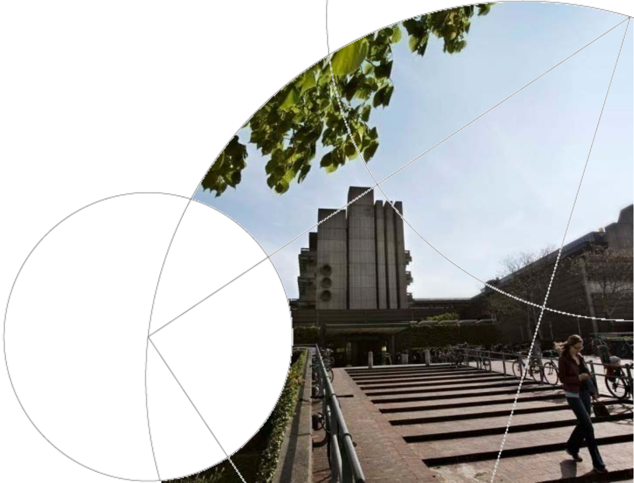
Survival analysis – Cox models

Theis Lange
Department of Biostatistics, University of Copenhagen
&
Center for Statistical Science, Peking University.

Mail: thlan@sund.ku.dk



Follow-up from Tuesday

Two questions:

1) Math detail of log-rank test.

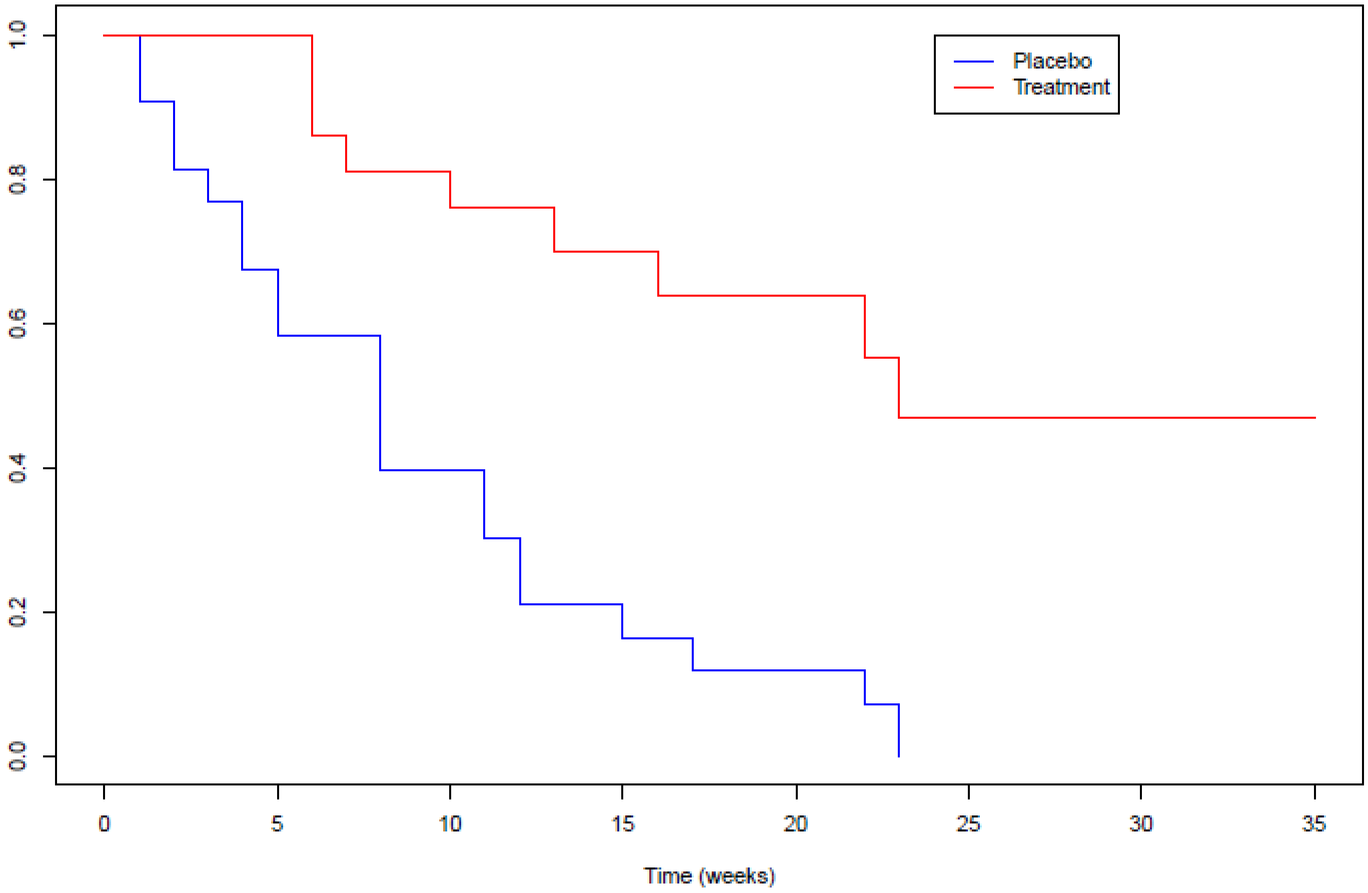
Solution on whiteboard.

2) Confidence band for KM-est.

Recall that var*((n-d)/n) = ((n-d)/n)*(1-(n-d)/n))*(1/n)*
Use delta-method with the function log(x)
Use delta-method with the function exp(x)
Result is this formula:

$$\widehat{\mathrm{Var}}(\widehat{S}(t)) = \widehat{S}(t)^2 \sum_{t_i \le t} \frac{d_i}{n_i(n_i - d_i)}$$

Recall: KM plots



Recall: The log-rank test in R

```
> survdiff(Surv(time, event)~placebo, data=remisData)
```

```
Call:
survdiff(formula = Surv(time, event) ~ placebo, data =
   remisData)

           N Observed Expected (O-E)^2/E (O-E)^2/V
placebo=0 21        9     19.3      5.46      16.8
placebo=1 21       21     10.7      9.77      16.8

 Chisq= 16.8  on 1 degrees of freedom, p= 4.17e-05
```

But what about getting a number for the effect size?

Why models for survival data?

- We want a parameter that describes the size of the difference between the two treatment groups.

- Not enough with p-value.

- Could we use usual parameters like:
  1. Mean survival time?
  2. Mediation survival time?
  3. Survival probability at say 90 days?

- First two does not work with censoring,
  the third only describes treatment effect at a single time point.

- We want to be able to include more than one covariate.

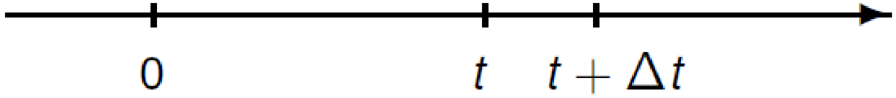
- Solution is the famous Cox model.

Recall: The hazard function

The hazard function (also referred to as hazard rate or intensity):

$$\lambda(t) \approx \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}$$

where the probability is read like: The conditional probability at time t of dying in the next short time interval $(t + \Delta t)$ given alive at t.



The hazard function provides a *local* description of the development.

Constant hazard model

The simplest model for the hazard would be

$$h(t) = h$$

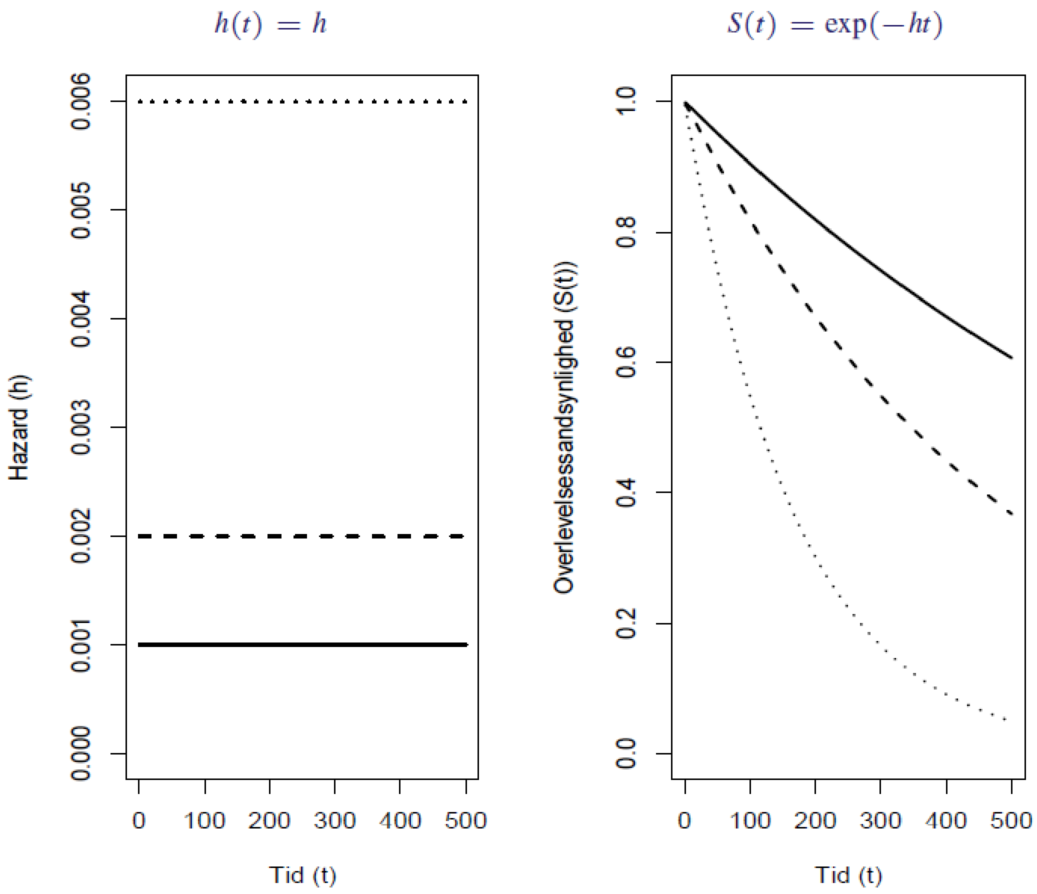
for h>0.

Then the survival function becomes

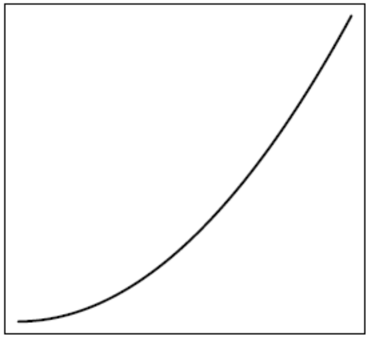
$$S(t) = \exp(-\int_0^t h(s)ds) = \exp(-\int_0^t hds) = \exp(-h \times t)$$

This model is know as the exponential survival model.

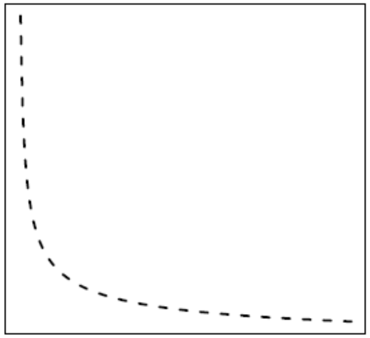
The exponential survival model



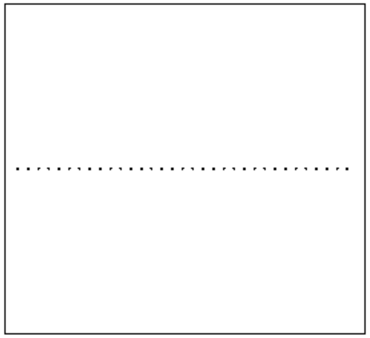
Other examples of hazard functions



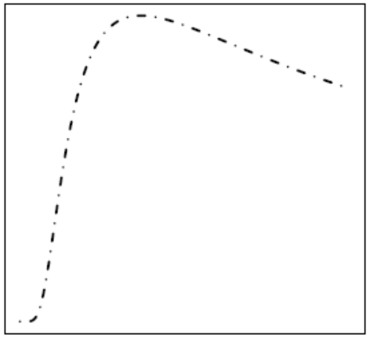
Leukaemia



Recovering



Healthy



Tuberculosis

Do we really have to chose in advance?

Hazard ratio (HR)

- Inspired by risk ratios we could calculate the rate between hazards.

- For remission data this would be

$$\frac{h^B(t)}{h^P(t)} = \frac{P(t \le T < t + d \mid T \ge t, \text{ Treated})}{P(t \le T < t + d \mid T \ge t, \ Placebo)}$$

- Interpretation is:
  For any time point the HR captures how much bigger/smaller the risk of death within a short time span is in treatment group compared to placebo.

- Note: HR can depend on time in general!

The Cox Proportional Hazards (PH) model

Let $X_i = (X_{i1}, X_{i2}, \dots, X_{ip})$ be a list of covariates for individual i.

The Cox PH model specifies the hazard for individual i as

$$\lambda_i(t) \quad = \quad \lambda_0(t) \exp(\beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip}).$$

If all covariates are 0 we get the *baseline hazard*

$$\lambda_i(t) \quad = \quad \lambda_0(t).$$

Only the baseline hazard depends on t.

The PH assumption is

$$\frac{\lambda_i(t)}{\lambda_j(t)} \quad = \quad \exp(\beta_1 (X_{i1} - X_{j1}) + \cdots + \beta_p (X_{ip} - X_{jp})).$$

i.e. constant over time.

Interpretation of the regression parameters

One binary covariate, e.g.

$$X_i = \begin{cases} 0 & \text{if individual } i \text{ is treated} \\ 1 & \text{if individual } i \text{ is not treated.} \end{cases}$$

The Cox model is

$$\lambda_i(t) = \lambda_0(t)\exp(\beta X_i) = \begin{cases} \lambda_0(t) & \text{if } i \text{ is treated} \\ \lambda_0(t)\exp(\beta) & \text{if } i \text{ is not treated.} \end{cases}$$

The hazard ratio (HR) or relative risk between non-treated and treated is

$$\frac{\lambda_0(t)\exp(\beta)}{\lambda_0(t)} = \exp(\beta).$$

Interpretation of the regression parameters

$$\text{HR} = \frac{\lambda_0(t)\exp(\beta)}{\lambda_0(t)} = \exp(\beta)$$

A treated patient has $\exp(\beta)$ the chance of relapsing compared to an untreated patient at *each* time point.

- HR < 1 ($\beta < 0$) treated relapse less than untreated
- HR $= 1$ ($\beta = 0$) treated and untreated have the same risk
- HR > 1 ($\beta > 0$) treated relapse more than untreated.

For a quantitative covariate (e.g. age, WBC)

$$\text{HR} = \frac{\lambda_0(t)\exp(\beta(X_{i1} + m))}{\lambda_0(t)\exp(\beta(X_{i1}))} = \exp(m\beta)$$

i.e. for each one-unit increase in the covariate, the HR is multiplied by $\exp(\beta)$.

Remission data - simple Cox

Define

$$\text{placebo} = \begin{cases} 0 & \text{if individual } i \text{ is treated} \\ 1 & \text{if individual } i \text{ is not treated.} \end{cases}$$

The simple Cox model is

$$\lambda_i(t) = \lambda_0(t)\exp(\beta \text{placebo}_i).$$

In R this model is fitted by:

```
library(survival)
coxFitObj1 <- coxph(Surv(time, event)~placebo, data=remisData)
summary(coxFitObj1)
```

Output on next slide.

Output from coxph-function in R

```
> summary(coxFitObj1)
Call:
coxph(formula = Surv(time, event) ~ placebo, data = remisData)

  n= 42, number of events= 30

          coef exp(coef) se(coef)     z Pr(>|z|)
placebo 1.5721    4.8169   0.4124 3.812 0.000138 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

        exp(coef) exp(-coef) lower .95 upper .95
placebo     4.817     0.2076     2.147     10.81

Concordance= 0.69  (se = 0.053 )
Rsquare= 0.322   (max possible= 0.988 )
Likelihood ratio test= 16.35  on 1 df,   p=5.261e-05
Wald test            = 14.53  on 1 df,   p=0.0001378
Score (logrank) test = 17.25  on 1 df,   p=3.283e-05
```

Remission data - Cox

Define

$$\text{female} = \begin{cases} 0 & \text{if individual } i \text{ is male} \\ 1 & \text{if individual } i \text{ is female.} \end{cases}$$

A possible Cox model is

$$\lambda_i(t) = \lambda_0(t) \exp(\beta_1 \text{placebo}_i + \beta_2 \text{female}_i + \beta_3 \text{logWBC}_i).$$

Baseline group: Males in treatment group with logWBC=0.

```
          coef exp(coef) se(coef)     z Pr(>|z|)
placebo 1.3909    4.0184   0.4566 3.046  0.00232
sex     0.2632    1.3010   0.4494 0.586  0.55817
logWBC  1.5936    4.9215   0.3300 4.829  1.37e-06
```

Is this model valid?

Assumptions for the Cox PH model

The ability of the Cox model to deal with many covariates comes from the regression structure,

$$\lambda_i(t) \quad = \quad \lambda_0(t) \exp(\beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip}).$$

- The effects of covariates are additive and linear on the log-risk scale:

$$\log(\lambda_i(t)) \quad = \quad \log(\lambda_0(t)) + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip}.$$

- If covariates interact with each other the regression model should include interaction terms
- Proportional hazards, i.e. the hazard ratio is constant over time

$$\frac{\lambda_i(t)}{\lambda_j(t)} \quad = \quad f((\beta_1, \ldots, \beta_p); X_i, X_j)$$

Importance of the PH assumption

- Crucial to carefully examine the assumption of proportionality. If the proportionality is not fulfilled the estimate for Cox's regression model is an average effect over time.
- Not correcting properly for important time varying effects may lead to severe bias for other estimates.
- A deeper understanding of what may be going on the data is very valuable.

If the PH assumption is not fulfilled for X_1, we may formulate a *stratified* Cox PH model

$$\lambda_i(t) \quad = \quad \lambda_{0k}(t)\exp(\beta_2 X_{i2} + \cdots + \beta_p X_{ip})$$

where k denotes the level (strata) of variable X_1.

In R stratified Cox models are fitted using the wrapper-function strata inside the coxph-function. Example:

```
coxph(Surv(time, event)~placebo+strata(female),
         data=remisData)
```

Evaluating the PH assumption

Several approaches:

- Graphical.
- Goodness-of-fit test.
- Time dependent variables.

More details in:

Kleinbaum and Klein (2005). Survival analysis. A Self-Learning Text. Springer.

A graphical approach for evaluating the PH assumption

The survival curve for the Cox PH model is

$$S(t \mid X_i) \quad = \quad S_0(t)^{\exp(\beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip})}.$$

Thus

$$\begin{aligned} \log(-\log(S(t \mid X_i))) \quad &= \quad \log(-\log(S_0(t))) \\ &\qquad + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip}. \end{aligned}$$

For two individuals *i* and *j* the difference between the survival curves

$$\begin{aligned} \log(-\log(S(t \mid X_i))) \; - \; \log(\, -\log(S(t \mid X_i))) \\ = \beta_1 (X_{i1} - X_{j1}) + \cdots + \beta_p (X_{ip} - X_{jp}) \end{aligned}$$

does not depend on time *t*, i.e. the curves are parallel.

Evaluation of the PH assumption

Asesseing the PH assumption for X_1, we assume PH is fulfilled for $X_2, \ldots, X_p$ and consider these fixed.

For a *binary* covariate we obtain two curves

$$\log(-\log(S(t \mid X_1 = 0, X_2, \ldots, X_p))),$$
$$\log(-\log(S(t \mid X_2 = 1, X_2, \ldots, X_p))).$$

For a *categorical* with k levels we obtain k curves.

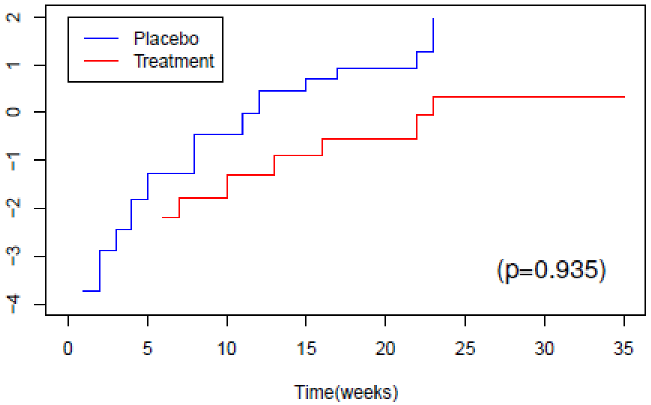
For *quantitative* X_1 we categorise X_1.

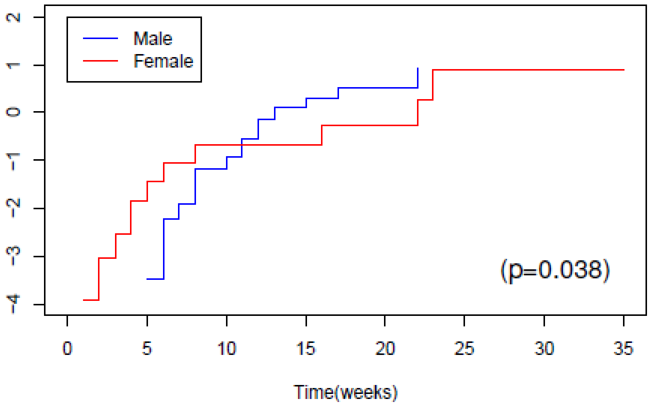
These models are fit by a Cox stratified on the levels of X_1:

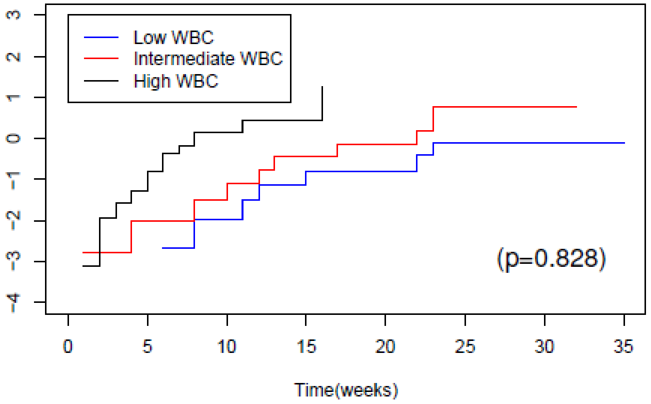
$$\lambda_i(t) \quad = \quad \lambda_{0k}(t) \exp(\beta_2 X_{i2} + \cdots + \beta_p X_{ip}).$$

Evaluation of the PH assumption for remission data

log-log-survival curves for remission data:







The p-values were found from a test based on Schoenfeld residuals.

PH problematic for sex

R-code to make plots on last slide

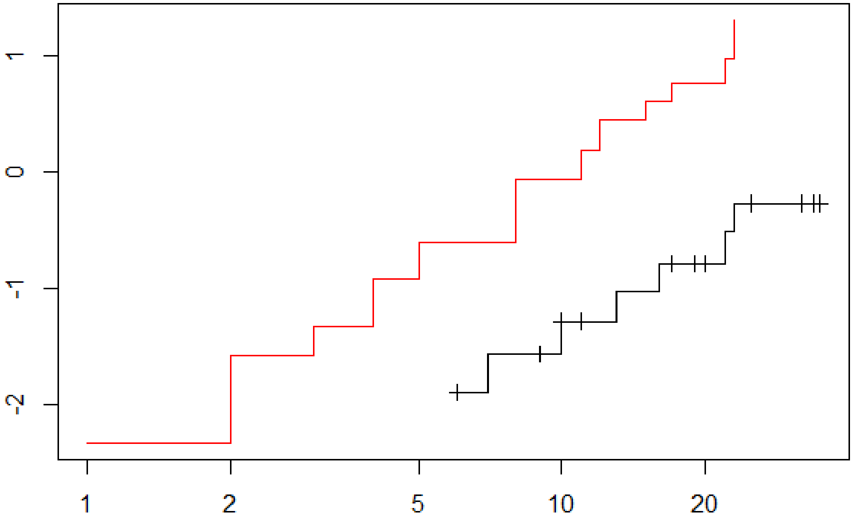
- The R function survfit can extract baseline hazard for each group defined by strata argument.

```
baselineFitObj1 <- survfit(coxph(Surv(time, event) ~
        strata(placebo), data=remisData))
```

- These can be plottet using plot and the argument fun="cloglog".

```
plot(baselineFitObj1 , col=c("black", "red"), fun="cloglog")
```

- P-value for test of proportional hazards can be obtained using the function **cox.zph**.



Departure from linearity

For continuous variables the linearity on the log-rate scale must be assessed. Having a single covariate X we may:

Categorise X into categories

$$\lambda_i(t) \quad = \quad \lambda_0(t)\exp(\beta_1(X_i \in (a_0, a_1]) + \cdots + \beta_k(X_i \in (a_{k-1}, a_k))).$$

to have an idea of the functional form of the effect. Requires a large sample size.

Include the covariate squared (or other transformations)

$$\lambda_i(t) \quad = \quad \lambda_0(t)\exp(\beta_1 X_i + \beta_2 X_i^2)$$

and test $\beta_2 = 0$ to test for departure from linearity.

Evaluation of the linearity for remission data

Including WBC squared:

```
               coef  exp(coef)  se(coef)      z Pr(>|z|)
placebo    1.752103   5.766718  0.490278  3.574 0.000352
sex        0.112526   1.119102  0.491584  0.229 0.818942
WBC       -0.028359   0.972040  0.041498 -0.683 0.494366
WBC2       0.001005   1.001005  0.000448  2.243 0.024926
```

Including logWBC:

```
              coef   exp(coef) se(coef)       z Pr(>|z|)
placebo    1.39335   4.02834   0.45368    3.071  0.00213
sex        0.16962   1.18485   0.46678    0.363  0.71633
WBC        0.01499   1.01510   0.01620    0.925  0.35495
logWBC     1.12094   3.06772   0.59466    1.885  0.05943
```

Interactions

Consider the two binary covariates $\mathrm{placebo}$ and sex for the remission data. Define

$$\mathrm{placeboF}_i \quad = \quad \begin{cases} 1 & \text{if } i \text{ is a female in the placebo group} \\ 0 & \text{otherwise.} \end{cases}$$

The Cox model becomes

$$\lambda_i(t) \quad = \quad \lambda_0(t)\exp(\beta_1 \mathrm{placebo}_i + \beta_2 \mathrm{female}_i + \beta_3 \mathrm{placeboF}).$$

The effect of treatment group now depends on sex (and vice versa). The reference (or baseline) group is males in the treatment group.

Interactions - categorical variables

$$\lambda_i(t) \quad = \quad \lambda_0(t)\exp(\beta_1 \mathrm{placebo}_i + \beta_2 \mathrm{female}_i + \beta_3 \mathrm{placeboF}).$$

The effect of placebo among males:

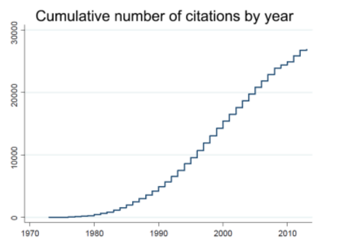
$$\frac{\lambda_0(t)\exp(\beta_1)}{\lambda_0(t)} \quad = \quad \exp(\beta_1)$$

The effect of placebo among females:

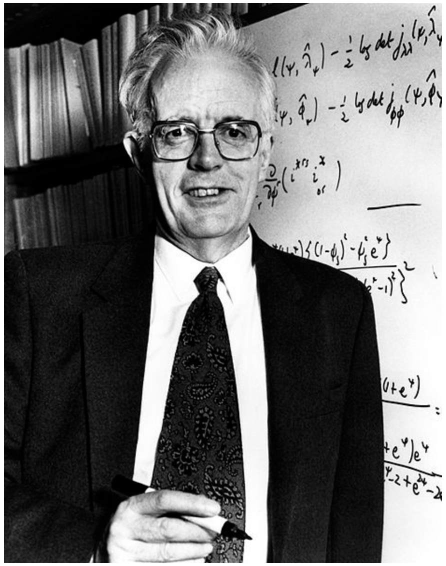
$$\frac{\lambda_0(t)\exp(\beta_1 + \beta_2 + \beta_3)}{\lambda_0(t)\exp(\beta_1)} \quad = \quad \exp(\beta_2 + \beta_3)$$

Output:

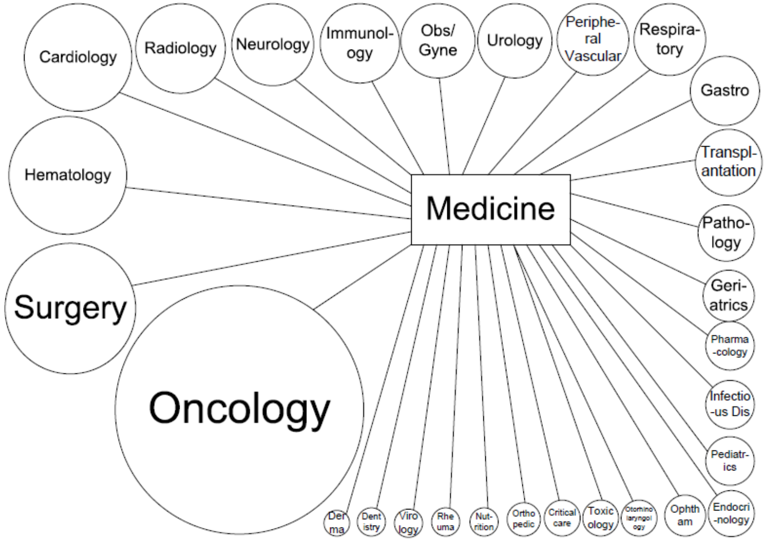
```
            coef exp(coef) se(coef)      z Pr(>|z|)
placebo   0.5867    1.7981   0.5420  1.082   0.2790
sex      -1.0726    0.3421   0.7014 -1.529   0.1262
placeboF  1.9059    6.7257   0.8148  2.339   0.0193
```


Cumulative number of citations by year

History of the Cox model

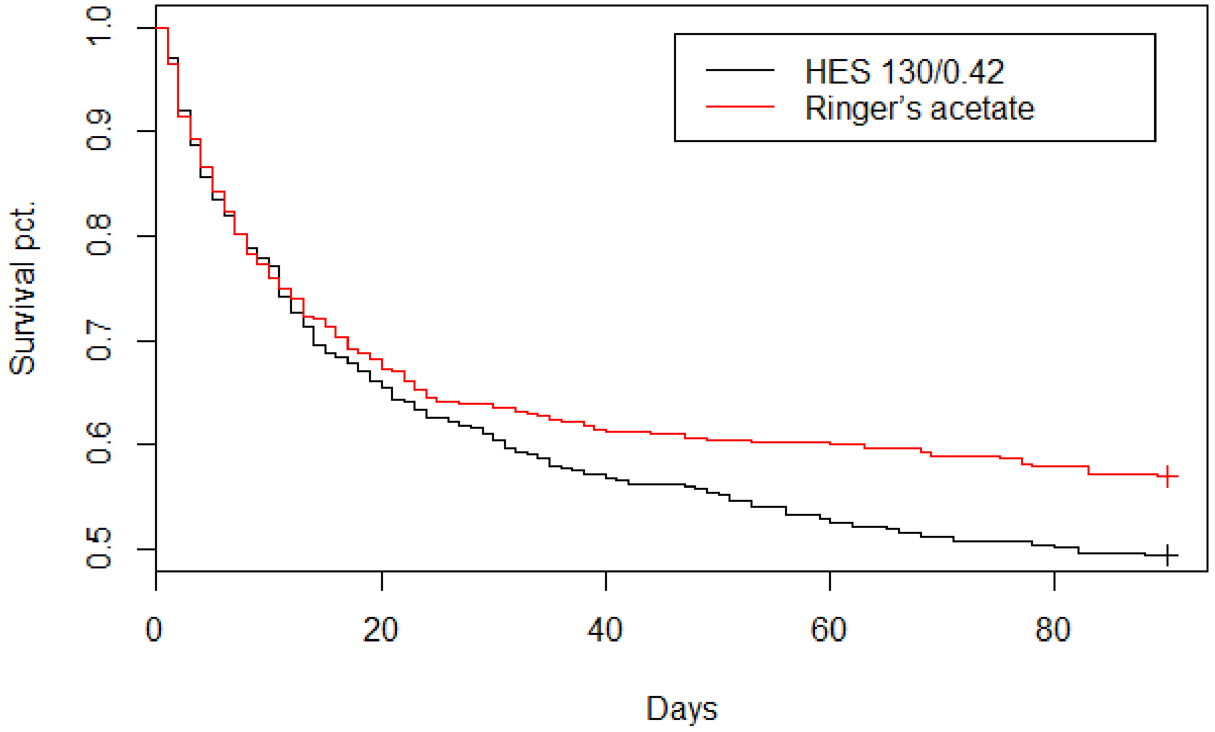


- Introduced in the 1972 paper “Regression Models and Life-Tables”, JRSS.
- One of the most cited statistics papers of all time.
- The model does not depend on time, only order of events.
- The central objective function is called the partial likelihood function for the same reason.
- Current asymptotic theory is based on counting process theory.
- Sir David Cox (born 1924) is still working within statistics



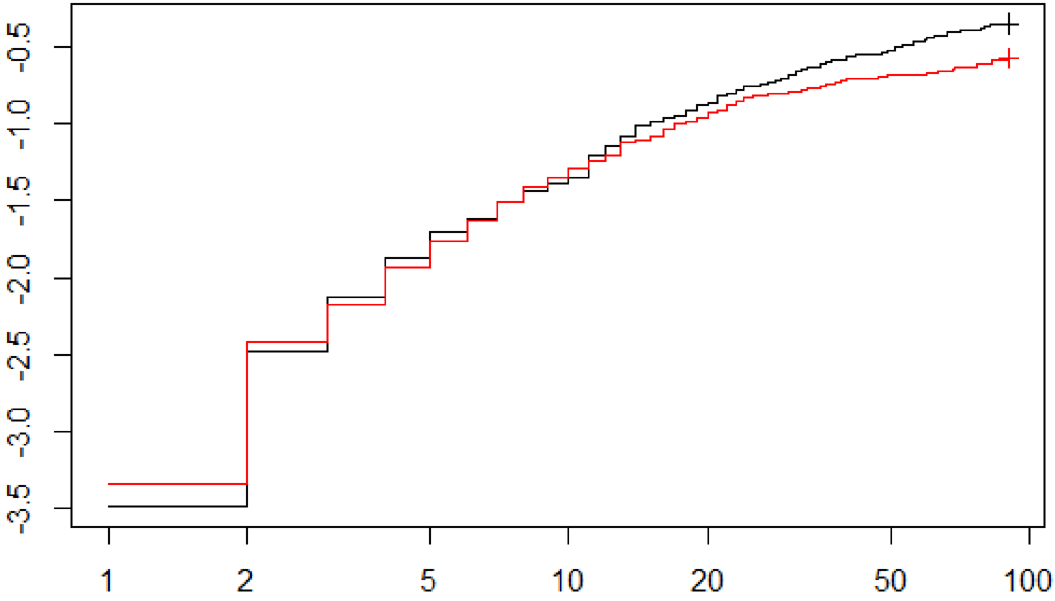
A case: Cox model applied to 6S data

Recall that we had the following survival curves for the 6S trial



6S: Do we have proportional hazards?

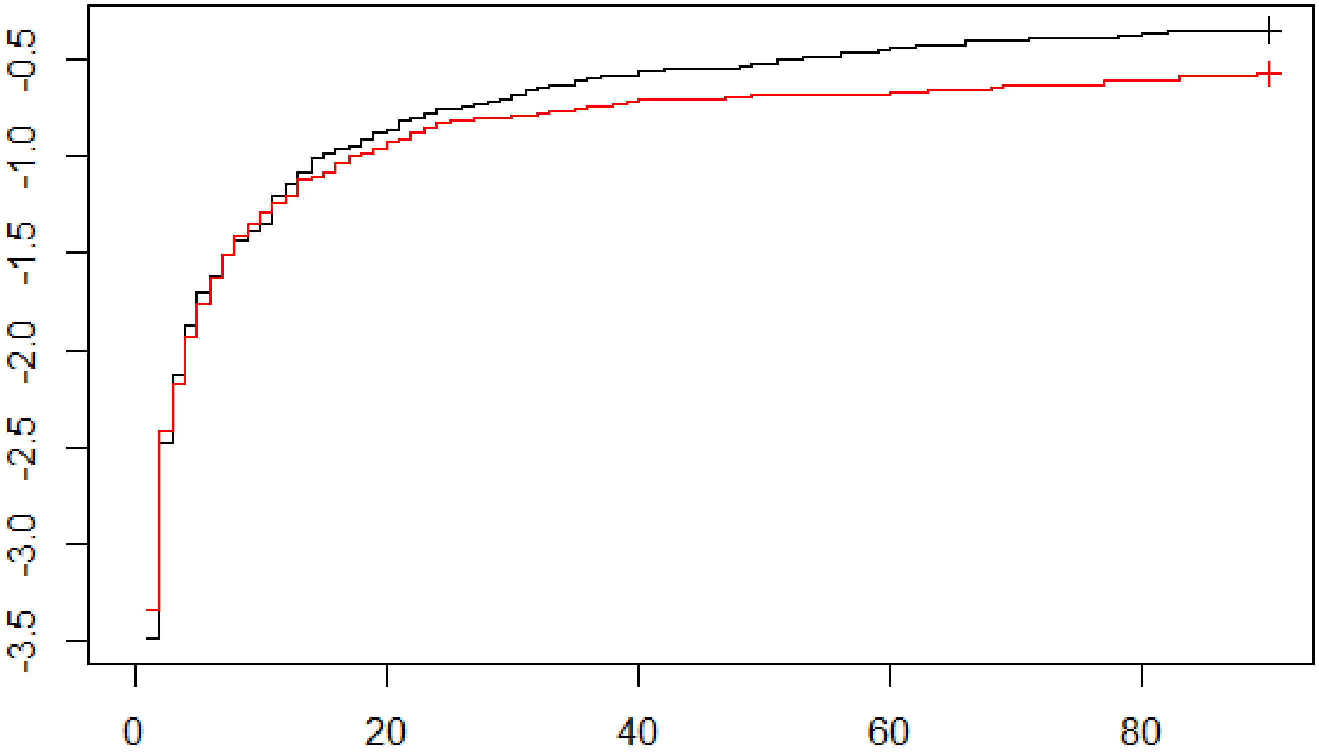
```
fit1 <- survfit(Surv(time_to_death,
   mortality_90days)~intervention, data=kidneyData)
plot(fit1, col=c("black", "red"), fun="cloglog")
```



Does not look too good. But would like non-log x-axis.

6S: Do we have proportional hazards?

```
fit1 <- survfit(Surv(time_to_death,
   mortality_90days)~intervention, data=kidneyData)
myFun <- function(q) return(log(-log(q)))
plot(fit1, col=c("black", "red"), fun=myFun)
```



6S: Do we have proportional hazards?

We could also do a formal test:

```
> coxfit1 <- coxph(Surv(time_to_death,
   mortality_90days)~intervention, data=kidneyData)
> cox.zph(coxfit1)
```

```
                rho chisq      p
intervention -0.0957  3.43 0.0641
```

So just OK.

6S: Alternative to proportional hazards

- We could also try to fit two different HR before and after some time point.

- How to pick the change point?

- In R you can get the estimates by:

```
> coxfit2 <- coxph(Surv(time_to_death, mortality_90days) ~
           intervention + tt(intervention), data=kidneyData,
           tt=function(x,t,...) x*(t<=21))
> summary(coxfit2)
```

```
                       coef exp(coef) se(coef)      z Pr(>|z|)
intervention        -0.4772    0.6205   0.2049 -2.329   0.0198 *
tt(intervention)     0.3929    1.4813   0.2379  1.652   0.0986 .
```

- So significant effect in period after 21 days.
- What is HR estimate in first 21 days?

6S: Alternative to proportional hazards

- You can also get the HR in period after day 21 by simply subsetting the data set:

```
> coxfit3 <- coxph(Surv(time_to_death,  mortality_90days) ~
intervention, data=subset(kidneyData, time_to_death>21))
> summary(coxfit3)
```

```
                coef exp(coef) se(coef)      z Pr(>|z|)
intervention -0.4772    0.6205   0.2049 -2.329   0.0198 *
--
```

- Not as easy to get HR before time 21 – WHY?