More on Cox-regression

STK4080 H16

1. Repetition
2. Left truncation
3. Time-dependent covariates
4. Stratified Cox-regression
5. Residuals - Model check
6. How to handle departures from prop.haz. assumption

Repetition Cox-regression

With covariate x_i the survival time T_i has hazard

$$\alpha(t|x_i) = \exp(\beta' x_i)\alpha_0(t)$$

where $\alpha_0(t)$ is called basis (underlying) hazard and where β is a regressionsparameter.

With D_i indicator for death for individual i, $\tilde{T}_i$ right censoring time and $\mathcal{R}(t) = \{i : \tilde{T}_i \geq t\}$ = the risk set (right before) time t we estimate β by maximizing the Cox' partial likelihood

$$L(\beta) = \prod_{i=1}^{n} \left[\frac{\exp(\beta' x_i)}{\sum_{k \in \mathcal{R}(\tilde{T}_i)} \exp(\beta' x_k)} \right]^{D_i} = \prod_{i=1}^{n} \left[\frac{\exp(\beta' x_i)}{S^{(0)}(\beta, t)} \right]^{D_i}$$

where $S^{(0)}(\beta, t) = \sum_{k \in \mathcal{R}(\tilde{T}_i)} \exp(\beta' x_k)$

Rep. Cox-regr., contd.

In counting process notation we saw that we may express the score

$$U(\beta) = \frac{\partial \log(L(\beta))}{\partial \beta} = \sum_{i=1}^{n} \int \left[x_i - \frac{S^{(1)}(\beta, t)}{S^{(0)}(\beta, t)} \right] dN_i(t).$$

Also $L(\beta)$ may be treated as a regular likelihood. Thus

$$\hat{\beta} \sim \mathrm{N}(\beta, I(\hat{\beta})^{-1}).$$

where $I(\beta) = -\frac{\partial^2 \log(L(\beta))}{\partial \beta^2}$ and for two nested models with maximum log-partial likelihood respectively $l^\star$ and $\hat{l}$ we have

$$\mathrm{LRT} = 2(l^\star - \hat{l}) \sim \chi^2_q$$

if the $l^\star$-model has q more parameters than the $\hat{l}$-model and that $\hat{l}$-model is true (H_0-model).

Estimation of cumulative hazard $A_0(t) = \int_0^t \alpha_0(s)ds$

The common estimator for $A_0(t)$ is the Breslow-estimator

$$\hat{A}_0(t) = \sum_{\tilde{T}_i \leq t} \frac{D_i}{\sum_{k \in \mathcal{R}(\tilde{T}_i)} \exp(\hat{\beta}' x_k)} = \int_0^t \frac{dN_\bullet(s)}{S^{(0)}(\hat{\beta}, s)}$$

Note the similarity with the Nelson-Aalen estimator.

Given $\hat{A}_0(t)$ it is simple to estimate cumulative hazard for an individual with hazard $\alpha(t|x_i) = \exp(\beta' x_i)\alpha_0(t)$ as

$$\hat{A}(t|x_i) = \exp(\hat{\beta}' x_i)\hat{A}_0(t)$$

The survival function $S(t|x_i) = \mathrm{P}(T_i > t|x_i)$ can be estimated by

$$\hat{S}(t|x_i) = \exp(-\hat{A}(t|x_i)) = \exp(-\hat{A}_0(t)\exp(\hat{\beta}' x_i)) = \hat{S}_0(t)^{\exp(\hat{\beta}' x_i)}.$$

Cox-regression for left truncated and rightcens. data

We may use the same likelihood expression in counting process notation. We may estimate β by maximizing the partial likelihood, or solving

$$U(\beta) = \frac{\partial \log(L(\beta))}{\partial \beta} = \sum_{i=1}^{n} \int \left[x_i - \frac{S^{(1)}(\beta, t)}{S^{(0)}(\beta, t)} \right] dN_i(t) = 0$$

where

$$S^{(0)}(\beta, t) = \sum_{i=1}^{n} Y_i(t) \exp(\beta' x_i).$$

We need only remember that $Y_i(t)$ need not be non-decreasing, but

$$Y_i(t) = I(L_i < t \leq T_i).$$

No arguments have used the property of non-decreasing $Y_i(t)$ (might be a good exercise to check).

R: Cox-reg. with l.trunc. and r.cens.

The program need information on left truncation time L_i in addition to $(\tilde{T}_i, D_i, x_i)$.

Example: Time until death psychiatric patients. Data:

$x_i = 1$ or 2 for men / women

$L_i =$ age of first time admitted to psychiatric ward (year)

$T_i =$ age at death/censoring (year)

$D_i =$ indicator of death.

```
> coxph(Surv(ageonset,agedeath,death)~sex)
    coef exp(coef) se(coef)     z    p
sex 0.39      1.48     0.61 0.639 0.52

Likelihood ratio test=0.43  on 1 df, p=0.514  n= 26
```

Time dependent covariates $x_i(t)$

Risk factors may depend on time:

- $x_i(t) =$ smoker at time t (yes/no)
- $x_i(t) =$ cum. no. cigarettes (pack years) smoked at age t
- $x_i(t) =$ no. years since quitting smoking

It may furthermore be that the proportional hazards model does not fit the data well with the included covariates, but that a valid model is given by.

$$\alpha(t|x_i) = \exp(\beta_1 x_i + \beta_2 x_i t)\alpha_0(t)$$

so that the effect of x_i becomes larger (smaller)according to $\beta_2 > 0 (< 0)$. May code this by introducing new covariates, for instance $x_{i2}(t) = t x_i$.

Cox-regression allows for time dependent $x_i(t)$!

Model: $\alpha_i(t) = \exp(\beta' x_i(t))\alpha_0(t)$. Will simply solve

$$U(\beta) = \frac{\partial \log(L(\beta))}{\partial \beta} = \sum_{i=1}^{n} \int \left[x_i - \frac{S^{(1)}(\beta, t)}{S^{(0)}(\beta, t)} \right] dN_i(t) = 0$$

(as before) with the difference that

$$S^{(0)}(\beta, t) = \sum_{i=1}^{n} Y_i(t) \exp(\beta' x_i(t)).$$

depend on time dependent $x_i(t)$.

Similarly to left truncation all theory goes through exactly the same way as before. For the sake of repetition we check $\mathrm{E}[U(\beta)] = 0$.

However: We need to assume that $x_i(t)$ is predictable!

$U(\beta)$ martingale with expectation 0. As before

$\frac{\partial S^{(0)}(\beta,t)}{\partial \beta} = \sum_{i=1}^n x_i(t) Y_i(t) \exp(\beta' x_i(t)) = S^{(1)}(\beta, t)$. Thus the score

$$U(\beta) = \sum_{i=1}^{n} \int [x_i(t) - \frac{S^{(1)}(\beta, t)}{S^{(0)}(\beta, t)}] dN_i(t)$$

With predictable covariates $x_i(t)$ the model can be written

$$dN_i(s) = Y_i(t) \exp(\beta' x_i(t)) \alpha_0(t) dt + dM_i(t)$$

(with standard interpretations) which gives

$$\begin{aligned} U(\beta) = & \sum_{i=1}^n \int [x_i(t) - \frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}] Y_i(t) \exp(\beta' x_i(t)) \alpha_0(t) dt \\ & + \sum_{i=1}^n \int [x_i(t) - \frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}] dM_i(t) \\ = & \sum_{i=1}^n \int [x_i(t) - \frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}] dM_i(t) \end{aligned}$$

because

$$\sum_{i=1}^{n} \int [x_i(t) - \frac{S^{(1)}(\beta, t)}{S^{(0)}(\beta, t)}] Y_i(t) \exp(\beta' x_i(t)) \alpha_0(t) dt = 0.$$

This since $\sum_{i=1}^{n} [x_i(t) - \frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}] Y_i(t) \exp(\beta' x_i(t))$

$$= S^{(1)}(\beta, t) - \frac{S^{(1)}(\beta, t)}{S^{(0)}(\beta, t)} S^{(0)}(\beta, t) = 0$$

and so

$$U(\beta) = \sum_{i=1}^{n} \int [x_i(t) - \frac{S^{(1)}(\beta, t)}{S^{(0)}(\beta, t)}] dM_i(t)$$

is a sum of integrals wrt. martingales, and itself a martingale with $\mathrm{E}[U(\beta)] = 0$. just as for time constant and right censored data!

Time dependent covariates in R

R only allows for time dependent covariates **constant on intervals**, i.e. step functions.

Assume that $x_i(t) = x_j$ on interval $< L_{ij}, U_{ij}]$ for $j = 1, 2, \ldots, J_i$.

Need to represent this individual J_i times in the data fil as left truncated data with

- L_{ij} as left trunkcation time
- U_{ij} as right censoring time
- $D_{ij} = D_i I(\text{ event in interval } j)$ as indicator
- x_j as covariate value

Note: Did not assume $L_{i,j+1} = U_{ij}$, and individuals may well disappear and reenter later.

Stratified Cox-regression

Assume that the population is divided into $s = 1, 2, \ldots, S$ strata so that person $i = 1, 2, \ldots, n_s$ within stratum s with covariate x_{is} has hazard

$$\alpha(t|x_{is}) = \exp(\beta' x_{is})\alpha_{0s}(t)$$

where

- $\alpha_{0s}(t)$ is the baseline in stratum s
  (typically $\alpha_{0s}(t)/\alpha_{0s'}(t)$ varies with t.)
- Effects of covariates β are the same in all strata

We get a partial likelihood from each stratum:

$$L_s(\beta) = \prod_{i=1}^{n_s} \left[\frac{\exp(\beta' x_{is})}{\sum_{k \in \mathcal{R}_s(T_{is})} \exp(\beta' x_{ks})} \right]^{D_{is}}$$

with D_{is} = indicator for individual is in stratum s etc.

Stratified Cox-regression, contd.

All strata give information on β. If we assume that the strata are independent (weak assumption) we may then combine this information by maximizing the stratified partial likelihood

$$L(\beta) = \prod_{s=1}^{S} L_s(\beta)$$

The corresponding score function becomes

$$U(\beta) = \frac{\partial \log(L(\beta))}{\partial \beta} = \sum_{s=1}^{S} \frac{\partial \log(L_s(\beta))}{\partial \beta} = \sum_{s=1}^{S} U_s(\beta)$$

where $U_s(\beta)$ is the score function from stratum s. These all have expectation 0, i.e. $\mathrm{E}[U(\beta)] = 0$.

Stratified Cox-regression, III

We find information I_s from stratum s and $I(\beta)$ total information becomes, due to independence,

$$\text{Var}[U(\beta)] = \sum_{s=1}^{S} \text{E}[I_s(\beta)] = \text{E}[I(\beta)]$$

Thus may use the stratified partial likelihood as a regular likelihood.

- The stratified partial likelihood is useful when the proportional model does not hold for a categorical variable.
- Stratify on this variable and keep regression model for other covariates
- In particular this is useful when the stratification variable is a confounder and the main interest is on the other variables.

Stratified Cox-regression i R

Uses the Melanoma-data. Stratifies on grouped tumor thickness by command `strata`:

```
coxph(Surv(lifetime,dead)~ulcer+sex+age+strata(grthick),data=mel)

         coef exp(coef) se(coef)      z      p
ulcer -0.9480     0.388  0.32572 -2.910 0.0036
  sex  0.4074     1.503  0.27351  1.490 0.1400
  age  0.0063     1.006  0.00837  0.753 0.4500

Likelihood ratio test=13.2  on 3 df, p=0.00426  n= 205
```

What did we gain from this?

Compares with the Cox-regression where thickness is a categorical variable

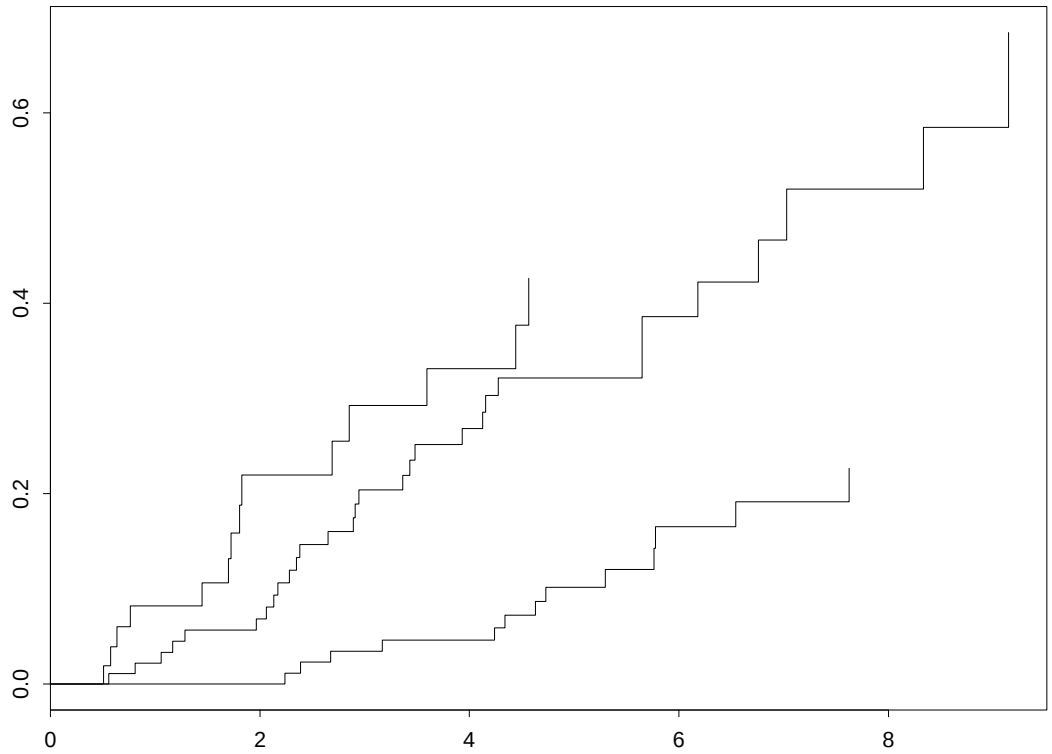
```
> coxph(Surv(lifetime,dead)~ulcer+sex+age+factor(grthick),data=mel)

                      coef exp(coef) se(coef)     z      p
            ulcer -0.9562     0.384  0.32407 -2.95 0.0032
              sex  0.3416     1.407  0.27127  1.26 0.2100
              age  0.0103     1.010  0.00845  1.22 0.2200
factor(grthick)2  1.0440     2.841  0.36538  2.86 0.0043
factor(grthick)3  1.1207     3.067  0.41641  2.69 0.0071

Likelihood ratio test=45.3  on 5 df, p=1.27e-08  n= 205
```

The results are only marginally different, but we have modeled in a much more flexible way.

There is hardly a change in the standard errors, and the added flexibility did not lead to loss in efficiency.

Baseline in each stratum in R.

```
> nycox<-coxph(Surv(lifetime,dead)~ulcer+sex+age+strata(grthick),data=m
> plot(survfit(nycox),fun="cum.haz")
```



The proportional hazards assumption might be checked from these plots (or rather the log-cumulative hazard plots).

Better methods are presented later in the lectures.

How can the Cox-model fail?

$$\alpha(t|x_i) = \exp(\beta' x_i)\alpha_0(t)?$$

The Cox-model is flexible wrt the baseline $\alpha_0(t)$, but otherwise strict with respect to how the hazard depend on covariates, for instance

- We may have specified covariate x_{ik} wrong, correct alternative may be f.ex. $x'_{ik} = \log(x_{ik})$ or $x''_{ik} = \sqrt{x_{ik}}$.
- We do not have a proportional model. The effect may vary with time, f.ex. $\alpha(t|x_i) = \exp(\beta(t)' x_i)\alpha_0(t)$ where $\beta(t)$ is a function of time.

Martingale residuals

With a specified model for the hazard, say a proportional hazards model $\alpha_i(t) = \exp(\beta' x_i)\alpha_0(t)$ we that

$$M_i(t) = N_i(t) - \int_0^t Y_i(s) \exp(\beta' x_i)\alpha_0(s)ds$$

is a martingale we expectation zero. Inserting the Cox-estimator $\hat{\beta}$ for β, the Breslow estimator $d\hat{A}_0(s) = \frac{dN_\bullet(s)}{S^{(0)}(\hat{\beta},s)}$ for $\alpha_0(s)ds$ and the maximal right censored survival time τ we get the so called **martingale residuals**

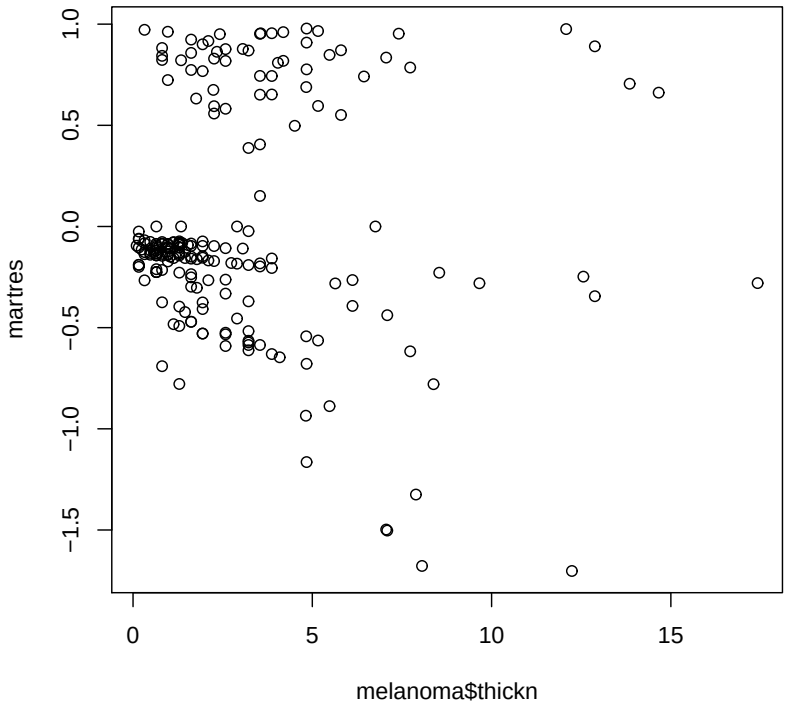
$$\hat{M}_i = N_i(\tau) - \int_0^\tau Y_i(s) \exp(\hat{\beta}' x_i)d\hat{A}_0(s)$$

which have the structure "Observed"-"Expected".

Example: Melanoma data

We saw that $\log(\text{tumorsize})$ was a better covariate than tumorsize directly. We will check if this can be discovered from martingale residuals.

```
coxfit<-coxph(Surv(lifetime,status==1)~sex+ulcer+thickn, data=melanoma)
martres<-coxfit$residuals
plot(melanoma$thickn,martres)
```



Example: Melanoma data, contd.

As often it is difficult to read off residual plots directly. We calculate mean mart. resid. for groups of tumor thickness

```
> grthickn<-melanoma$grthick
> lm(martres~factor(grthickn)-1)

Coefficients:
factor(grthickn)1  factor(grthickn)2  factor(grthickn)3
         -0.05661              0.14225           -0.09167
> summary(lm(martres~factor(grthickn)))

                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       -0.05661    0.05279  -1.072    0.285
factor(grthickn)2  0.19886    0.08680   2.291    0.023 *
factor(grthickn)3 -0.03506    0.11082  -0.316    0.752
---
F-statistic: 3.149 on 2 and 202 DF,  p-value: 0.04502
```

Note: This analysis is not strictly correct, from ABG one may figure out a correct test. But it shows that the martingale residuals are largest for the second group.

GAM: Generalized Additive Models

(Hastie & Tibshirani (1990): By smoothing techniques we may fit

$$\text{Linear model:} \qquad Y = \alpha + f(x) + \varepsilon$$

$$\text{Logistic model:} \quad \log(\tfrac{p}{1-p}) = \alpha + f(x)$$

where $f(x)$ is some smooth function. This may be done in R with the library `gam` that may need to be downloaded from CRAN.

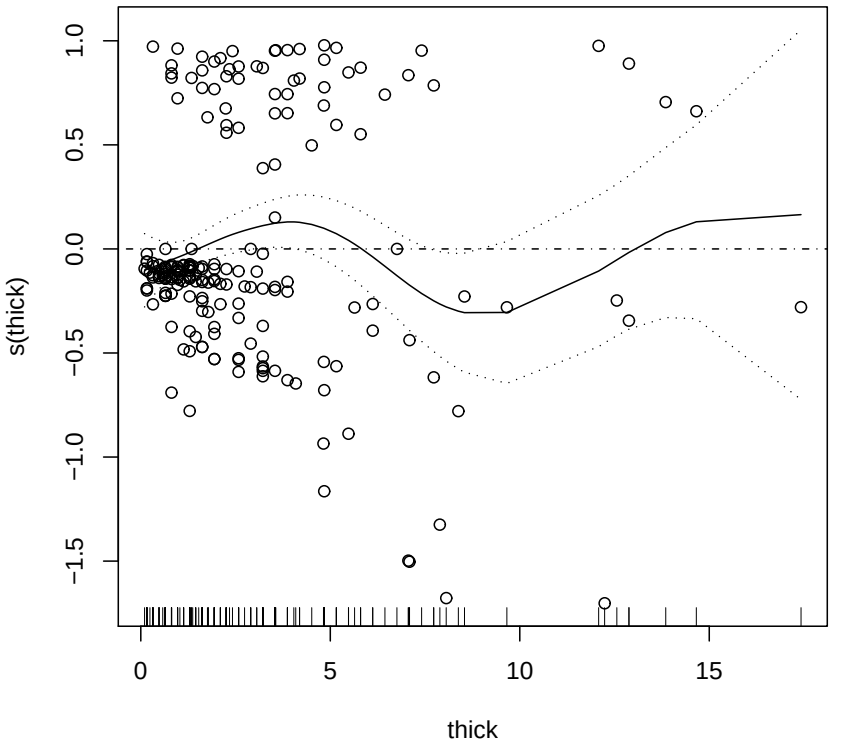
Similar models may be fitted for survival data under specifications

$$\alpha(t|x) = \exp(f(x))\alpha_0(t)$$

Example: Melanoma data, GAM-plot

We add a gam-curve with 95% CI to the scatter plot of martingale residuals vs. thickness.

```
thick<-melanoma$thickness)
library(gam)
plot(gam(martres~s(thick)),se=T,ylim=c(min(martres),max(martres)))
points(thick,martres)
```



GAM for Cox-regression in R

- Regression-spline: R-syntax

  ```
  coxph(Surv(time,status)~ns(z,df=4))
  ```

- Cubic smoothing-spline (penalized partial likelihood):

  R-syntax:

  ```
  coxph(Surv(time,status)~pspline(z,df=4))
  ```

Penalized log-partial-likelihood:

Maximize, for given smoothing parameter λ,

$$\log(L(f)) - \lambda \int (f''(x))^2 dx$$

where

- $L(f) =$ Partial likelihood
- $\lambda =$ "penalty"-term for curvature of $f(x)$

In particular:

- $\lambda = \infty$: No curvature, $f(x) = \beta x$ straight line
- $\lambda = 0$: No smoothing

Maximation problem actually has a simple numerical solution. The penalty term λ corresponds to a certain degree of freedom and may be interpreted as df = no. covariates in mod., though df may be any real number > 0.

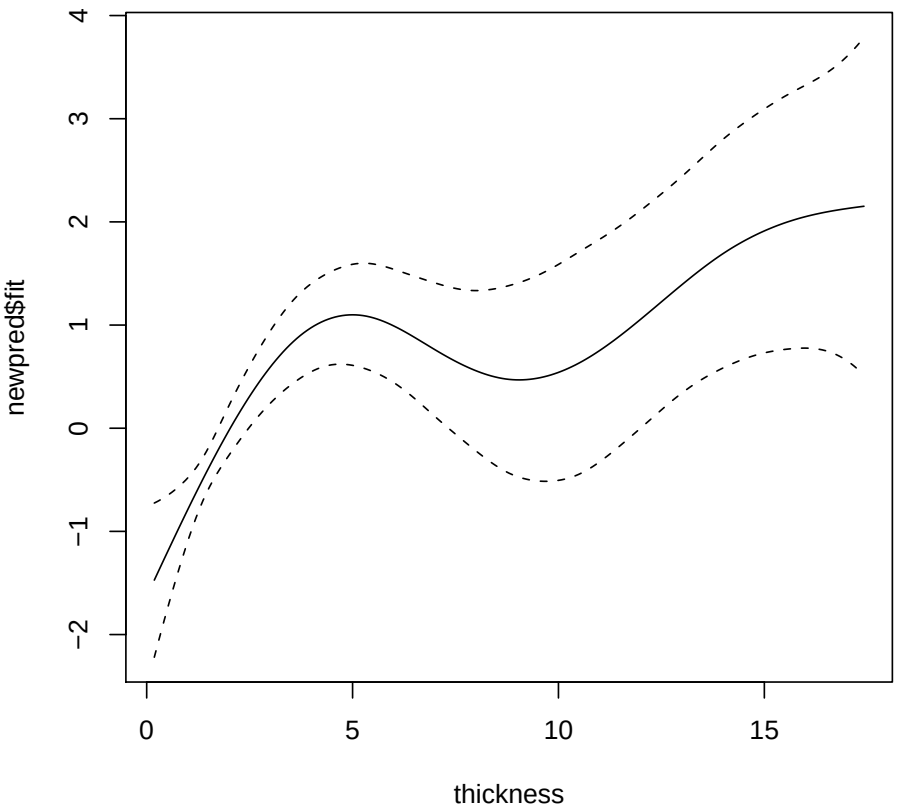
Example: Penalized log-partial-likelihood:

Melanoma-data, thickness, R-commands:

```
newcoxfit<-coxph(Surv(lifetime,status==1)~pspline(thickn,df=4),data=mel

newmel<-melanoma
thckspacing<-min(melanoma$thickn)
             +(1:205)*(max(melanoma$thickn)-min(melanoma$thickn))/205
newmel$thickn<-thckspacing
newpred<-predict(newcoxfit,newmel,type="terms",term=1,se=T)

mi<-min(newpred$fit-1.96*newpred$se.fit)
ma<-max(newpred$fit+1.96*newpred$se.fit)
plot(thckspacing,newpred$fit,type="l",ylim=c(mi,ma),xlab="thickness")
lines(thckspacing,newpred$fit+1.96*newpred$se.fit,lty=2)
lines(thckspacing,newpred$fit-1.96*newpred$se.fit,lty=2)
```

Example: Penalized log-partial-likelihood:

Melanoma-data, thickness, Plot of $\hat{f}(x)$ against $x =$ thickness:



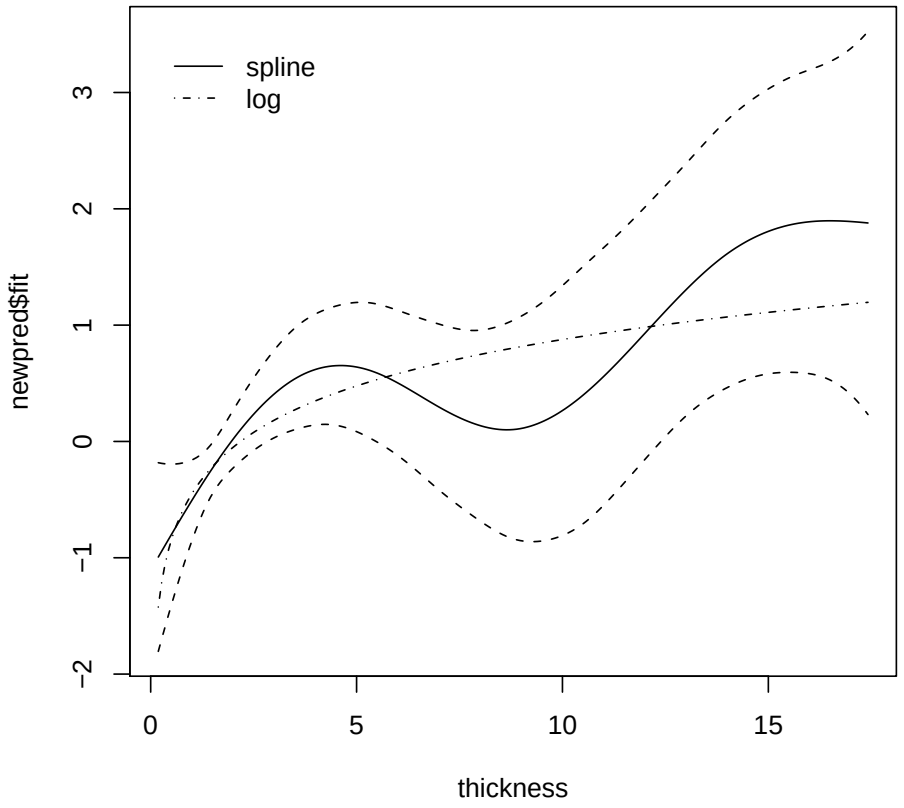
The plot shows what we also found with grouping of thickness, it is the smallest tumors that have smaller risk.

May smooth with more covariates

$$\alpha(t|x) = \exp(\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + f_4(x_4))\alpha_0(t)$$

Ex: Melanoma, $x_1 =$ sex, $x_2 =$ ulc., $x_3 =$ age, $x_4 =$ tumorth.

```
newcoxfit<-coxph(Surv(lifetime,status==1)~sex+ulcer+age
                                          +pspline(thickn,df=4),data=me
newpred<-predict(newcoxfit,newmel,type="terms",term=4,se=T)
```



Non-proportional hazards

The old-fashioned way of checking departure from proportionality is based on the following: With $\alpha(t|x) = \exp(\beta x)\alpha_0(t)$ we have

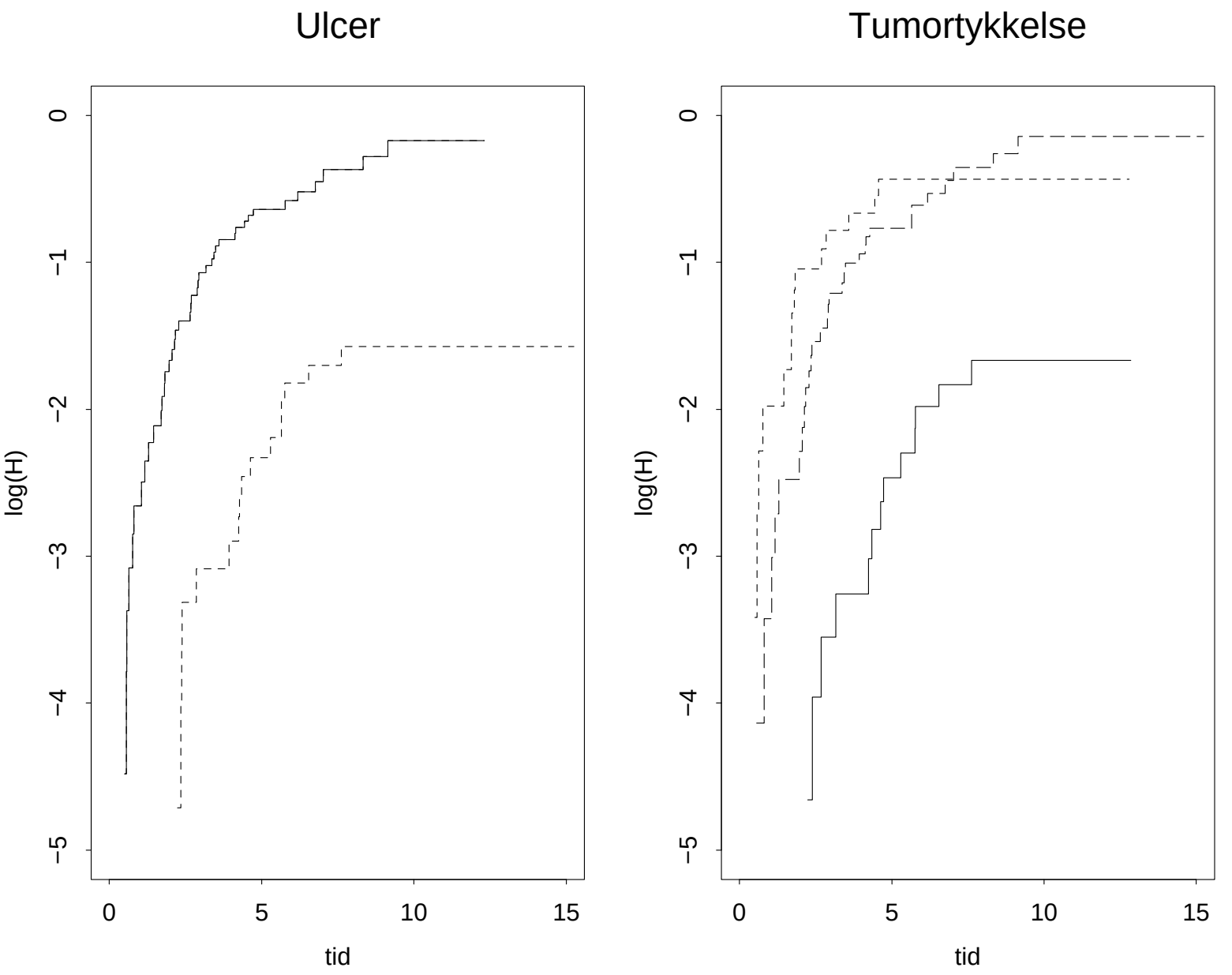
$$\log(A(t|x)) = \beta' x + \log(A_0(t))$$

i.e. $\log(A(t|x))$ for different x and $\log(A_0(t))$ should be parallel lines.

Thus if x is the only covariate and is categorical we may plot

- $\log$ of Nelson-Aalen for every level of x
- If the lines are parallel then proportionality OK

Ex. Melanoma: Checks tumor-thickness and ulceration



Multivariable models

If we have decided that $x_1, x_2, \ldots, x_p$ should be included in the model and want to check if the categorical covariate x_{p+1} satisfies proportionality we should rather

- Fit a stratified Cox-model with x_{p+1}-levels as strata and $x_1, x_2, \ldots, x_p$ as covariates.
- Plot $\log(\hat{A}_s(t, \bar{x}))$ against t for different levels s of x_{p+1}
- Are the lines parallel?

Again, it might be an advantage to look at $\log(\hat{A}_s(t, \bar{x})) - \log(\hat{A}_1(t, \bar{x}))$.

Schoenfeld residuals

The Schoenfeld-residuals are given by, for $\tilde{T}_j$ such that $D_j = 1$

$$x_{jk} - \frac{\sum_{i=1}^{n} x_{ik} Y_i(\tilde{T}_j) \exp(\hat{\beta}' x_i)}{\sum_{i=1}^{n} Y_i(\tilde{T}_j) \exp(\hat{\beta}' x_i)}$$

There is thus one residual for event time and for each component of the covariates.

Since $\frac{Y_i(\tilde{T}_j) \exp(\hat{\beta}' x_i)}{\sum_{i=1}^{n} Y_i(\tilde{T}_j) \exp(\hat{\beta}' x_i)}$ sums to one they may be thought of as point mass probabilities for some distribution.

The interpretation of the distribution is that of the covariates x_j given that individual j experienced the event. And so

$$\frac{\sum_{i=1}^{n} x_{ik} Y_i(\tilde{T}_j) \exp(\hat{\beta}' x_i)}{\sum_{j=1}^{n} Y_i(\tilde{T}_j) \exp(\hat{\beta}' x_i)} = \bar{x}_k(\tilde{T}_j)$$

More Schoenfeld

becomes the expectation in this distribution.

The Schoenfeldt-residuals at time $\tilde{T}_j$ may also be written $x_j - \bar{x}(\tilde{T}_j)$.

Furthermore, with $U(\beta) =$ the scorefunction,

$$0 = U(\hat{\beta}) = \sum_{\tilde{T}_j : D_j = 1} [x_j - \bar{x}(\tilde{T}_j)]$$

a sensible property for a residual.

Sometimes component k for $x_{jk} - \bar{x}_k(\tilde{T}_j)$ shows a clear tendency for positive values over intervals (and negative over others). In the "positive" intervals there is the greater risk connected to component k than in the "negative".

Schoenfeld, contd.

We could have done a local Cox-regression within the "positive" interval. In such case the $\bar{x}(t)$ would tend to be larger.

Thus the Schoenfeld-residuals give information as to whether the prop.haz. assumption holds.

They may even be used to estimate how the hazard ratio varies over time. Let

$$V(t) = \frac{S^{(2)}(\hat{\beta}, t)}{S^{(0)}(\hat{\beta}, t)} - \left[\frac{S^{(1)}(\hat{\beta}, t)}{S^{(0)}(\hat{\beta}, t)}\right]^2$$

The observed information is given as $\int V(t)dN(t)$ and in particular $V(t)$ can be interpreted as the variance of x given event at t.

Scaled Schoenfeld-residual

If the true model equals $\alpha(t|x) = \exp(\beta(t)x)\alpha_0(t)$ for some function $\beta(t)$ we have as a 1.order approximation

$$\beta(\tilde{T}_j) \approx \hat{\beta} + V(\tilde{T}_j)^{-1}(x_j - \bar{x}(\tilde{T}_j))$$

i.e. varying around the scaled Schoenfeld-residual

$$V(\tilde{T}_j)^{-1}(x_j - \bar{x}(\tilde{T}_j))$$

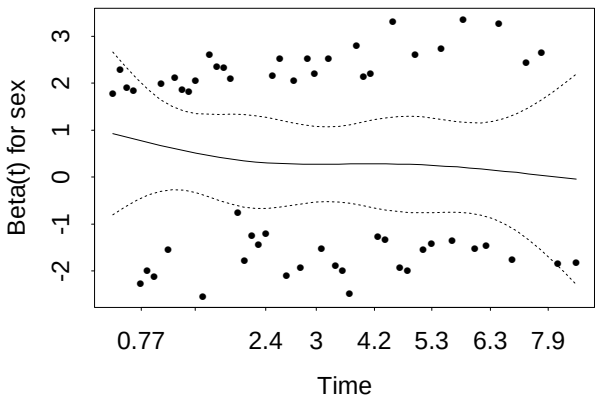
that measures the departure from mean risk $\hat{\beta}$.

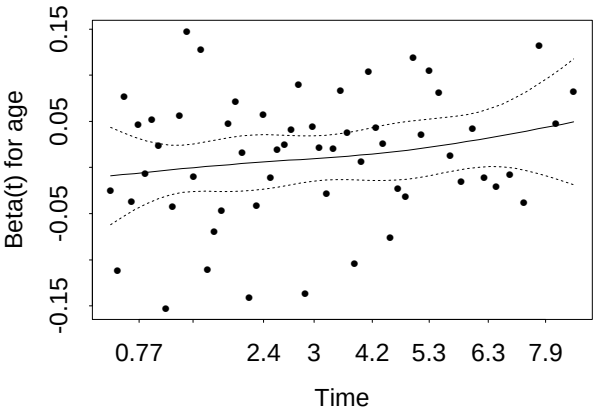
Plot of this quantity - smoothed over time - may give a picture of how the risk varies.

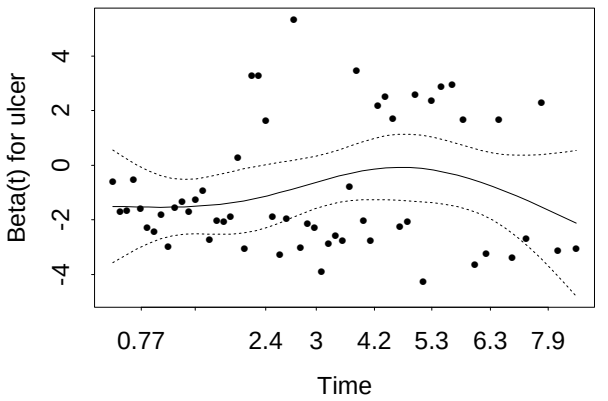
Ex: Melanoma data

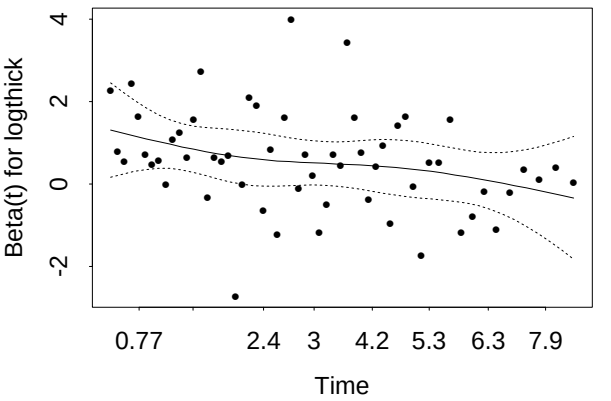
Calculates and smooths scaled Schoenfeld-residuals for all 4 covariates in the melanom data.

```
coxfit<-coxph(Surv(lifetime,dead)~sex+age+ulcer+logthick,data=mel)
plot(cox.zph(coxfit))
```









Test for prop. assumption

can be based on Schoenfeld-residuals: The tests are directed to departures from the model given by

$$\alpha(t|x) = \exp((\beta_0 + \theta g(t))x)\alpha_0(t)$$

for specified functions $g(t)$. Actually they are score-tests for a new covariate $x'(t) = g(t)x$.

If the model is extended by a new time-dependent covariate $x_{p+1}(t) = x_p g(t)$ the score for the coefficient β_{p+1} is given

$$U_{p+1} = \sum_j g(\tilde{T}_j)[x_p - \bar{x}_p(\tilde{T}_j)]$$

as a weighted sum of the Schoenfeld-residuals.

Test for proportionality

Example: Melanoma data: KM transform.

Sex	$\chi^2_1 = 0.5$	p=0.46
Ulceration	$\chi^2_1 = 0.7$	p=0.40
Age	$\chi^2_1 = 2.8$	p=0.09
log(Tumor thickness)	$\chi^2_1 = 4.1$	p=0.04

Indication for departure for tumor thickness

R-syntax:

```
coxfit<-coxph(Surv(lifetime,dead)~sex+age+ulcer+logthick,data=mel)
cox.zph(coxfit)
             rho  chisq      p
     sex -0.095  0.536 0.4642
     age  0.200  2.828 0.0927
   ulcer  0.116  0.717 0.3972
logthick -0.299  4.079 0.0434
  GLOBAL     NA 10.450 0.0335
```

Strategies when proportional hazard fails

- Stratified Cox-regression
- Separate analyzes on disjoint time intervals
- Time-dependent covariates
- Alternative regression models
  - Accelerated failure time models
  - Additive models

Ex. Stratified Cox-regression

Weak departure wrt. thickness. Stratifies on `grthick`:

```
> coxstrat<-coxph(Surv(lifetime,dead)~sex+age+ulcer+strata(grthick),dat
> coxstrat

        coef exp(coef) se(coef)      z      p
  sex  0.4074     1.503  0.27351  1.490 0.1400
  age  0.0063     1.006  0.00837  0.753 0.4500
ulcer -0.9480     0.388  0.32572 -2.910 0.0036

Likelihood ratio test=13.2  on 3 df, p=0.00426  n= 205
> cox.zph(coxstrat)
           rho  chisq     p
   sex -0.0232 0.0313 0.860
   age  0.1178 1.0581 0.304
 ulcer  0.1037 0.5619 0.453
GLOBAL      NA 1.5924 0.661
```

But possibly the stratification changed other estimates somewhat?

Separate intervals

We may split the time interval i 2 and make separate

Cox-regressions within each interval:

Ex: Melanoma data. Half of death before $\tau = 3$
Analysis on $[0, 3 >$: Uses events only if `lifetime` < 3

```
coxph(Surv(lifetime,dead*(lifetime<3))~sex+age+ulcer+logthick,data=mel)
             coef exp(coef) se(coef)     z      p
     sex  0.4654     1.593    0.336  1.38 0.1700
     age  0.0106     1.011    0.010  1.06 0.2900
   ulcer -1.1979     0.302    0.449 -2.67 0.0076
logthick  0.6628     1.940    0.231  2.87 0.0041
```

Analysis on $[3, \infty >$: Uses only events with `lifetime` > 3

```
coxph(Surv(lifetime,dead*(lifetime>3))~sex+age+ulcer+logthick,data=mel)

             coef exp(coef) se(coef)      z      p
     sex  0.2204     1.247   0.3531  0.624 0.5300
     age  0.0332     1.034   0.0122  2.718 0.0066
   ulcer -0.5140     0.598   0.3833 -1.341 0.1800
logthick  0.2378     1.268   0.2148  1.107 0.2700
```

Time dependent covariats

If similar parameter estimates for age, sex and ulceration on these intervals we may fit a common model

$$\log\left(\frac{\alpha(t|x)}{\alpha_0(t)}\right) = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 I(t < 3) + \beta_5 x_4 I(t \geq 3)$$

i.e. Cox-regression with time dependent covariates

- $x_4 I(t < 3)$
- $x_4 I(t \geq 3)$

Cox-regression with time dependent covariates

Need to set up data-frame with time dependent covariates:

```
mel2 <- data.frame(intime = c(rep(0,205),rep(3,167)))
mel2$outtime <- c(pmin(mel$lifetime,3),mel$lifetime[mel$lifetime>3])
mel2$indi    <- c(mel$dead*(mel$lifetime<3),mel$dead[mel$lifetime>3])

mel2$sex    <- c(mel$sex,mel$sex[mel$lifetime>3])
mel2$ulcer    <- c(mel$ulcer,mel$ulcer[mel$lifetime>3])
mel2$age    <- c(mel$age,mel$age[mel$lifetime>3])

mel2$logtha <- c(mel$logthick,rep(0,167))
mel2$logthb <- c(rep(0,205),mel$logthick[mel$lifetime>3])

coxph(Surv(intime,outtime,indi)~sex+ulcer+age+logtha+logthb,data=mel2)
          coef exp(coef) se(coef)      z       p
   sex  0.3813     1.464  0.26901  1.417 0.16000
 ulcer -0.9845     0.374  0.32646 -3.016 0.00260
   age  0.0102     1.010  0.00823  1.239 0.22000
logtha  0.8985     2.456  0.24757  3.629 0.00028
logthb  0.2130     1.237  0.23346  0.912 0.36000

Likelihood ratio test=49  on 5 df, p=2.17e-09  n= 372
```

Advantages/disadvantages with strategies

1. Stratification
   - Easy
   - More difficult to show effect of stratification variable
   - Allows for only a few problem covariates
2. Separate intervals
   - Relatively easy
   - Choice of interval difficult/arbitrary
   - Looses power for covariates where the assumption is OK
   - Many parameter estimates

Advantages/disadvantages with strategies

3. Time dependent covariates

- Somewhat awkward to arrange (in R(?))
- Difficult choice of interval
- Only helpful when prop.haz. OK for most covar.

Consequences of departure from proportionality

- biased estimates of coefficients
- both for covariates where the assumption hold and fail
- biased survival estimates