temp<-data.frame(treatF=0,treatI=1)
lines(survfit(fit,newdata=temp),conf.int=F,lty=3)
legend(80,0.8,c('Kontrol','Fenbend','Ivermec'),lty=c(1:3),lwd=1.5)
```



# Delayed Entry and Cox

On 1 July 1973, 1499 in Fyn suffered from insulin-dependent diabetes.

Survival status for these patients was assessed January 1982.

Purpose: Study age-specific mortality of diabetes, so time-scale is age.

The cross-sectional sampling introduces a potential *length bias*, longer survivals have a higher probability of being sampled.

Solution: Use **delayed entry** technique, i.e., subjects are only at risk at age of entry and onwards.

```
coxph( Surv(age_73, age_dead, status) ~ covariates)
```

# Time-dependent covariates

Some covariates may change their value with time:  $X(t)$ .

Example: Time for horses in optimal training.

$$X(t) = \begin{cases} 1 & \text{Time since last race is } < 14 \text{ days} \\ 0 & \text{Otherwise} \end{cases}$$

horse	lane	start	stop	status	race	....
1	2	0	27	0	0	
1	2	27	35	1	1	
2	1	0	55	1	0	
3	1	0	40	0	0	
3	1	40	54	0	1	
3	1	54	83	0	0	
		.				
		.				
		.				

In R:

```
coxph( Surv(start, stop, status) ~ factor(lane)+race)
```

# Time-dependent covariate effects

- Treatment is effective for some time, but then effect levels off.
- Takes some time before treatment has an effect.

This is a time-dependent covariate effect. This may actually be fitted within the Cox-model using the package `timereg` developed by Martinussen and Scheike. Model:

$$\lambda_i(t) = \lambda_0(t) \exp(\beta_1(t)X_{i1} + \cdots + \beta_p(t)X_{ip}).$$

In R:

```
fit.time<-timecox(Surv(time/365,status==1)~sex+log(thickness)+ulcer)
```

Many technical aspects with this model, however. Need some further practical experience before recommending it for general use.

# Other models for event time data

- Parametric models
- Aalen's additive hazard model:

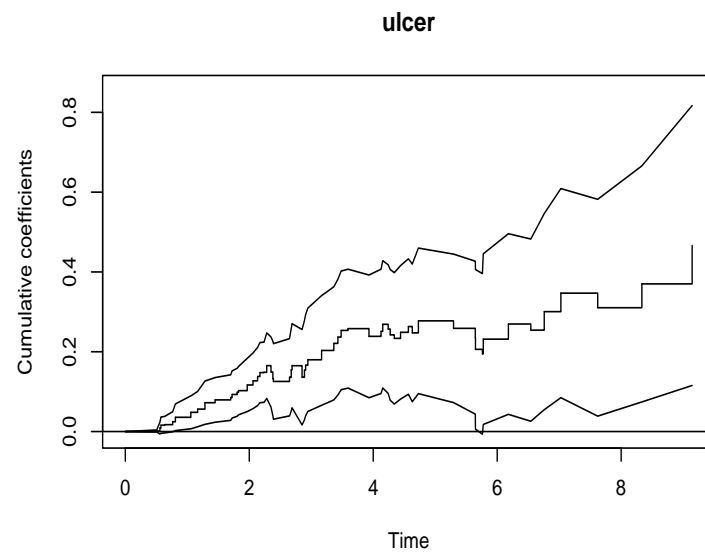
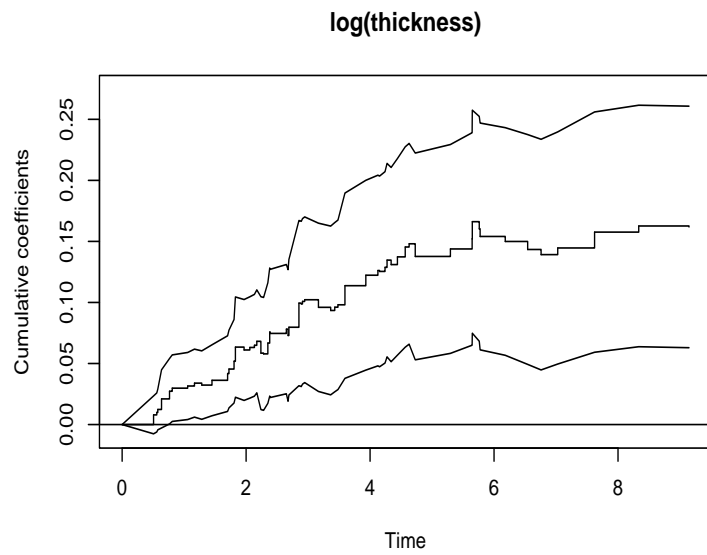
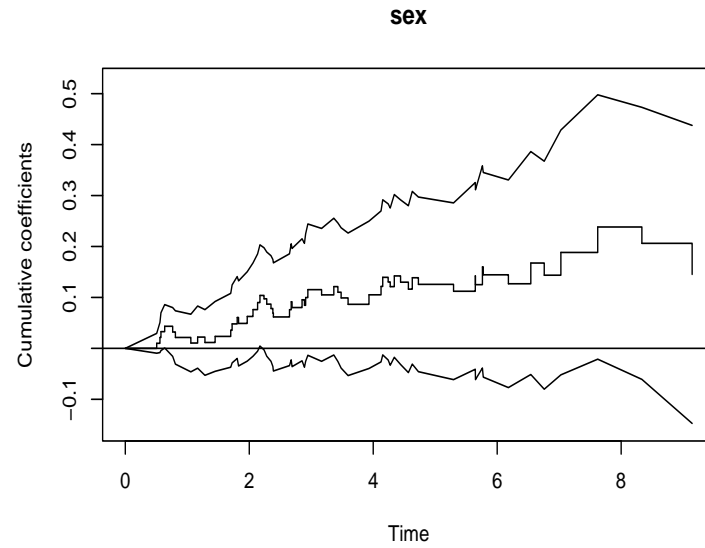
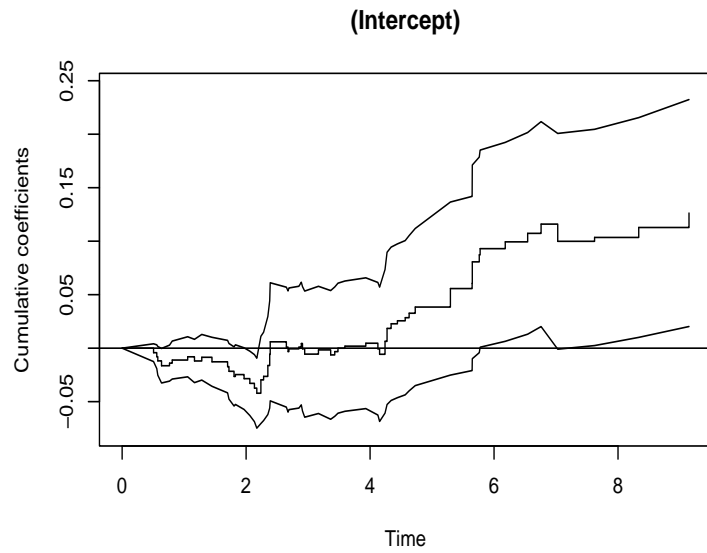
$$\lambda_i(t) = \beta_0(t) + \beta_1(t)X_{i1} + \cdots + \beta_p(t)X_{ip},$$

where time-dependent effects are incorporated. Models effects on additive scale instead on relative scale. One estimates  $B_j(t) = \int_0^t \beta_j(s) ds$ ,  $j = 0, \dots, p$ . In R using `timereg`:

```
fit.time<-aalen(Surv(time/365,status==1)~sex+log(thickness)+ulcer)
plot(fit.time)
```

and a lot more, see forthcoming book by Martinussen and Scheike.

- Goodness-of-fit tools.



# Cox-regression with clusters

Some variables play a special role, so-called cluster variables or random effects variables.

Examples:

- Worm-data: Cocoons come from same pair of worms. In total 47 pairs. Cluster=pair.
- Diabetic retinopathy. Time to blindness is observed on both eyes. Cluster=patient.

Often one is not interested the effect of the clusters but they may contribute variation that needs to be accounted for when judging for example treatment effect.

Treating data as being independent is cheating since they stem from from a smaller number of clusters.

# Worm-data

```
> kokoner[1:10,]
```

	treat	pair	age	time	status	kvart
1	K	1	5	15.99937	1	1
2	K	2	5	14.99687	1	1
3	K	1	6	14.99049	1	1
4	K	1	6	14.99950	1	1
5	K	3	6	13.99824	1	1
6	K	3	6	13.00182	1	1
7	K	3	6	17.00665	1	1
8	K	2	6	12.99521	1	1
9	K	2	6	10.99926	1	1
10	K	2	6	10.99527	1	1

# Diabetic Retinopathy Study

197 patients with "high-risk" diabetic retinopathy Each patient had one eye randomized to laser treatment and the other eye received no treatment. For each eye, the event of interest was the time from treatment to blindness

- laser treatment (Yes=1, no=0)
- treated eye: 1=right 2=left
- type of diabetes

id	time	status	trt_eye	treat	adult	age_dx
5	46.23	0	2	1	2	28
5	46.23	0	2	0	2	28
14	42.5	0	1	1	1	12
14	31.3	1	1	0	1	12
16	42.27	0	1	1	1	9
16	42.27	0	1	0	1	9
25	20.6	0	2	1	1	9
25	20.6	0	2	0	1	9

# Worm-data

```
> fit<-coxph(Surv(time,status==1)~factor(treat))
```

```
> fit
```

	coef	exp(coef)	se(coef)	z	p
factor(treat)1	0.166	1.18	0.0885	1.87	6.1e-02
factor(treat)2	0.598	1.82	0.0883	6.78	1.2e-11

Likelihood ratio test=47 on 2 df, p=6.17e-11 n= 894

```
> fit<-coxph(Surv(time,status==1)~factor(treat)+cluster(pair))
```

```
> fit
```

	coef	exp(coef)	se(coef)	robust se	z	p
factor(treat)1	0.166	1.18	0.0885	0.234	0.708	0.4800
factor(treat)2	0.598	1.82	0.0883	0.223	2.683	0.0073

Likelihood ratio test=47 on 2 df, p=6.17e-11 n= 894

# Random-effects models

Also called: Frailty models. Suppose again that we have clusters. For  $i$ th individual in  $k$ th cluster:

$$\lambda_{ik}(t) = Z_k \lambda_0(t) \exp(\beta_1 X_{i1} + \dots + \beta_p X_{ip}),$$

where  $Z_1, \dots, Z_K$  are independent **random variables**.

- Such models can to some extent be fitted in R
- May be of interest if one is interested in estimating the correlation within clusters, e.g., genetic effects.
- **Important** to notice, however, that the marginal model is not a Cox-model.
- If one is not interested in the cluster effect then use the `cluster`-option in `coxph`.

# Literature

- Andersen and Væth: Statistisk analyse af overlevelses data. [In danish, a bit old-fashioned].
- Collett: Modelling survival data in medical research. [Quite good, non-technical].
- Therneau and Grambsch: Modeling Survival Data: Extending the Cox Model. [Somewhat technical; R-code and SAS-code; not very precise mathematically].
- Andersen, Borgan, Keiding and Gill: Statistical models based on counting processes. [Technical, asymptotic results].
- Martinussen and Scheike: Dynamic regression models for survival data. [not yet published; modern approach; R-code].

# Exercise using R

We wish to build a model for the worm-data based on the variables `treatment`, `age.gr` and `pair`.

- (1) Do a stratified Cox-regression analysis with `age.gr` as stratification variable and `pair` as cluster variable.
- (2) Give relevant relative risks.
- (3) Compute relevant survival curves based on the fitted model. Plot them in one figure. Estimate median waiting times.
- (4) Summarize the results of the analysis.