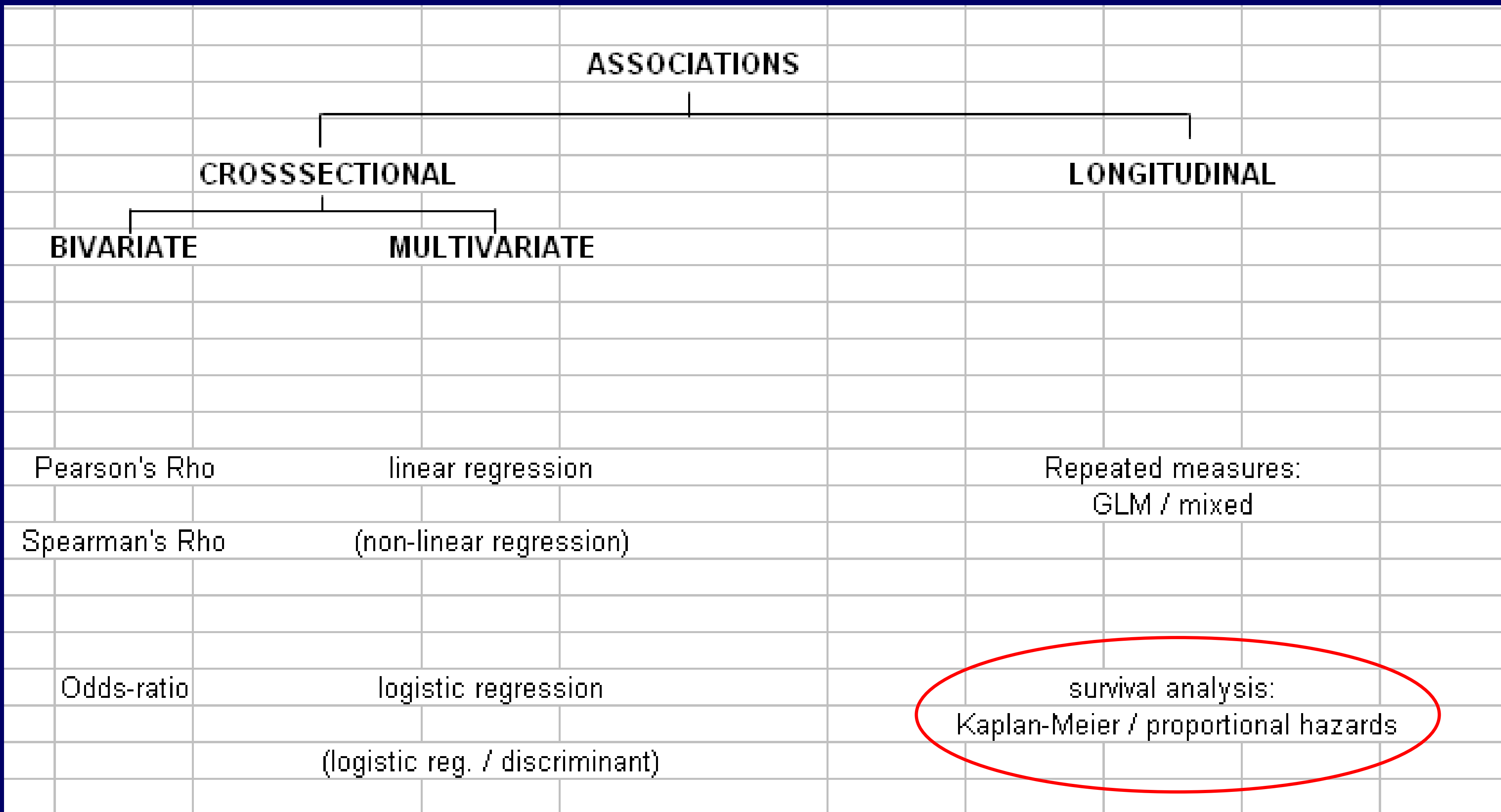
Practical biostatistics

Department of Clinical Epidemiology,
Biostatistics and Bioinformatics

AMC

Cox Survival Analyse

Day 8/9 Survival analysis



Cox-regression: between dis- and encouraging

Altman

- *This section is more complex than the others in the book.* (p. 387)
- *Expert statistical advice should be sought for carrying out Cox regression on survival data.* (p. 393)

Harris & Albert

- *The Cox model is basically very simple.* (p. 93)

What is survival analysis ?

- A serie of statistical procedures to analyse a specific type of data:

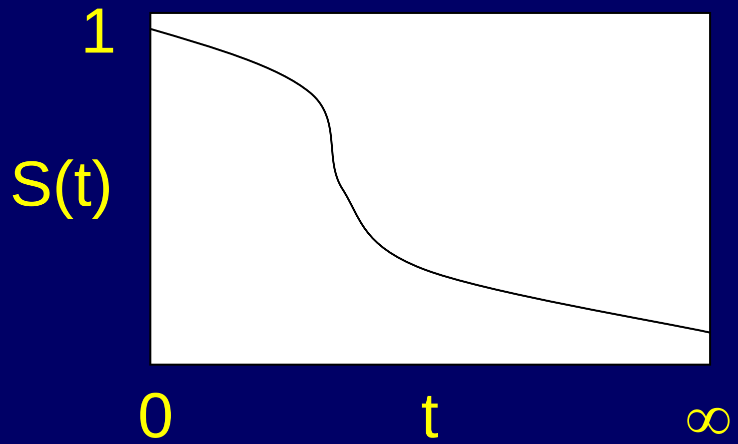
1. Outcome: time till event
2. Study design:
   start follow-up   TIME →   event
3. Event: deceased, sick, relapse, recovery, etc. (assumption: event occurs just once and is unique)
4. Time: years, months, days, weeks, seconds, ….

Time until event

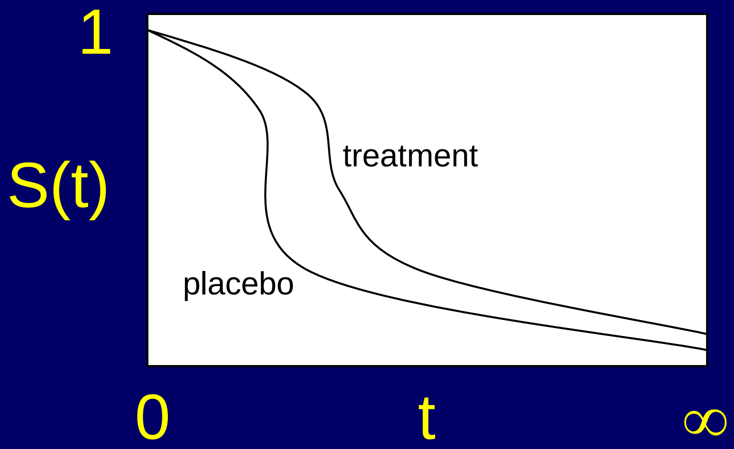
Start	**Event**	**"Survival"**
Anti-hypertensive drug	Myocard infarct	*Disease free period*
Antibiotic	Recovery	*Duration of the infection*
HIV-infection	Deceased	*Survival time*
Statine	Dementia	*Disease free period*

Aim of survival analysis

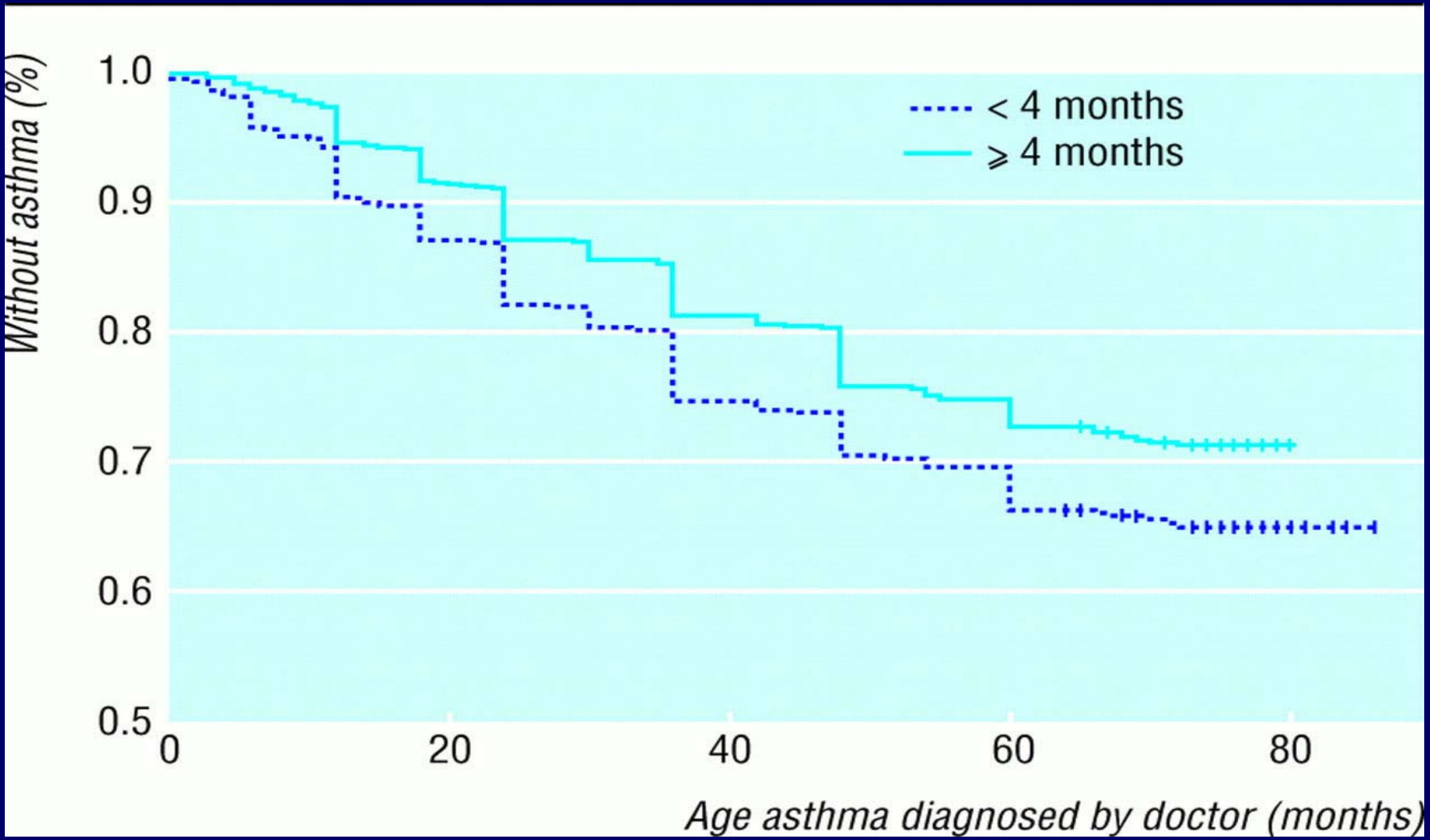
Estimation and interpretation of survival curves



Comparison of survival curves



Kaplan Meier survival functions for age at diagnosis of asthma stratified by duration of exclusive breast feeding (log rank statistic 10.70, df=1, P=0.001).

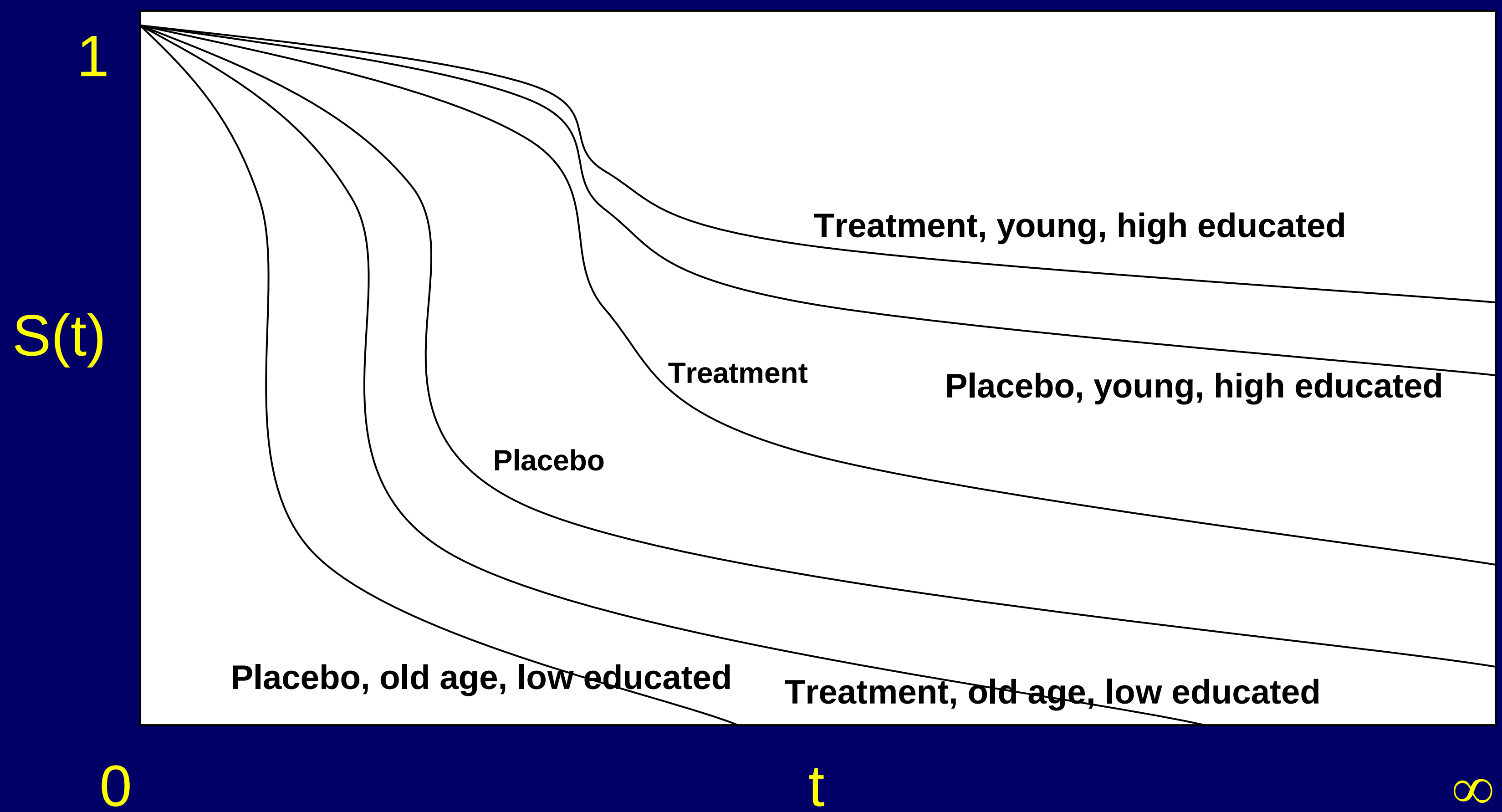


Vertical bars denote censoring events
Oddy, Holt, Sly, et al. BMJ 1999

Aim Cox regression analysis

To study the independent effects of a number of explanatory variables on survival

Several influences



Qualitative or quantitative

- Log-rank test:
  - Is there a difference, if so in what direction?
- Cox-regression:
  - How large is the difference and what is the influence of other predictive factors?

Regression analysis

Kind of regression	Dependent variable y	Outcome
Linear	continuous	regression coefficient β
Logistic	binary	odds ratio e^{β}
Cox	time until event (censoring)	hazard ratio e^{β}

Formulas

Linear regression:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$$

Logistic regression:

$$\ln(\text{odds}) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$$

$$OR = e^{\beta}$$

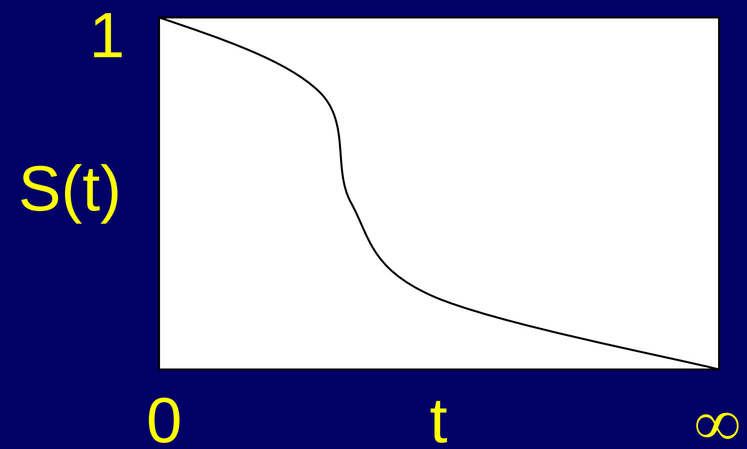
Cox regression:

$$\ln(h(t)) = \ln(h_0(t)) + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$$

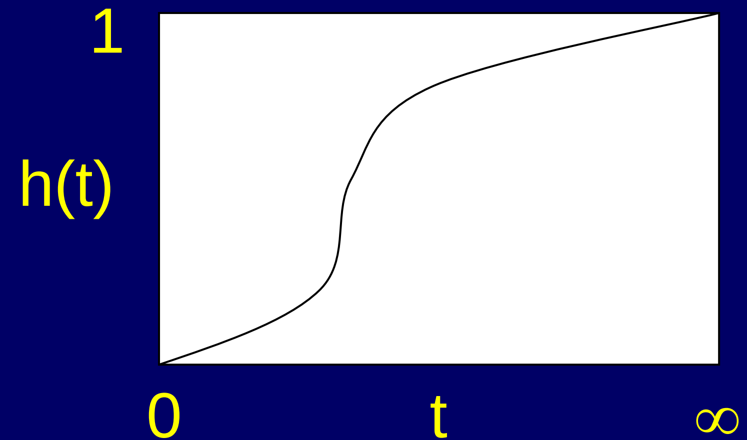
$$HR = e^{\beta}$$

Survival vs hazard

Cumulative
survival



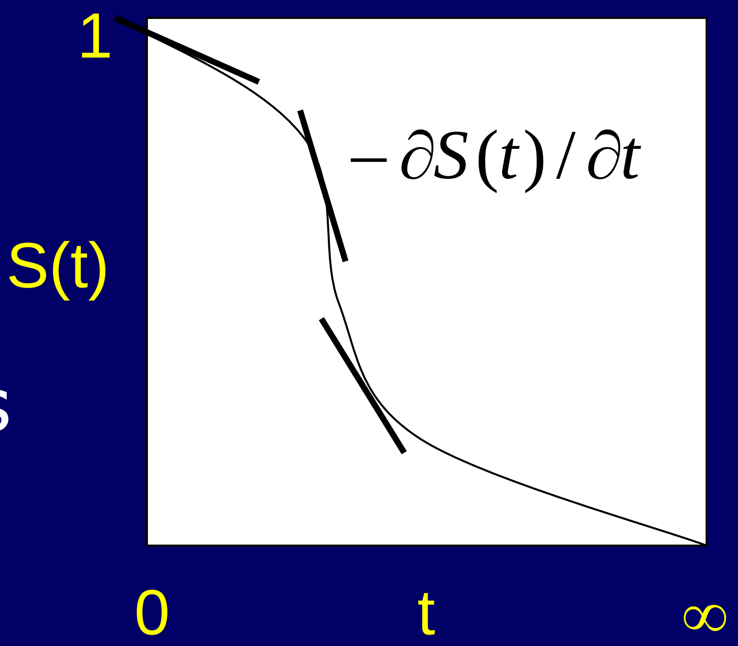
Cumulative
hazard rate



From S(t) to h(t); hazard rate

Hazard rate is the instant event rate in relation to the proportion survivors until time t, so:

$$h(t) = -\frac{\partial S(t)/\partial t}{S(t)} = -\frac{S'(t)}{S(t)}$$



Cox regression model (log-linear)

patient i

$$h_i(t) = h_0(t) * e^{\beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_p x_{pi}}$$

hazard function

predictors

basis hazard function

coefficients

Elements of the Cox-model

$$h_0(t)$$

- Basis hazard function
- Dependent of time
- Can take any form (non-parametric)
- Valid for an average patient for all predictors

$$e^{\beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_p x_{pi}}$$

- Coefficients and predictors
- Independent of time, so: constant
- regression coefficients have to be estimated (parametric) and tested for significance
- if $\beta=0$, no effect

Hazard ratio (HR): constant over time

IF:

$$h_i(t) = h_0(t) * e^{\beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_p x_{pi}}$$

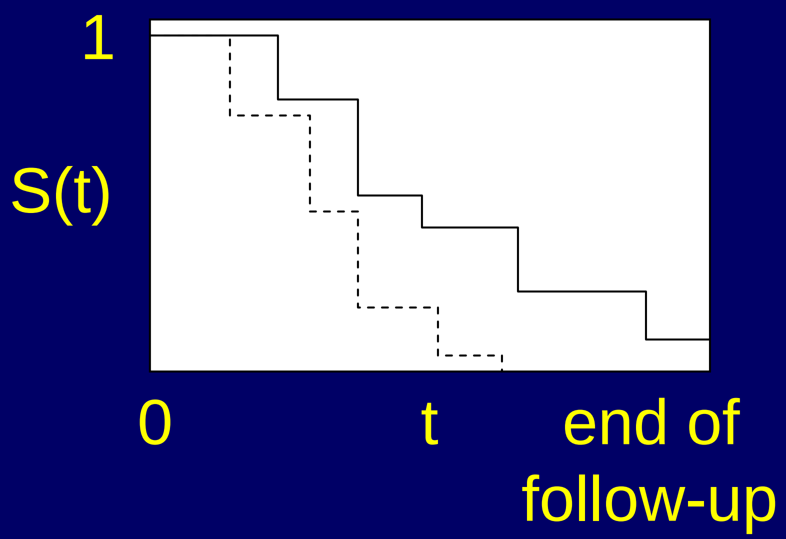
Then:

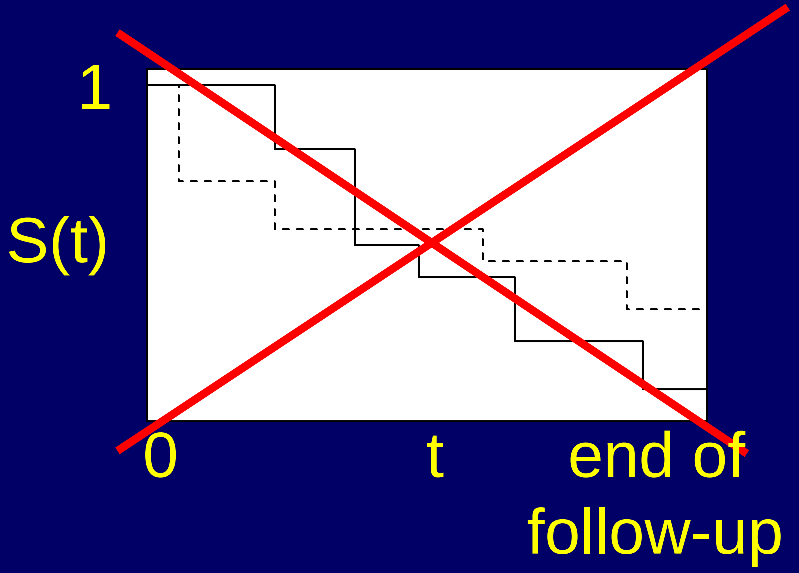
$$h_i(t)/h_0(t) = e^{\beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_p x_{pi}}$$

$$= \text{constant}$$

Assumption: proportional hazards (PH)

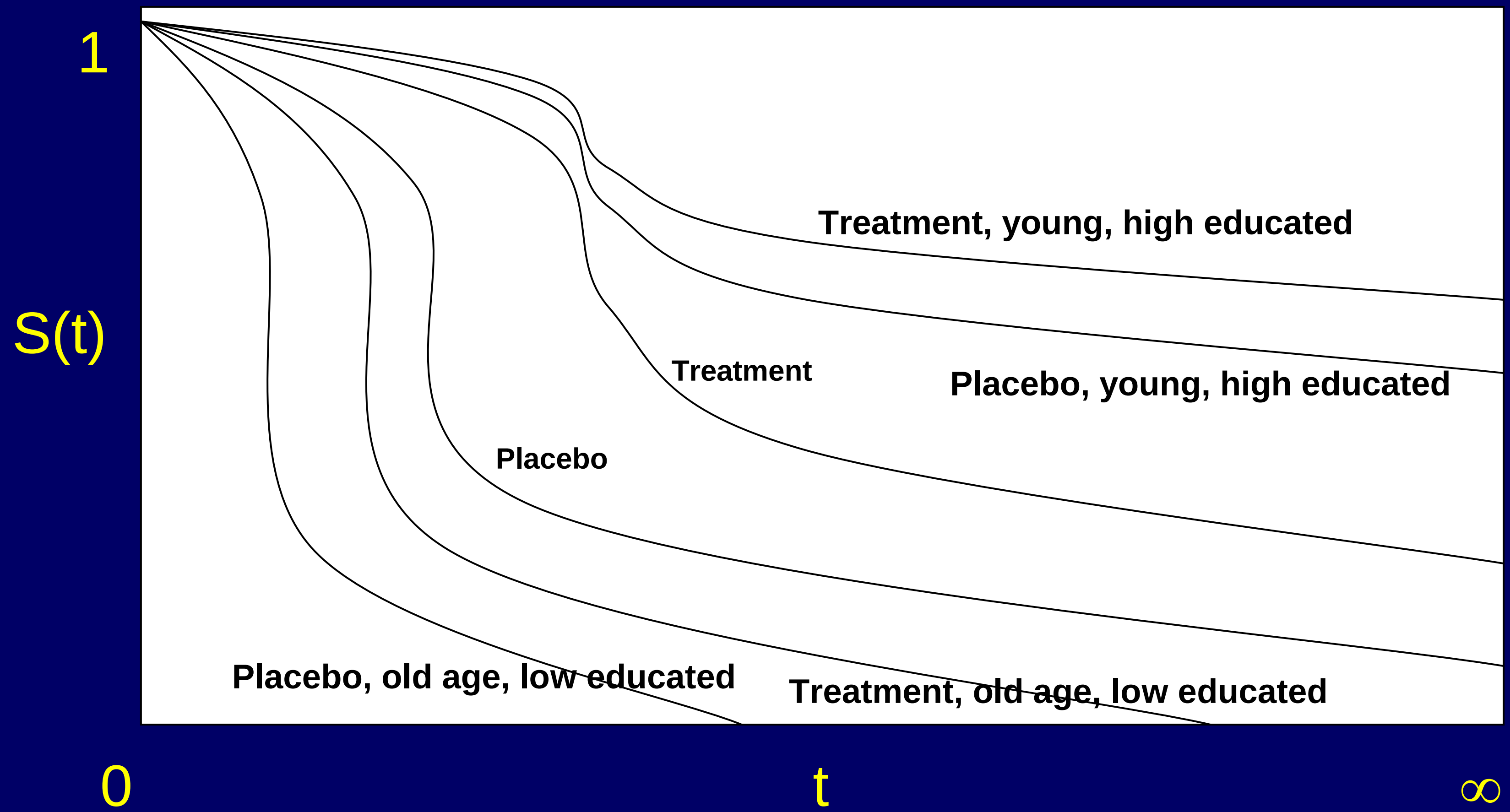
PH assumption





Assumption: proportional hazards (PH), so:
<u>survival curves are not allowed to cross each other</u>

Several influences



Example

1 $$h_i(t) = h_0(t) * e^{\beta_1 \text{treatment}_i + \beta_2 \text{age}_i + \beta_3 \text{education}_i}$$

2 $$h_i(t) = h_0(t) * e^{\beta_1 \text{treatment}_i + \beta_2 \text{age}_i}$$

3 $$h_i(t) = h_0(t) * e^{\beta_1 \text{treatment}_i}$$

Best model

- Comparison of the models with likelihood-ratio test:

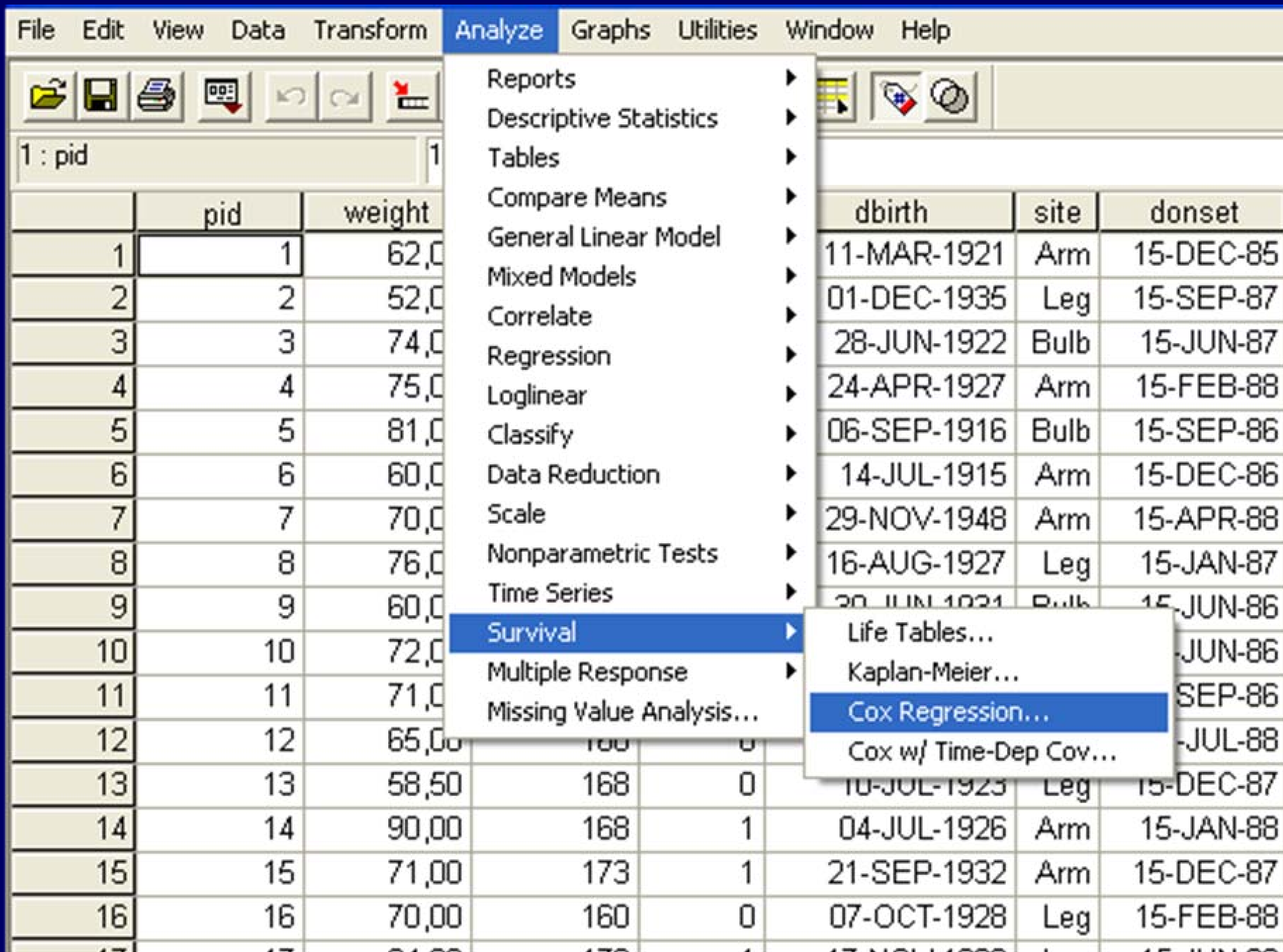
$$-2\log\frac{L_{\text{model1}}}{L_{\text{model2}}} = -2\left(\log L_{\text{model1}} - \log L_{\text{model2}}\right)$$

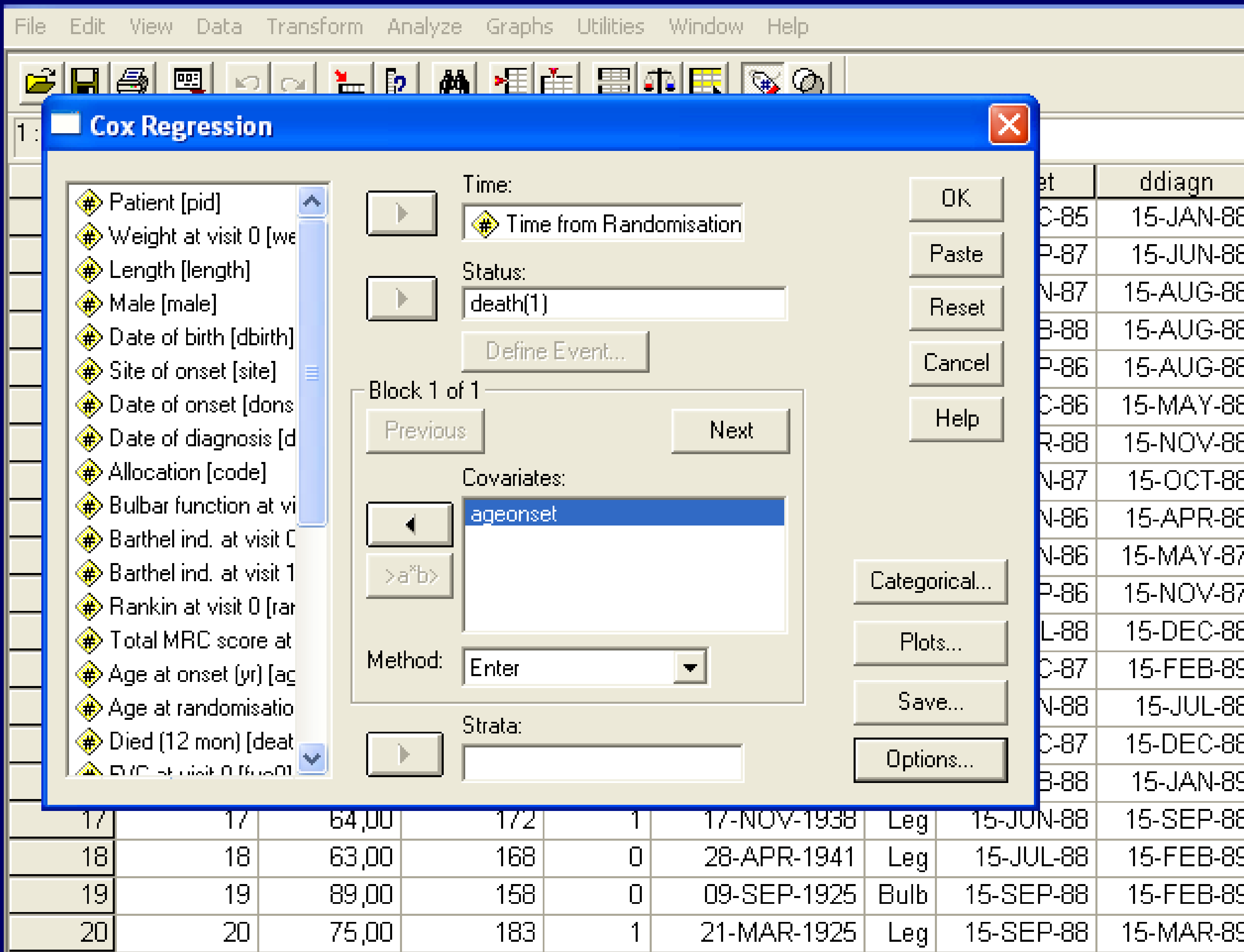
- This test behaves like a χ^2 with df=q,
- q is the number of additional covariates

Best model

Comparison of the models with likelihood-ratio test:

Closer to zero is better


File Edit View Data Transform Analyze Graphs Utilities Window Help
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Mixed Models
Correlate
Regression
Loglinear
Classify
Data Reduction
Scale
Nonparametric Tests
Time Series
Survival
Multiple Response
Missing Value Analysis...
Life Tables...
Kaplan-Meier...
Cox Regression...
Cox w/ Time-Dep Cov...
1 : pid
pid
weight
dbirth
site
donset


File Edit View Data Transform Analyze Graphs Utilities Window Help
Cox Regression
Patient [pid]
Weight at visit 0 [we
Length [length]
Male [male]
Date of birth [dbirth]
Site of onset [site]
Date of onset [dons
Date of diagnosis [d
Allocation [code]
Bulbar function at vi
Barthel ind. at visit 0
Barthel ind. at visit 1
Rankin at visit 0 [rar
Total MRC score at
Age at onset (yr) [ag
Age at randomisatio
Died (12 mon) [deat
Time:
Time from Randomisation
Status:
death(1)
Define Event...
Block 1 of 1
Previous
Next
Covariates:
ageonset
>a*b>
Method:
Enter
Strata:
OK
Paste
Reset
Cancel
Help
Categorical...
Plots...
Save...
Options...
ddiagn

Block 0: Beginning Block

Omnibus Tests of Model Coefficients

-2 Log Likelihood
720,496

Block 1: Method = Enter

Omnibus Tests of Model Coefficients[a,b]

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
710,499	9,324	1	,002	9,997	1	,002	9,997	1	,002

a. Beginning Block Number 0, initial Log Likelihood function: -2 Log likelihood: 720,496

b. Beginning Block Number 1. Method = Enter

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95,0% CI for Exp(B)	
							Lower	Upper
ageonset	,036	,012	9,193	1	,002	1,036	1,013	1,061

Interpretation HR

- Suppose, model 3 is the best description of 'reality', and β_1=-0.7. What is the hazard ratio (relative risk) for a patient treated with new treatment (treatment=1) and for a patient treated with placebo (treatment=0)?

$$\frac{h_1(t)}{h_2(t)} = \frac{h_0(t) * e^{-0.7*\text{treatment}_1}}{h_0(t) * e^{-0.7*\text{treatment}_2}}$$

$$= e^{-0.7*(\text{treatment}_1 - \text{treatment}_2)}$$

$$= e^{-0.7*(1-0)} = 0.496$$

Check PH assumption

- survival curves: no crossing !
- log - log survival curves: parallel !

Violation of PH

→ stratified Cox regression

→ introduction time-dependent variables

In summary

- Use Cox-regression to study the independent effects of explanatory variables on survival
- Results: Hazard ratio’s (relative risks)
- Choose best model using likelihood ratio test
- Check PH-assumption