UPPSALA UNIVERSITET

Analysis of Survival Data

Lecture 5:
Regression for survival data

$$\boxed{h(t \mid \mathbf{Z}) = h_0(t)\exp(\boldsymbol{\beta}^t\mathbf{Z})}$$

Cox, D.R. (1972), Regression Models and Life Tables, *Journal of the Royal Statistical Society,* B34

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{D} \underset{\uparrow}{L_i} = \prod_{i=1}^{D} \frac{\exp\left(\sum_{k=1}^{p}\beta_k Z_{(i)k}\right)}{\sum_{\underset{\uparrow}{j\in R(t_i)}} \exp\left(\sum_{k=1}^{p}\beta_k Z_{jk}\right)}$$

Partial likelihood for the *i*th event time

Risk set at time t_i

Inger Persson

Program L5

- **Regression for survival data**
  - Cox’s proportional hazards regression
    - Partial maximum likelihood
    - Interpretation of estimated coefficients
    - Continuous vs categorical covariates
    - Ties
    - Local tests
    - Model building
    - Time-dependent covariates

Comparing two or more groups

When the groups to be compared are similar, except for the grouping variable, the K samples tests (Log-rank, Gehan’s, Gray’s test, etc.) can be used.

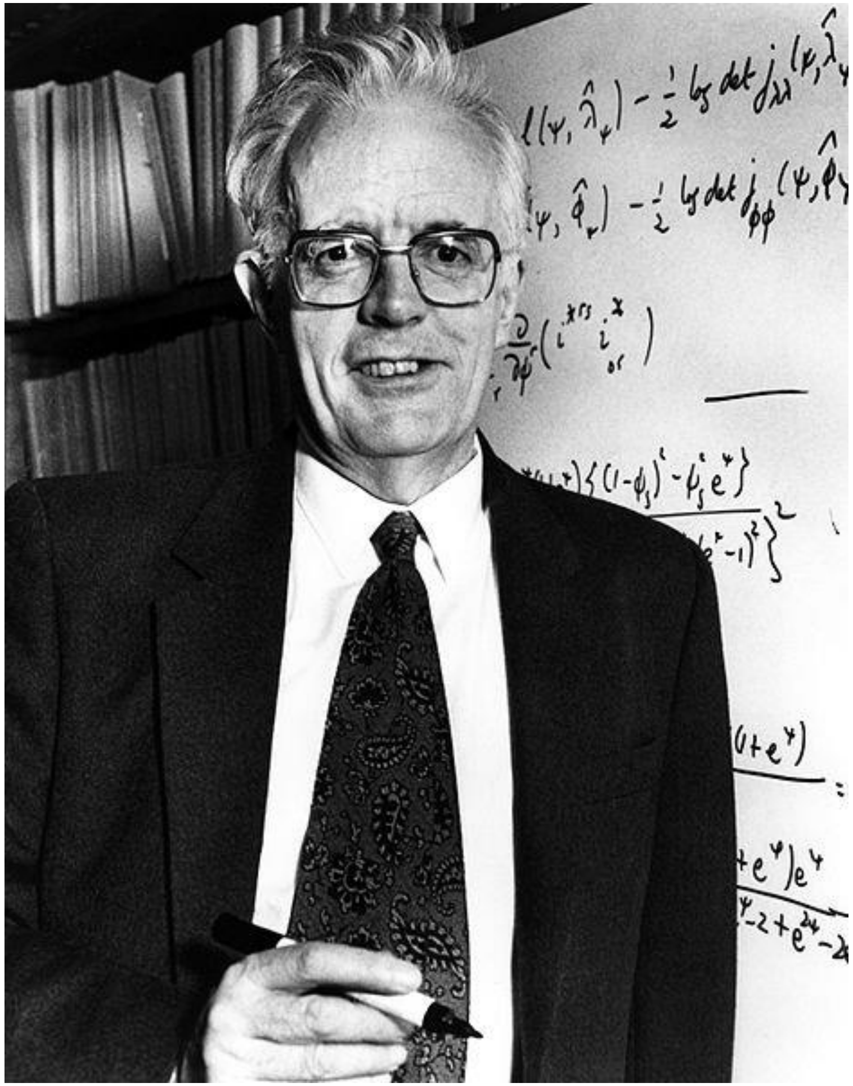
When there are other covariates that affect the event rates in the K different populations, stratified tests can be used. But stratified tests will not provide information about the size of the effect that the covariates might have on the outcome.

Alternative: regression

Regression models

Continuous data	Linear regression
Dichotomous data	Logistic regression
No. of events	Poisson regression
Survival data	Cox regression

Sir David Cox



Cox, D.R. (1972), Regression Models and Life Tables, *Journal of the Royal Statistical Society,* B34

Still one of the most frequently cited journal articles in statistics and medicine.

Cox regression (a.k.a. proportional hazards regression, PH regression)

X = time to some event (e.g. survival time)

Data:

$(T_j, \delta_j, \mathbf{Z}_j(t))$ $\qquad j = 1, \ldots, n$

T = observed time (time to event or censoring)

δ = event indicator (1=event, 0=right censored obs.)

$\mathbf{Z}(t)$= vector of p covariates

We'll start focusing on fixed covariates (that do not depend on t)

Cox's proportional hazards model

Vector of regression parameters

Vector of covariates (explanatory variables)

$$h(t\,|\,\mathbf{Z}) = \underset{\uparrow}{h_0}(t)\underbrace{c(\overset{\downarrow}{\boldsymbol{\beta}}{}^t\,\overset{\swarrow}{\mathbf{Z}})}_{\text{Known function}}$$

Baseline hazard

The baseline hazard rate $h_0(t)$ is an unknown (arbitrary) function, giving the hazard function for the standard set of conditions $\mathbf{Z} = \mathbf{0}$.

A semi-parametric model

Making special assumptions about the baseline hazard $h_0(t)$ leads to parametric models, e.g. the exponential and Weibull distributions.

The advantage of Cox’ model is the fact that such assumptions can be avoided.

Cox’s approach is said to be **semi-parametric**.

UPPSALA
UNIVERSITET

Common choice of *c*

$$h(t \mid \mathbf{Z}) = h_0(t)c(\boldsymbol{\beta}^t\mathbf{Z})$$

$h(t \mid \mathbf{Z}) > 0$ → *c* is chosen so it never can be negative

Common model for $c(\boldsymbol{\beta}^t\mathbf{Z})$: $\exp(\boldsymbol{\beta}^t\mathbf{Z})$

↓

$$\boxed{h(t \mid \mathbf{Z}) = h_0(t)\exp(\boldsymbol{\beta}^t\mathbf{Z})}$$

UPPSALA UNIVERSITET

Example: 3 explanatory variables

p = 3 covariates (explanatory variables)

$$\boldsymbol{\beta}^t = (\beta_1\ \beta_2\ \beta_3) \qquad \mathbf{Z}(t) = \begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \end{pmatrix}$$

$$\boldsymbol{\beta}^t\mathbf{Z} = \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3 = \sum_{k=1}^{p} \beta_k Z_k$$

$$h(t \mid \mathbf{Z}) = h_0(t)\exp\left(\sum_{k=1}^{3} \beta_k Z_k\right)$$

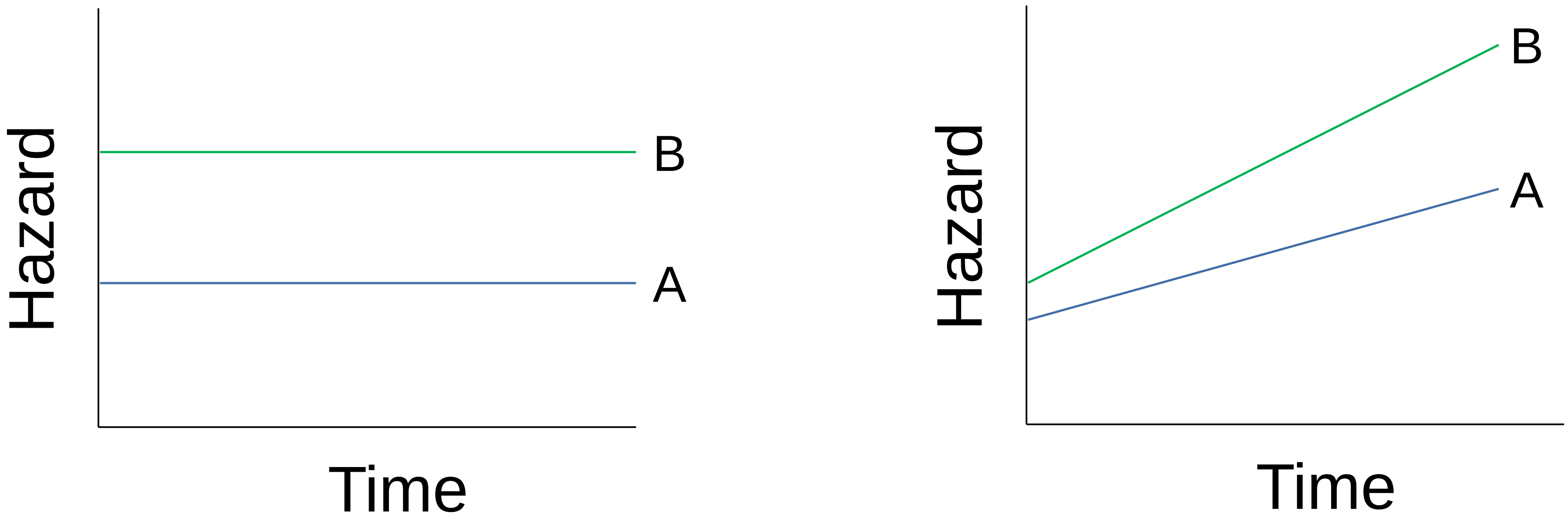
Comparing two individuals (proportional hazards)

One way to compare two individuals with covariate values $\mathbf{Z}$ and $\mathbf{Z}^*$:

$$\frac{h(t\,|\,\mathbf{Z})}{h(t\,|\,\mathbf{Z}^*)} = \frac{\cancel{h_0(t)}\exp\left(\sum_{k=1}^{p}\beta_k Z_k\right)}{\cancel{h_0(t)}\exp\left(\sum_{k=1}^{p}\beta_k Z_k^*\right)} = \exp\left(\sum_{k=1}^{p}\beta_k (Z_k - Z_k^*)\right)$$

The ratio between two hazards is constant, independent of t. This means that the hazard rates are **proportional**.

Proportional hazards



The ratio between the two hazards is constant.

UPPSALA UNIVERSITET

Relative risk

$\exp\left(\sum_{k=1}^{p}\beta_k(Z_k - Z_k^*)\right)$ is called the **relative risk** (hazard ratio)

Describes the risk of an individual with covariates $\mathbf{Z}$ experiencing the event, compared to an individual with covariates $\mathbf{Z}^*$.

Example: Relative risk for women compared to men

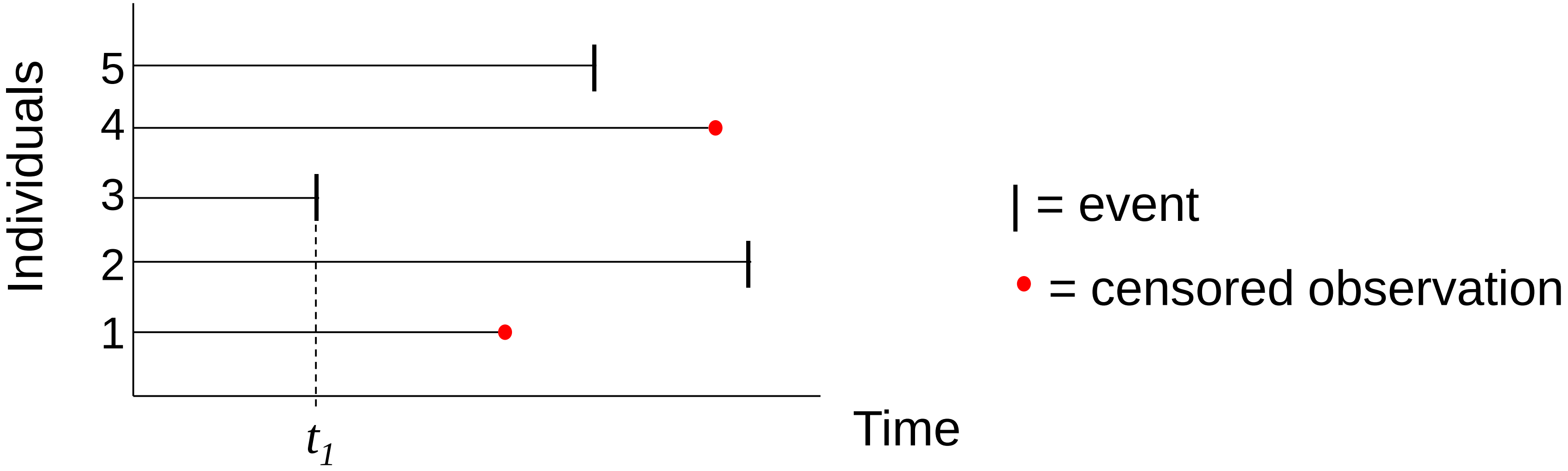
$$Z_1 = \begin{cases} 0 & \text{if man} \\ 1 & \text{if woman} \end{cases}$$

If all other covariates have the same value, the hazard ratio

$$\frac{h(t \mid \mathbf{Z})}{h(t \mid \mathbf{Z}^*)} = \exp(\beta_1)$$

describes the risk of experiencing the event for women compared to the risk for men

Example: estimating the parameters β



Given that a person experiences the event at t_1, what is the probability that it is individual no. 3?

We want a model that gives a large probability that it is individual no. 3.

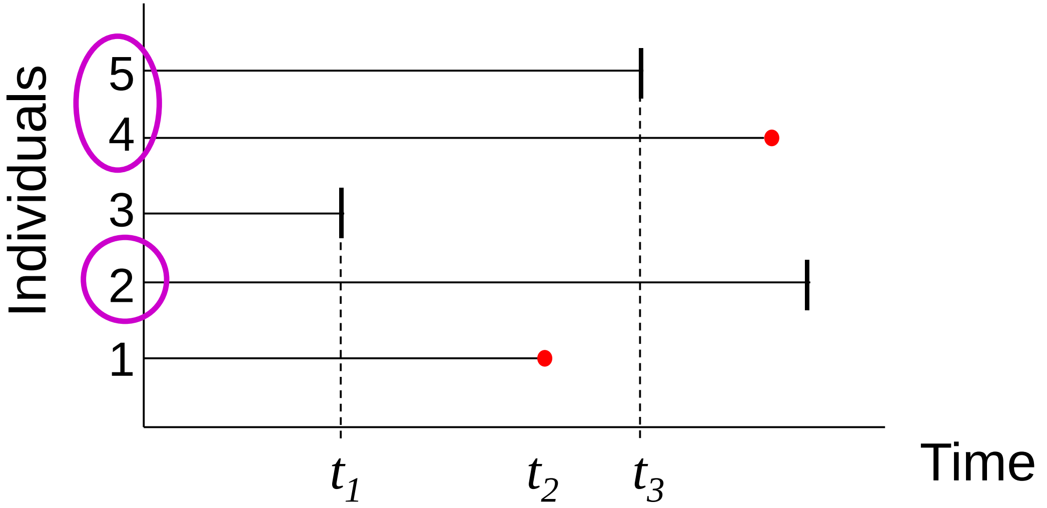
Example: estimating the parameters β

h_3 = the probability that individual no. 3 experiences the event at t_1, given that he/she has not experienced the event before that.

$$L_1 = \frac{h_3}{h_1 + h_2 + h_3 + h_4 + h_5} \longleftarrow \text{The risk set}$$

= probability (likelihood) that no. 3 experiences the event, compared to all the individuals at risk

Example: estimating the parameters β



| = event

● = censored observation

There is a term in the likelihood for each event, not for each individual.

Next event is at time t_3.

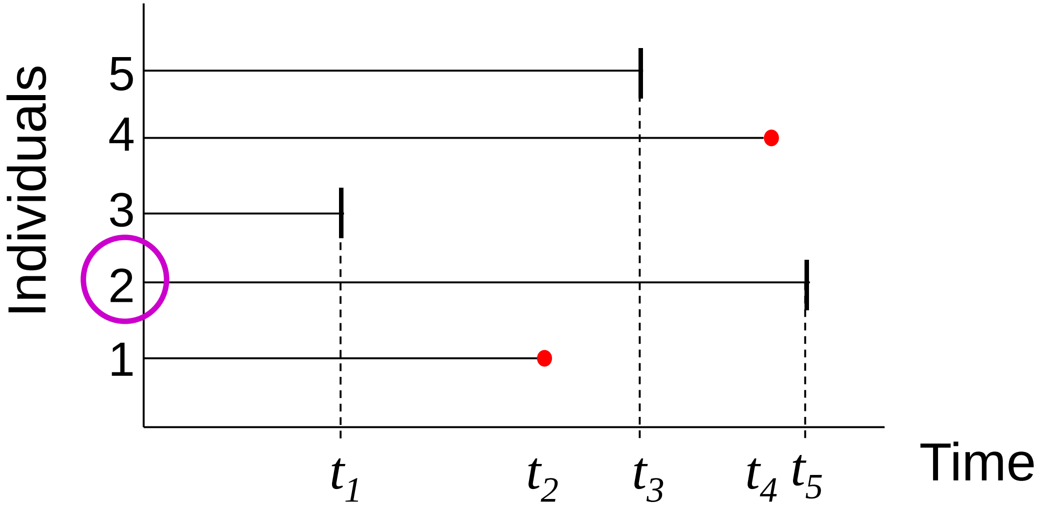
At that time there are three individuals still at risk.

Example: estimating the parameters β

$$L_2 = \frac{h_5}{h_2 + h_4 + h_5}$$

= probability (likelihood) that individual no. 5 experiences the event, compared to all the individuals still at risk

Example: estimating the parameters β



| = event

● = censored observation

Next event is at time t_5.

At that time there is only one individual still at risk.

Example: estimating the parameters β

$$L_3 = \frac{h_2}{h_2} = 1$$

= probability (likelihood) that individual no. 2 experiences the event, compared to all the individuals still at risk

The “partial likelihood”

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{D} \underset{\uparrow}{L_i} = \prod_{i=1}^{D} \frac{\exp\left(\sum_{k=1}^{p} \beta_k Z_{(i)k}\right)}{\underset{\uparrow}{\sum_{j \in R(t_i)}} \exp\left(\sum_{k=1}^{p} \beta_k Z_{jk}\right)}$$

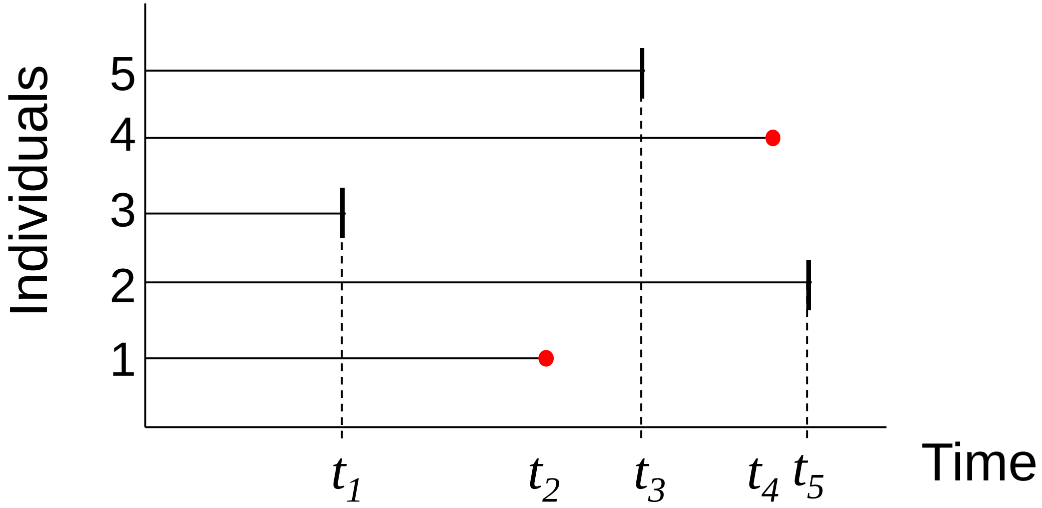
Partial likelihood for
the *i*th event time

Risk set
at time t_i

Inference for β is based on this **partial likelihood**.

This partial likelihood method estimates β only, not the hazard.

Example: partial maximum likelihood



Suppose p = 1

$$Z = \begin{cases} 0 & \text{if man} \\ 1 & \text{if woman} \end{cases}$$

Suppose individuals no. 1, 3, and 5 are women.

Example: partial maximum likelihood

First event time, individual 3: $h(t_1|Z) = h_0(t_1)e^{\beta Z_3} = h_0(t_1)e^{\beta}$

$$L_1 = \frac{h_3}{h_1 + h_2 + h_3 + h_4 + h_5} = \frac{e^{\beta}}{e^{\beta} + 1 + e^{\beta} + 1 + e^{\beta}} = \frac{e^{\beta}}{2 + 3e^{\beta}}$$

Second event time (t_3), individual 5:

$$L_2 = \frac{h_5}{h_2 + h_4 + h_5} = \frac{e^{\beta}}{1 + 1 + e^{\beta}} = \frac{e^{\beta}}{2 + e^{\beta}}$$

Estimating the parameters β: (partial) maximum likelihood

Estimates of β are found by maximizing the partial likelihood

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{D} \frac{\exp\left(\sum_{k=1}^{p} \beta_k Z_{(i)k}\right)}{\sum_{j \in R(t_i)} \exp\left(\sum_{k=1}^{p} \beta_k Z_{jk}\right)}$$

Even though this is not a likelihood in the traditional sense, it is treated as one, and inference is carried out by usual means.

Estimating the parameters β: (partial) maximum likelihood

The maximization of the partial likelihood (solving for the most likely values of β) cannot be done analytically, numerical methods must be employed.

The maximization can be done using a Newton-Raphson technique (or some other iterative method for optimization).

Estimating the parameters β: (partial) maximum likelihood

The partial likelihood does not depend upon the baseline hazard rate $h_0(t)$, which means that inference on the effects of explanatory variables (covariates) can be made without any knowledge about the baseline hazard.

In these analyses, the baseline hazard $h_0(t)$ is treated as a nuisance parameter function.

Characteristics of Cox proportional hazards regression

- Does not require that you choose some particular probability model to represent survival times, and is therefore more robust than parametric methods
- Semi-parametric (parametric assumptions can be avoided)
- Can accommodate both discrete and continuous measures of event times

Characteristics of Cox proportional hazards regression, cont’d

- Easy to incorporate time-dependent covariates (covariates that may change in value over the course of the observation period)
- Cox regression models the effect of covariates on the hazard rate but leaves the baseline hazard rate unspecified
- Estimates relative rather than absolute risk

Assumptions

Assumptions of the Cox model:

- random sample(s) (for inference)
- independent observations
- noninformative censoring
- right censored or left truncated data
- large sample (common rule of thumb: ≥ 10 events/cov.)*
- proportional hazards

**NOTE: This is a recommendation, not a strict rule. See e.g.*

- *Peduzzi et al., Importance of Events Per Independent Variable in Proportional Hazards regression Analysis, J Clin Epidemiol, 1995; Vol 48, No. 12*
- *Vittinghoff & McCulloch, Relaxing the Rule of Ten Events per Variable in Logistic and Cox Regression, Am J Epidemiol, 2006; Vol. 165, No. 6*

Proportional hazards assumption

The Cox regression model assumes that the hazard rates are proportional.

This assumption can be checked by a number of tests and graphical methods (we’ll learn more later on).

UPPSALA
UNIVERSITET

Program L5

- **Regression for survival data**
  - Cox's proportional hazards regression
    - Partial maximum likelihood
    - Interpretation of estimated coefficients
    - Continuous vs categorical covariates
    - Ties
    - Local tests
    - Model building
    - Time-dependent covariates

Example: Hodgkin’s disease

A small study comparing the effectiveness of allogeneic transplants versus autogeneic transplants for Hodgkin’s disease or non-Hodgkin’s lymphoma is presented in section 1.10.

Allogeneic transplant: from a matching donor

Autogeneic transplant: your own bone marrow is cleansed and returned after a high dose of chemotherapy

Is there a difference in disease-free survival between allogeneic and autogeneic transplants?

Example: Hodgkin’s disease

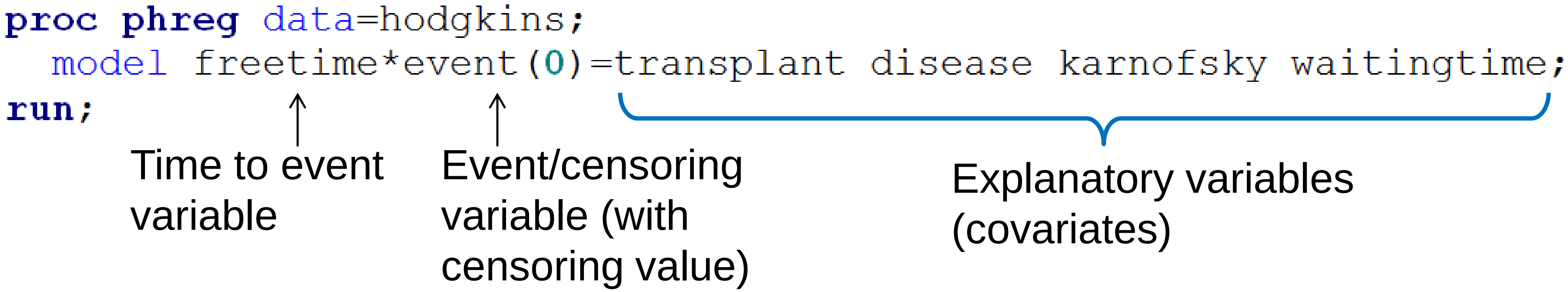
Variables:

freetime = time to death or relapse (days)
transplant = type of transplant (0=allogeneic, 1=autogeneic)
event = event indicator (1=dead or relapse,
0=alive without relapse)
disease = disease type (0=non-Hodgkin’s lymphoma,
1=Hodgkin’s disease)
karnofsky = pretransplant Karnofsky score, 0-100
(higher score = less functional impairment)
waitingtime = waiting time from diagnosis to transplant
(months)

UPPSALA
UNIVERSITET

Example: Hodgkin’s disease

To estimate the hazard ratios using the Cox proportional hazards model, use the **phreg** procedure.



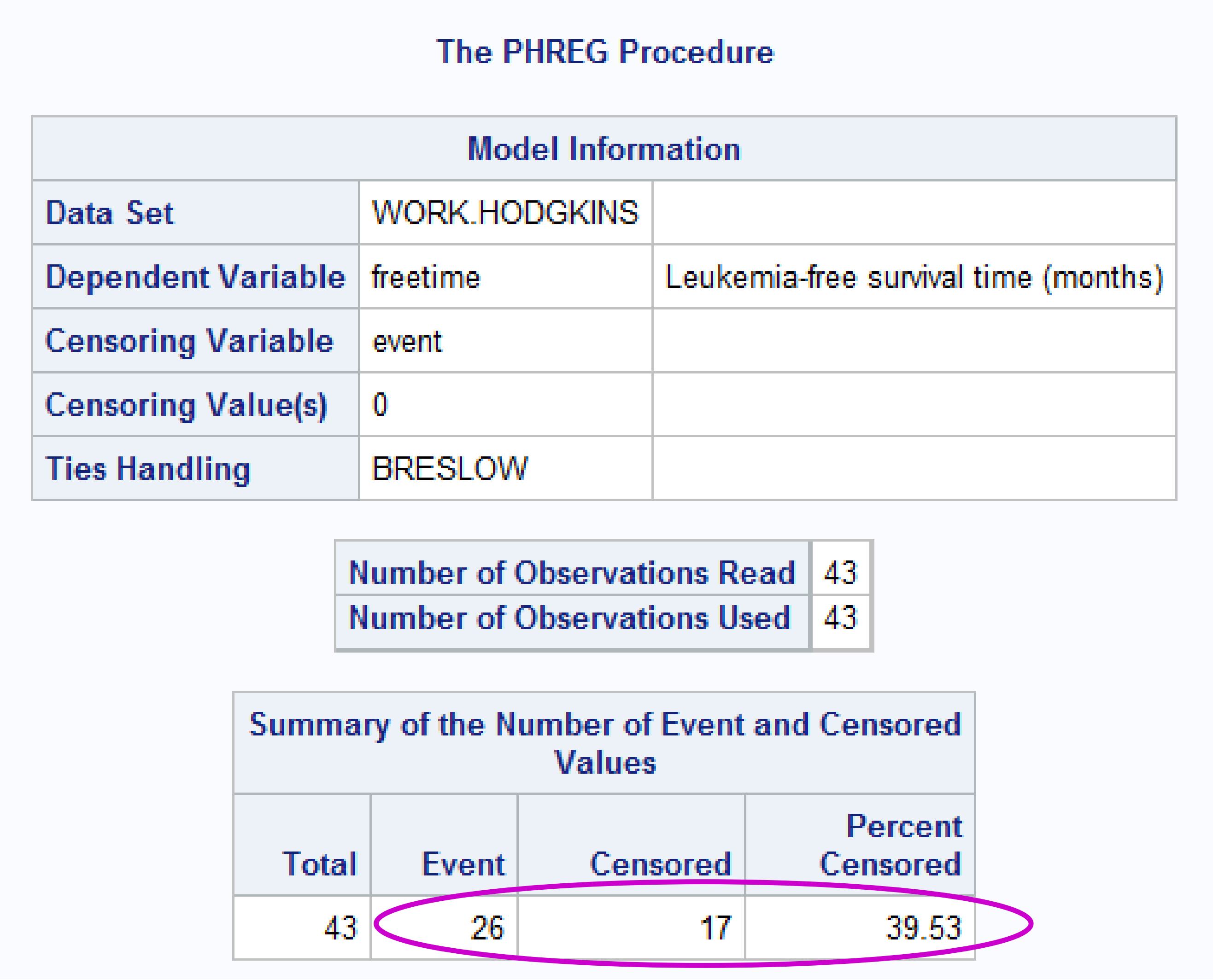
```
proc phreg data=hodgkins;
  model freetime*event(0)=transplant disease karnofsky waitingtime;
run;
```

Time to event variable

Event/censoring variable (with censoring value)

Explanatory variables (covariates)

Example: Hodgkin’s disease



The PHREG Procedure

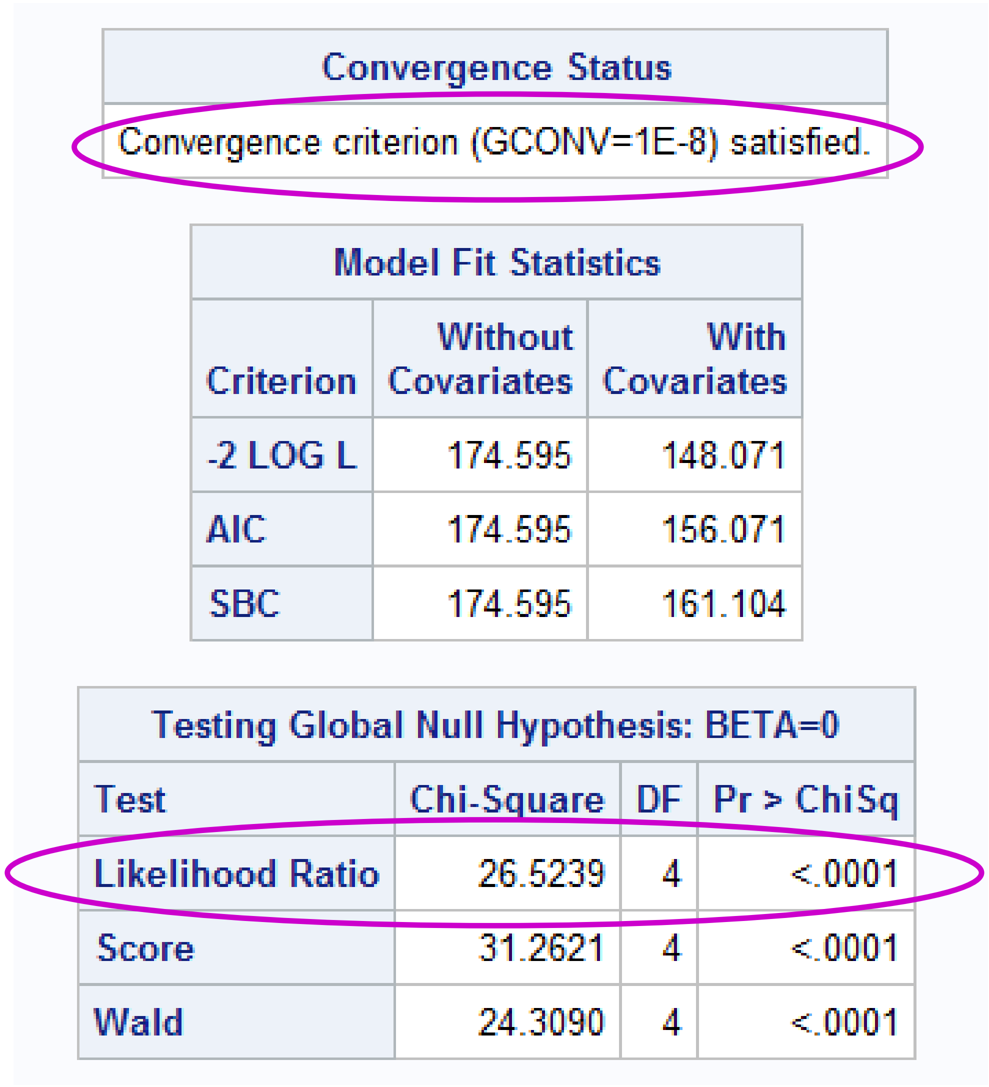
Model Information		
Data Set	WORK.HODGKINS	
Dependent Variable	freetime	Leukemia-free survival time (months)
Censoring Variable	event	
Censoring Value(s)	0	
Ties Handling	BRESLOW	

Number of Observations Read	43
Number of Observations Used	43

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
43	26	17	39.53

Example: Hodgkin’s disease



Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	174.595	148.071
AIC	174.595	156.071
SBC	174.595	161.104

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	26.5239	4	<.0001
Score	31.2621	4	<.0001
Wald	24.3090	4	<.0001

Convergence criterion

The iterative process of maximizing the (partial) likelihood is declared converged when the relative change in log likelihoods between successive steps is less than 0.0001.

Global tests

Three common tests of $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ (e.g. $H_0 : \boldsymbol{\beta} = 0$)

1) **Likelihood ratio test**, based on the likelihood ratio (how many times more likely the data are under the model with compared to without covariates)

2) **Wald's test**, based on asymptotic normality of the (partial) maximum likelihood estimates

3) **Scores test**, based on efficient scores (same scores used when finding the partial maximum likelihood estimates), asymptotically *p*-variate normal.

Global tests

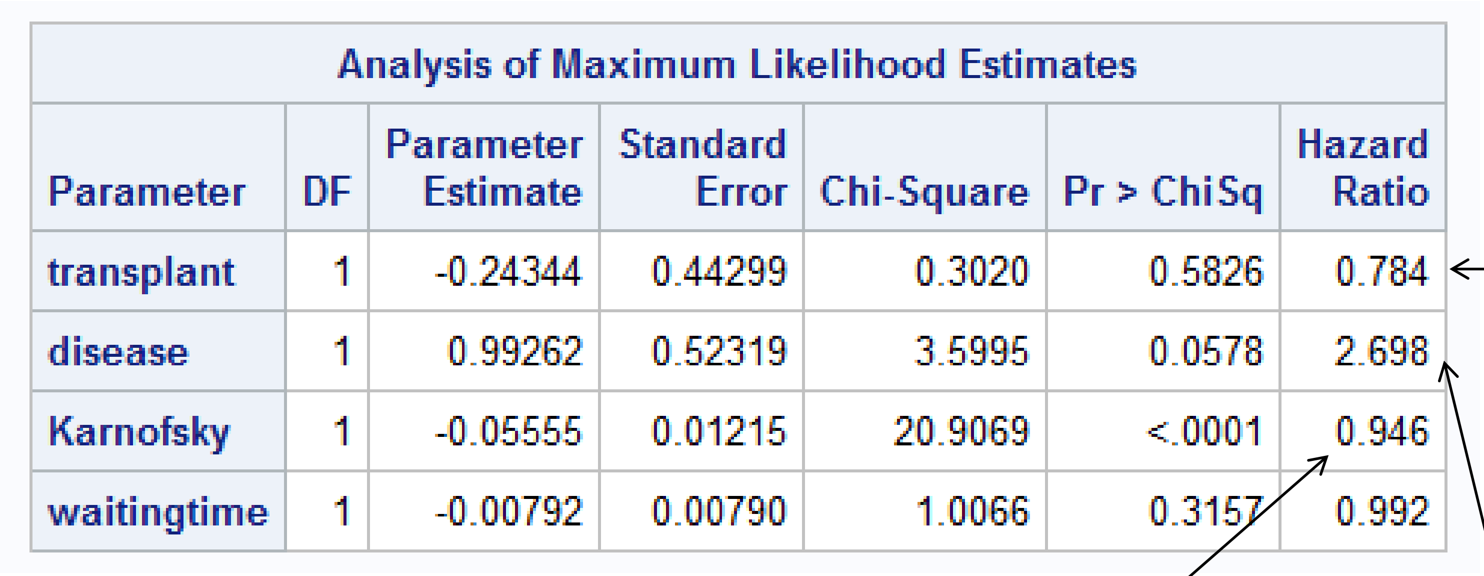
The scores test is identical to the log-rank test if there are no ties between event times.

The Wald and Scores tests are both approximations of (and asymptotically equivalent to) the Likelihood ratio test.

With regards to size (α) and power, Li et al (1996) showed that the likelihood ratio test outperforms the Wald test especially for small samples. The scores test is not recommended, it tends to inflate the size of the test.

All three tests assume that the hazard rates are proportional.

Example: Hodgkin’s disease



Analysis of Maximum Likelihood Estimates						
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
transplant	1	-0.24344	0.44299	0.3020	0.5826	0.784
disease	1	0.99262	0.52319	3.5995	0.0578	2.698
Karnofsky	1	-0.05555	0.01215	20.9069	<.0001	0.946
waitingtime	1	-0.00792	0.00790	1.0066	0.3157	0.992

Hazard ratio = 0.784

Interpretation:

The risk of dying or relapsing for autogeneic transplanted patients is 78.4% of the same risk for allogeneic transplanted, on average.

Hazard ratio = 0.946

Interpretation:

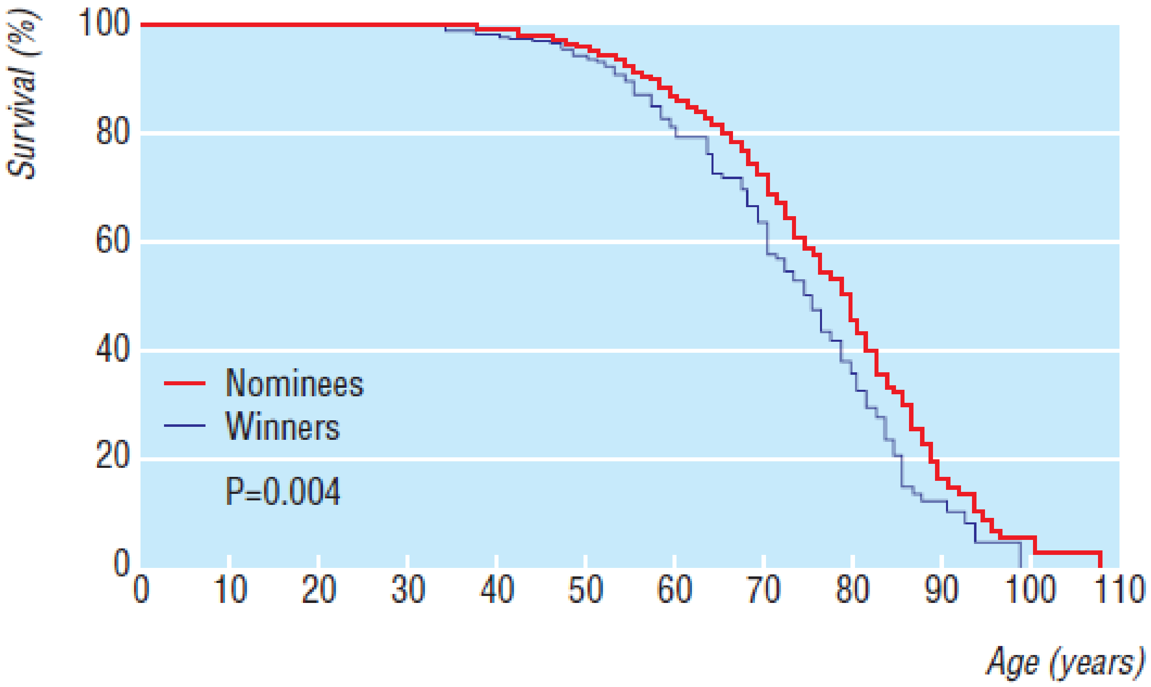
The risk of dying or relapsing decreases by 5.4% with each one-unit increase in the Karnofsky score, on average. Equivalently, the risk decreases by 24.2% with each 5-unit increase of Karnofsky score, on average ($0.946^5 = 0.758$).

Hazard ratio = 2.698

Interpretation:

The risk of dying or relapsing is 2.7 times higher for patients with Hodgkin’s disease than for patients with Non-Hodkin’s lymphoma, on average.

Example: Study of longevity (life length) in academy award winning screenwriters



Survival of winners and nominees of academy awards for screenwriting. The graph shows the percentage of each group alive, plotted by using the Kaplan-Meier technique. Primary statistical analysis is based on a log rank test comparing winners to nominees (n=185, deaths=112 and n=610, deaths=316, respectively)

p. 1494

Ref: D.A. Redelmeier, S.M. Singh (2001). *Longevity of screenwriters who win an academy award: longitudinal study.* BMJ 2001; 323: 1491-6

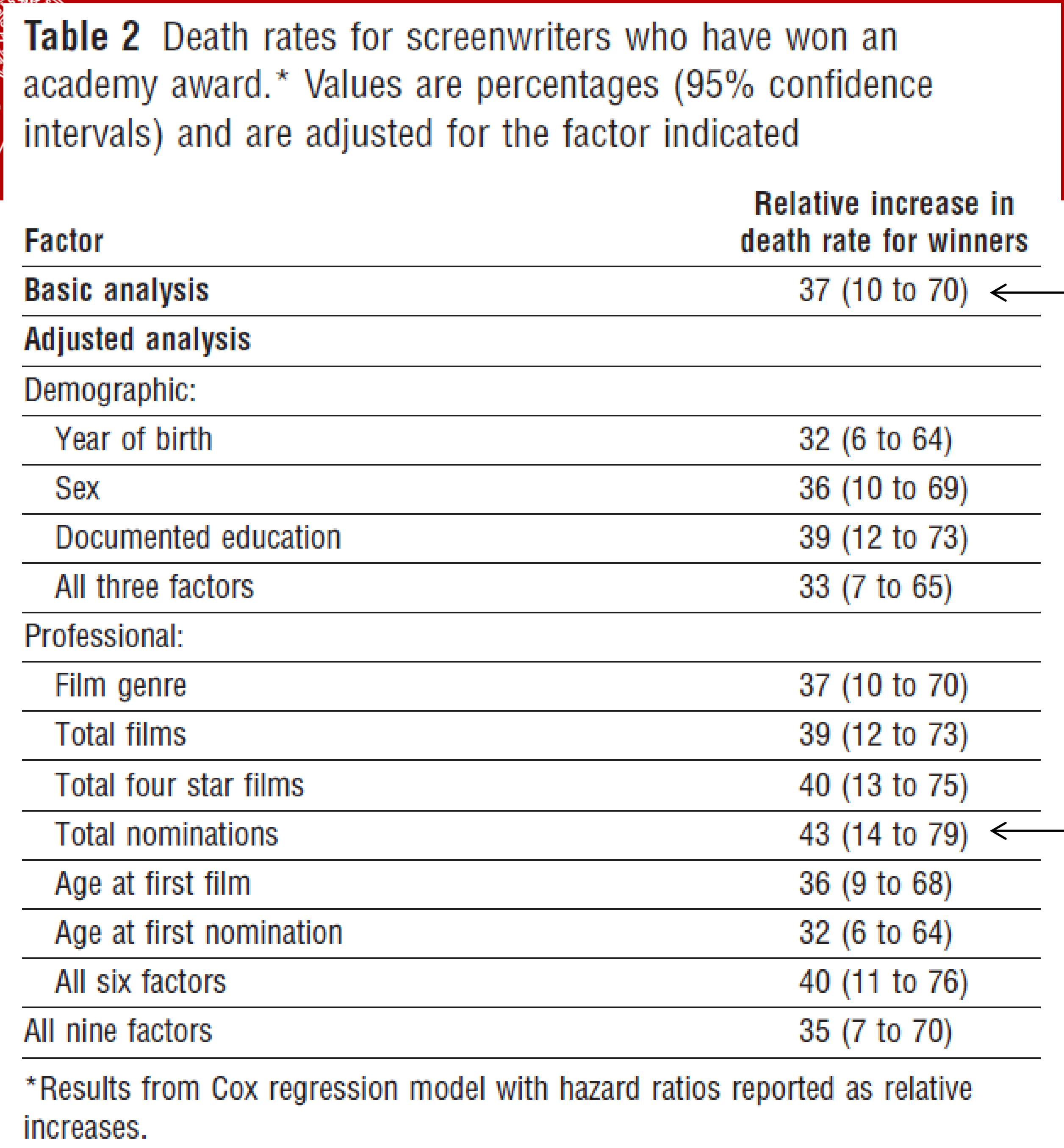


Table 2 Death rates for screenwriters who have won an academy award.* Values are percentages (95% confidence intervals) and are adjusted for the factor indicated

Factor	**Relative increase in death rate for winners**
Basic analysis	37 (10 to 70)
Adjusted analysis	
Demographic:	
Year of birth	32 (6 to 64)
Sex	36 (10 to 69)
Documented education	39 (12 to 73)
All three factors	33 (7 to 65)
Professional:	
Film genre	37 (10 to 70)
Total films	39 (12 to 73)
Total four star films	40 (13 to 75)
Total nominations	43 (14 to 79)
Age at first film	36 (9 to 68)
Age at first nomination	32 (6 to 64)
All six factors	40 (11 to 76)
All nine factors	35 (7 to 70)

*Results from Cox regression model with hazard ratios reported as relative increases.

Hazard ratio = 1.37

Interpretation:

The risk of dying is on average 37% higher for winners compared to nominees

Hazard ratio = 1.43

Interpretation:

The risk of dying is on average 43% higher for every extra nomination

Hazard ratio confidence intervals

Confidence intervals can of course be calculated around hazard ratios estimated by Cox proportional hazards regression.

Two methods available:

1) **Wald** (standard confidence intervals). May work poorly for maximum-likelihood estimation.

2) **Profile-likelihood**, inverts a likelihood-ratio test to obtain a CI for the parameter in question. Applicable to all likelihood-based statistical analyses.

Example: Hodgkin’s disease

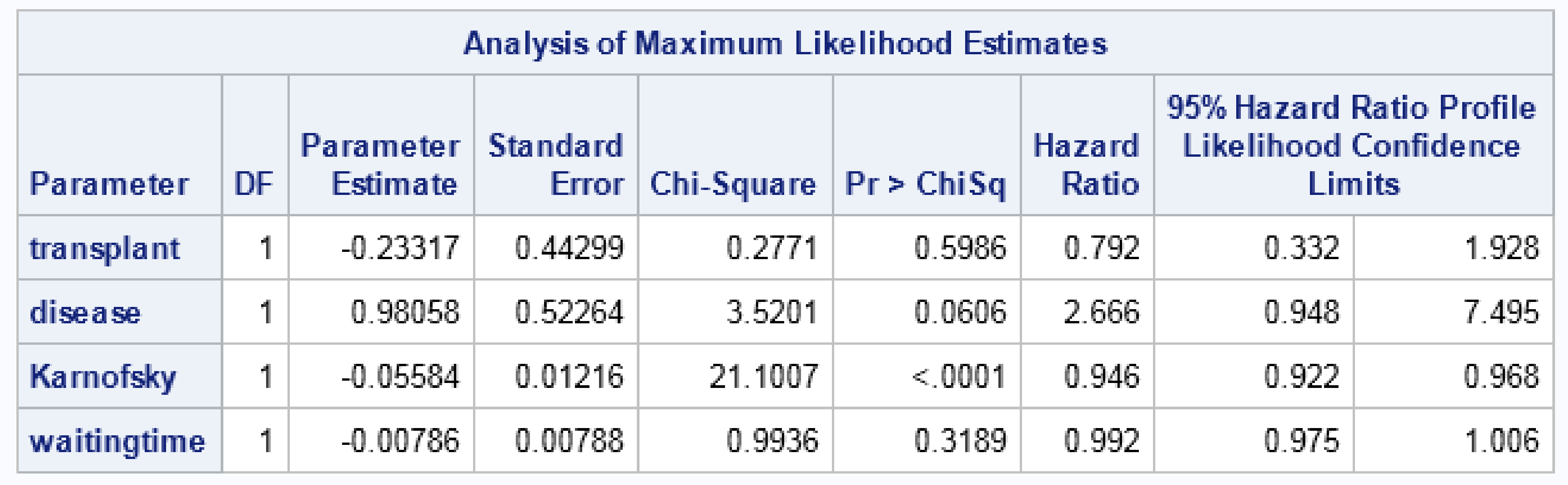
Use proc phreg and the **risklimit** option.

```
proc phreg data=hodgkins;
  model freetime*event(0)=transplant disease karnofsky
            waitingtime/risklimit=pl;
run;
```

Produces confidence intervals for hazard ratios

Profile-likelihood intervals

Example: Hodgkin’s disease



Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Profile Likelihood Confidence Limits	
transplant	1	-0.23317	0.44299	0.2771	0.5986	0.792	0.332	1.928
disease	1	0.98058	0.52264	3.5201	0.0606	2.666	0.948	7.495
Karnofsky	1	-0.05584	0.01216	21.1007	<.0001	0.946	0.922	0.968
waitingtime	1	-0.00786	0.00788	0.9936	0.3189	0.992	0.975	1.006

UPPSALA
UNIVERSITET

Program L5

- **Regression for survival data**
  - Cox's proportional hazards regression
    - Partial maximum likelihood
    - Interpretation of estimated coefficients
    - Continuous vs categorical covariates
    - Ties
    - Local tests
    - Model building
    - Time-dependent covariates

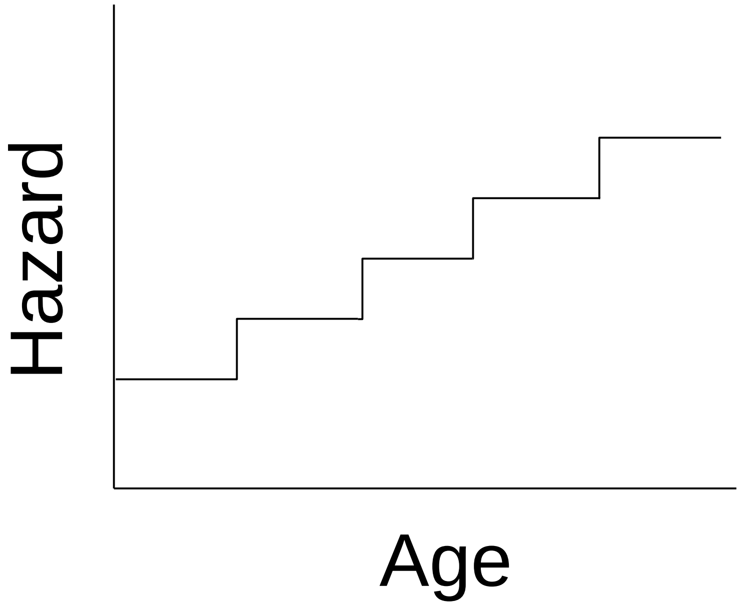
Continuous vs. categorical variables

Continuous variables can be used as they are in the regression model.

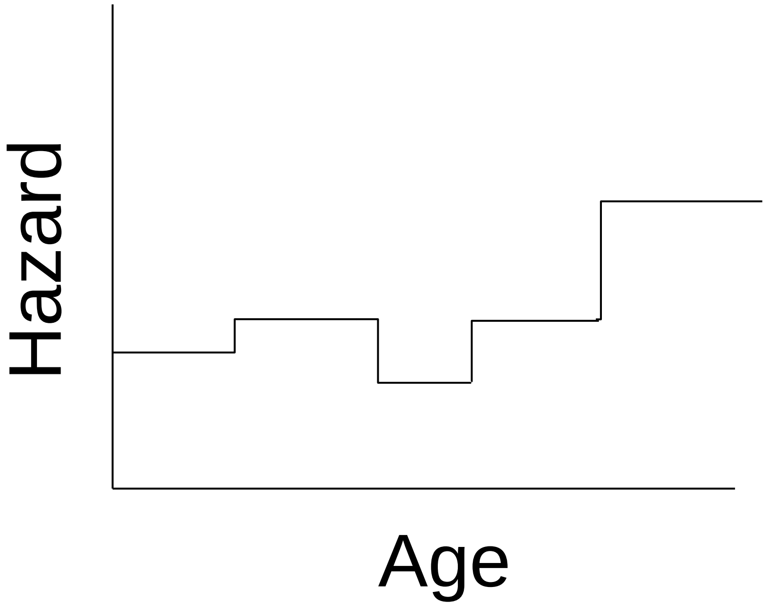
The regression estimate of β will then describe the effect of a one-unit increase of the explanatory variable.

If the hazard does not increase/decrease continuously it might be better to categorize the covariate.

Example: Age covariate



Continuously increasing hazard



Not continuously increasing hazard

Dummy variables and interaction effects

Dummy variables (for categorical variables) and interaction variables can be used with Cox's proportional hazards model, just as for linear models.

Categorization of covariates

There are different ways of categorizing a continuous variable.

One way is to divide into groups of equal size (the same number of observations in each group).

The optimal strategy is to determine cut points based on scientific reasoning.

UPPSALA UNIVERSITET

Avoid dichotomization of covariates

Avoid dichotomization (using only two categories)!

Dichotomizing is a way of effectively losing a great deal of information, with a serious loss of power to detect real relationships.

Dichotomizing may also increase the probability of false positive results.

Further reading: P. Royston, D.G. Altman, W. Sauerbrei (2006). Dichotomizing continuous preditors in multiple regression: a bad idea. *Statistics in medicine*, 2006; 25: 127-141.

Transformation of variables

Keep the continuous variable continuous, if possible, to avoid loss of information.

Variables can be transformed (using logarithms, squares, etc) to find a stronger relationship between explanatory and dependent variables.

Ties

Events occur at D times, $t_1 < t_2 < \cdots < t_D$

At time t_i there are d_i events (there can be ties between event times).

Common methods of constructing the partial likelihood when ties are present:

1) Exact
2) Breslow (approximation)
3) Efron (approximation)
4) Cox’s discrete

“Exact” method of handling ties

Time is treated as a continuous variable, and ties are assumed being a result of imprecise measurements of time.

Assumes there is a true unknown order of events in time.

Calculates the exact probability of all possible orderings of events.

Complex computations, but usually no noticeable extra computer time.

Kalbfleisch, J. D. and Prentice, R. L. (1980), *The Statistical Analysis of Failure Time Data*, New York: John Wiley & Sons

UPPSALA UNIVERSITET

Breslow’s method of handling ties

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{D} \frac{\exp\left(\boldsymbol{\beta}^t \mathbf{s}_i\right)}{\left(\sum_{j \in R_i} \exp\left(\boldsymbol{\beta}^t \mathbf{Z}_j\right)\right)^{d_i}}$$

Set of all individuals at risk just prior to time t_i

$$\mathbf{s}_i = \sum_{j \in \mathrm{D}_i} \mathbf{Z}_j$$

Set of all individuals who experience the event at time t_i

Works well when there are few ties.

The default method in SAS.

Efron’s method of handling ties

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{D} \frac{\exp\left(\boldsymbol{\beta}^t \mathbf{s}_i\right)}{\prod_{j=1}^{d_i}\left(\sum_{k\in R_i}\exp\left(\boldsymbol{\beta}^t \mathbf{Z}_k\right) - \frac{j-1}{d_i}\sum_{k\in \mathrm{D}_i}\exp\left(\boldsymbol{\beta}^t \mathbf{Z}_k\right)\right)}$$

Set of all individuals at risk just prior to time t_i

$$\mathbf{s}_i = \sum_{j\in \mathrm{D}_i} \mathbf{Z}_j$$

Set of all individuals who experience the event at time t_i

Closer to the correct partial likelihood based on a discrete hazard model than Breslow’s likelihood.

Similar to Breslow’s likelihood when there are few ties.

Cox’s discrete method of handling ties

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{D} \frac{\exp\left(\boldsymbol{\beta}^t \mathbf{s}_i\right)}{\sum_{q\in Q_i} \exp\left(\boldsymbol{\beta}^t \mathbf{s}_q^*\right)}$$

Set of all subsets of d_i individuals who could be selected from the risk set R_i

$$\mathbf{s}_i = \sum_{j\in \mathrm{D}_i} \mathbf{Z}_j$$

Set of all individuals who experience the event at time t_i

$$\mathbf{s}_q^* = \sum_{j=1}^{d_j} \mathbf{Z}_{qj}$$

Assumes time is discrete (not very common in reality).

Gives exact estimates, no approximations are used.

A logistic model is assumed for the hazard rate, models proportional odds (not hazards).

Choice of ties handling method

When there are no ties, all methods give exactly the same results.

When there are few ties, the choice of method has a very small impact on the results.

When there are many ties, the Breslow and Efron approximations give poor results (coefficients are biased towards 0)

UPPSALA UNIVERSITET

Choice of ties handling method, cont’d

Base the choice of method on substantive grounds – are the tied events truly tied, or are they a result of imprecise measurement?

Prefer discrete or exact method over Breslow and Efron approximations.

If you have to use approximations: Prefer Efron’s method over Breslow’s.

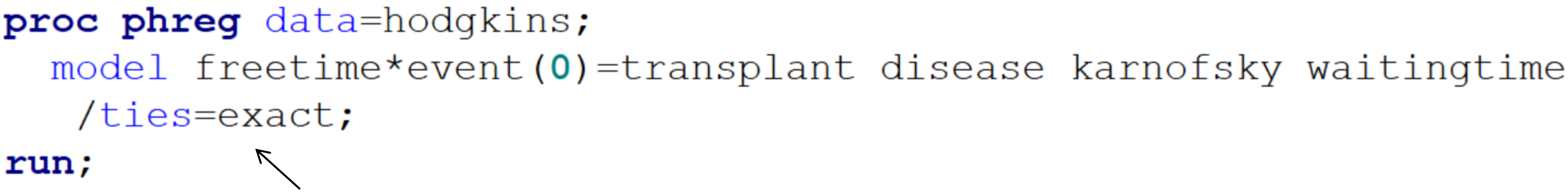
Example: Hodgkin’s disease

Variables:

freetime = time to death or relapse (days)
transplant = type of transplant (0=allogeneic, 1=autogeneic)
event = event indicator (1=dead or relapse,
0=alive without relapse)
disease = disease type (0=non-Hodgkin’s lymphoma,
1=Hodgkin’s disease)
karnofsky = pretransplant Karnofsky score, 0-100
(higher score = less functional impairment)
waitingtime = waiting time from diagnosis to transplant
(months)

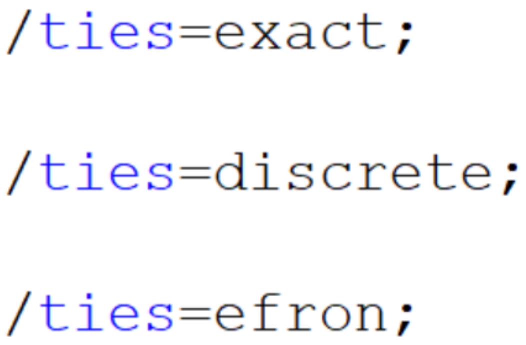
Example: Hodgkin’s disease

To choose the tie handling method, use the **ties** option.



```
proc phreg data=hodgkins;
  model freetime*event(0)=transplant disease karnofsky waitingtime
   /ties=exact;
run;
```

Choice of method

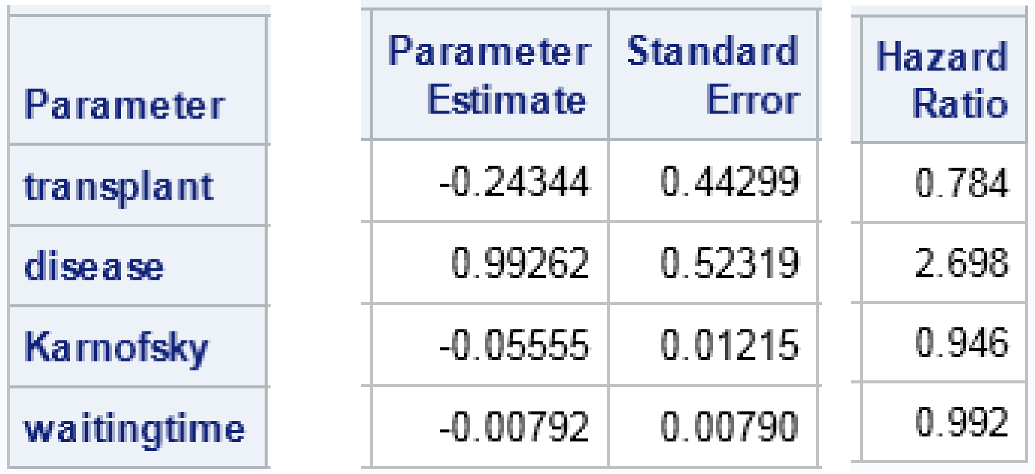


```
/ties=exact;

/ties=discrete;

/ties=efron;
```

UPPSALA UNIVERSITET

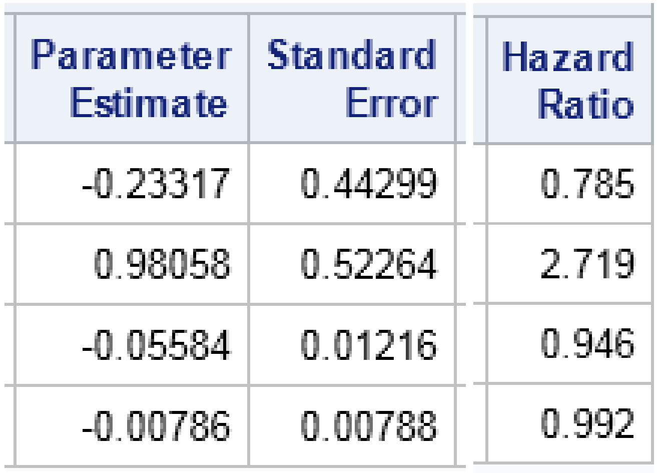
Example: Hodgkin’s disease

Breslow



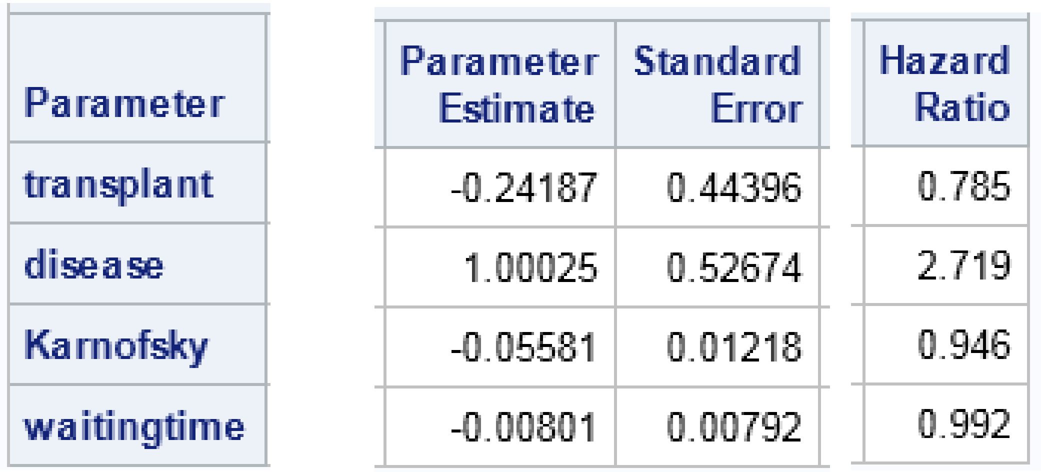
Parameter	Parameter Estimate	Standard Error	Hazard Ratio
transplant	-0.24344	0.44299	0.784
disease	0.99262	0.52319	2.698
Karnofsky	-0.05555	0.01215	0.946
waitingtime	-0.00792	0.00790	0.992

Efron



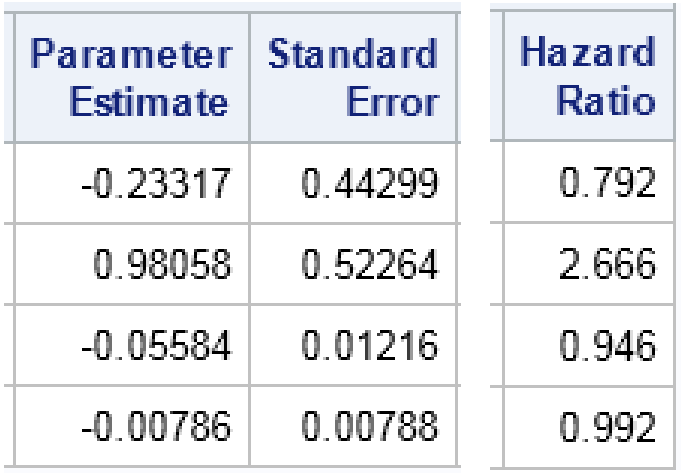
Parameter Estimate	Standard Error	Hazard Ratio
-0.23317	0.44299	0.785
0.98058	0.52264	2.719
-0.05584	0.01216	0.946
-0.00786	0.00788	0.992

Discrete (Cox)

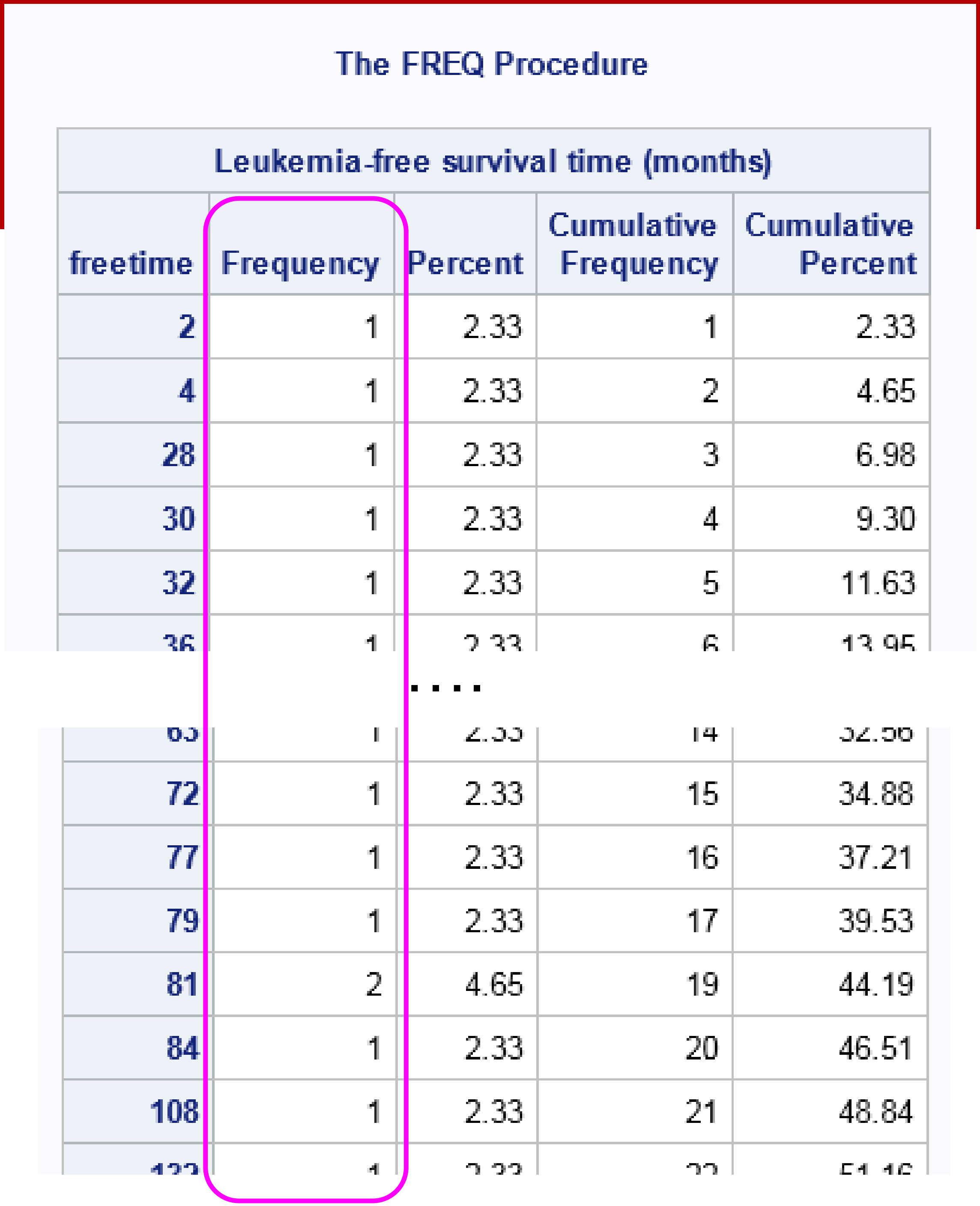


Parameter	Parameter Estimate	Standard Error	Hazard Ratio
transplant	-0.24187	0.44396	0.785
disease	1.00025	0.52674	2.719
Karnofsky	-0.05581	0.01218	0.946
waitingtime	-0.00801	0.00792	0.992

Exact



Parameter Estimate	Standard Error	Hazard Ratio
-0.23317	0.44299	0.792
0.98058	0.52264	2.666
-0.05584	0.01216	0.946
-0.00786	0.00788	0.992



The FREQ Procedure

Leukemia-free survival time (months)				
freetime	Frequency	Percent	Cumulative Frequency	Cumulative Percent
2	1	2.33	1	2.33
4	1	2.33	2	4.65
28	1	2.33	3	6.98
30	1	2.33	4	9.30
32	1	2.33	5	11.63
36	1	2.33	6	13.95
		....		
65	1	2.33	14	32.56
72	1	2.33	15	34.88
77	1	2.33	16	37.21
79	1	2.33	17	39.53
81	2	4.65	19	44.19
84	1	2.33	20	46.51
108	1	2.33	21	48.84
122	1	2.33	22	51.16

Program L5

- **Regression for survival data**
  - Cox’s proportional hazards regression
    - Partial maximum likelihood
    - Interpretation of estimated coefficients
    - Continuous vs categorical covariates
    - Ties
    - Local tests
    - Model building
    - Time-dependent covariates

Local tests

Local hypotheses can be tested, e.g.

$$H_0 : \beta_1 = 0 \, , \beta_2 = 0, \text{etc.}$$

$$H_0 : \beta_1 = \beta_3$$

$$H_0 : \beta_1 = \beta_3 = \beta_5$$

The Wald test is used in SAS (OK for large samples).

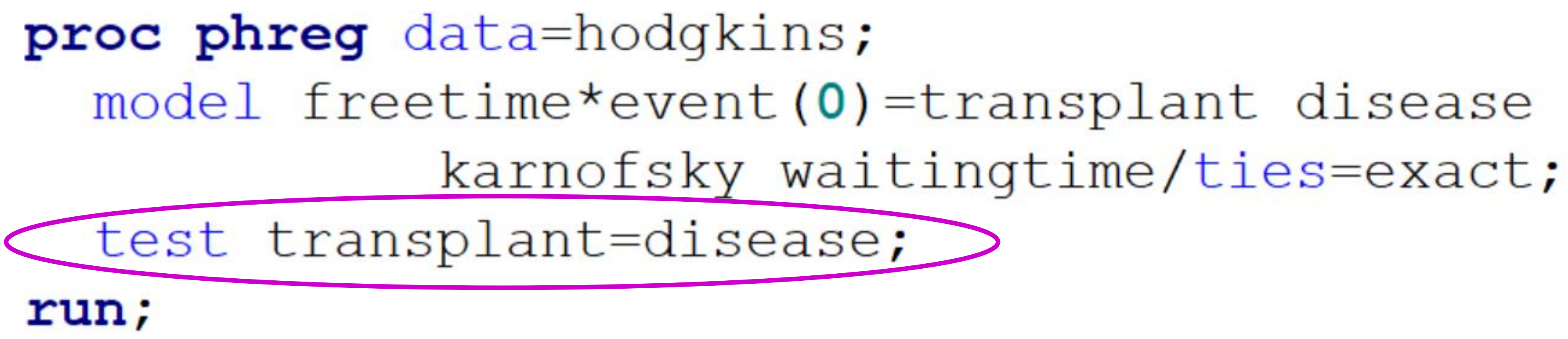
The Likelihood ratio test (and the scores test) can be calculated (SAS code in document *SAS examples.pdf*)

Example: Hodgkin’s disease

To test the local hypothesis that the effect of type of transplant is the same as the effect of disease type:

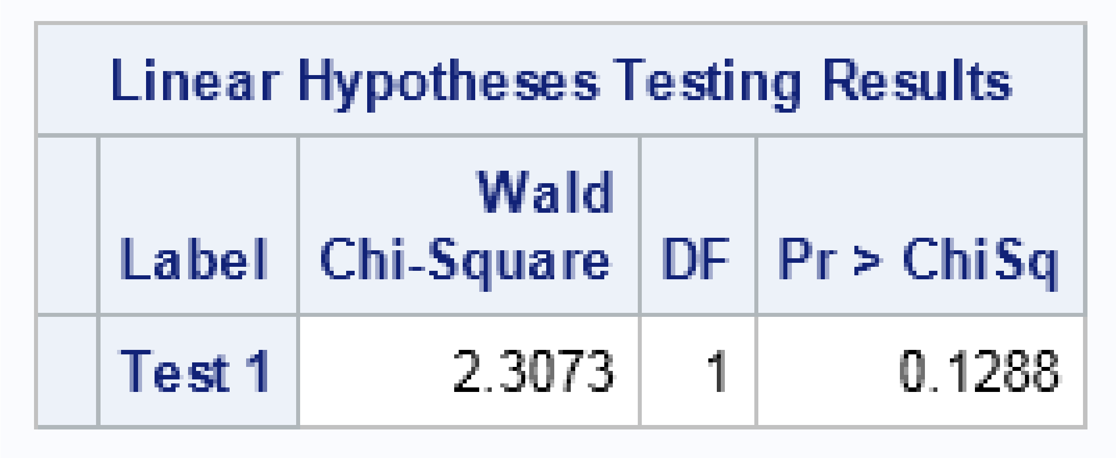
$$H_0 : \beta_{transplant} = \beta_{disease}$$

Use proc phreg and the **test** statement.



```
proc phreg data=hodgkins;
  model freetime*event(0)=transplant disease
              karnofsky waitingtime/ties=exact;
  test transplant=disease;
run;
```

Example: Hodgkin’s disease



Linear Hypotheses Testing Results

Label	Wald Chi-Square	DF	Pr > ChiSq
Test 1	2.3073	1	0.1288

Relative risks that don’t appear directly in the result table

You might be interested in relative risks not directly presented in the regression results.

Example:
three treatments are being compared; A, B, and C.

Treatment A is being used as the reference category, and dummy variables are created for treatments B and C.

Relative risks that don’t appear directly in the result table

You are interested in the risk of experiencing the event for patients receiving treatment B relative to the risk of experiencing the event for patients receiving treatment C. This will not appear directly with the regression results.

$$\frac{Risk_B}{Risk_C} = \frac{e^{\beta_B}}{e^{\beta_C}} = e^{(\beta_B - \beta_C)}$$

Relative risks that don’t appear directly in the result table

To find the confidence interval for this relative risk you need the standard error of b_B-b_C.

The variance-covariance matrix of the b_i’s can be estimated in SAS by using the **covb** option.

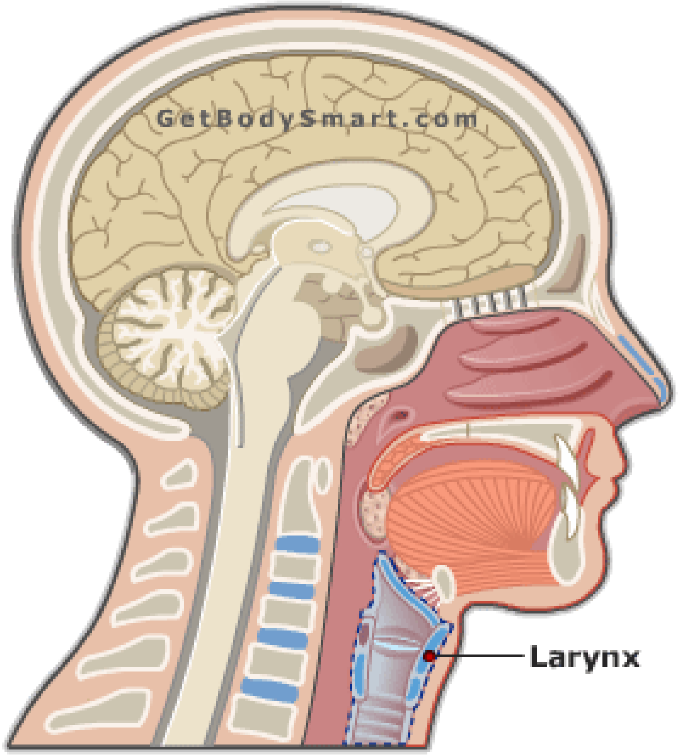
A confidence interval for (β_B- β_C) can be constructed by using V(b_B-b_C) = V(b_B) + V(b_C) – 2 Cov(b_B, b_C)

And a confidence interval for the relative risk $e^{(\beta_B-\beta_C)}$ is then obtained by exponentiating the lower and upper limits.

UPPSALA UNIVERSITET

Example: larynx cancer

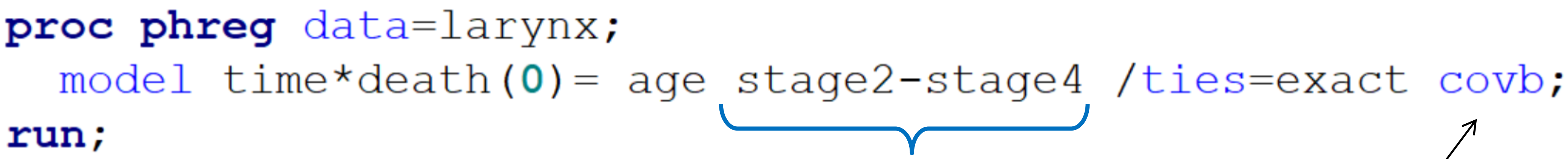
A study of 90 males with cancer of the larynx is described in section 1.8.



Example: larynx cancer

X = survival time from first treatment (years)

There are four stages of the disease, stage I – stage IV, ordered from least serious to most serious.

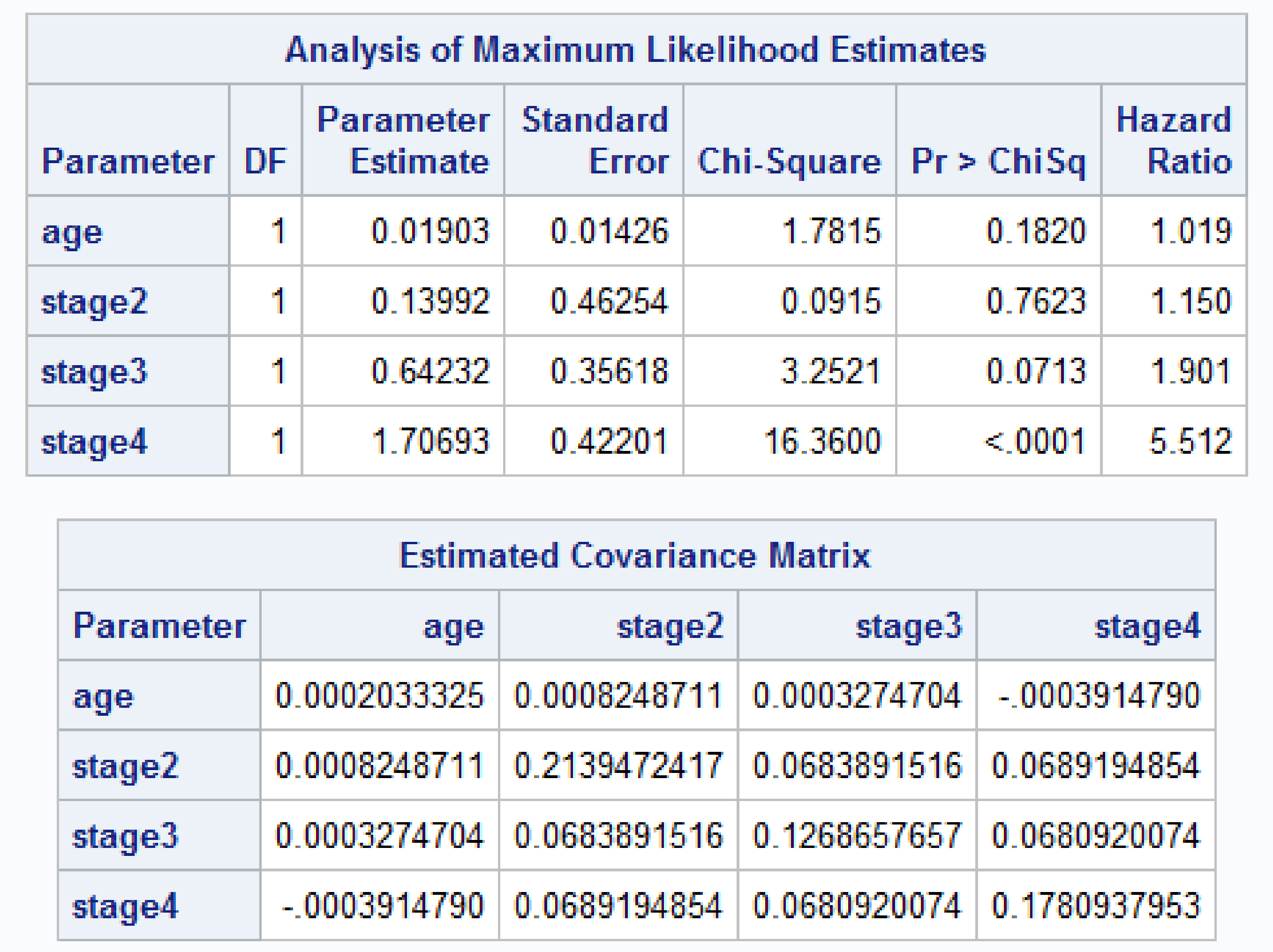


```
proc phreg data=larynx;
  model time*death(0)= age stage2-stage4 /ties=exact covb;
run;
```

Dummy variables for stages II, III and IV

Produces the variance-covariance matrix

Example: larynx cancer



Analysis of Maximum Likelihood Estimates						
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
age	1	0.01903	0.01426	1.7815	0.1820	1.019
stage2	1	0.13992	0.46254	0.0915	0.7623	1.150
stage3	1	0.64232	0.35618	3.2521	0.0713	1.901
stage4	1	1.70693	0.42201	16.3600	<.0001	5.512

Estimated Covariance Matrix				
Parameter	age	stage2	stage3	stage4
age	0.0002033325	0.0008248711	0.0003274704	-.0003914790
stage2	0.0008248711	0.2139472417	0.0683891516	0.0689194854
stage3	0.0003274704	0.0683891516	0.1268657657	0.0680920074
stage4	-.0003914790	0.0689194854	0.0680920074	0.1780937953

$$\frac{Risk_{IV}}{Risk_{III}} = \frac{5.512}{1.901}$$

$$= 2.9$$

Example: larynx cancer

95% CI for (β_{IV}- β_{III}) :

b_{IV}-b_{III} ± 1.96 sqrt(V(b_{IV}-b_{III}))

V(b_{IV}-b_{III}) = V(b_{IV}) + V(b_{III}) – 2 Cov(b_{IV}, b_{III}) =

= 0.17809 + 0.12687 – 2 × 0.06809 =

= 0.16878

1.70693 - 0.64232 ± 1.96 sqrt(0.16878)

1.06461 ± 0.80522

[0.2594; 1.8698]

Example: larynx cancer

95% CI for the relative risk $e^{(\beta_{IV}-\beta_{III})}$**:**

[exp(0.2594); exp(1.8698)]

[1.296; 6.487]

There is a significant difference in the risk of dying for stage IV compared to stage III (the relative risk is significantly different from 1)

UPPSALA
UNIVERSITET

Program L5

- **Regression for survival data**
  - Cox’s proportional hazards regression
    - Partial maximum likelihood
    - Interpretation of estimated coefficients
    - Continuous vs categorical covariates
    - Ties
    - Local tests
    - Model building
    - Time-dependent covariates

Model building

Different ways of building a model:

1) Hypothesis based modelling. The time distribution is predicted by explanatory variables selected to fit a specific hypothesis.

2) Predicting the distribution of time by selecting from a number of explanatory variables with no particular prior hypothesis in mind.

UPPSALA UNIVERSITET

Hypothesis based modelling

When a particular hypothesis is in mind, the explanatory variables fitting that hypothesis are included in the model.

Other explanatory variables can also be added, variables that can be seen as adjusters or confounders (variables that might affect the relationship between the hypothesis based variables and the outcome).

E.g. demographic variables (age, sex, etc.) can be confounders, or the severity of a patient’s illness, the size of a tumour, etc.

Hypothesis based modelling – forward selection approach

1) Fit the model including only the explanatory variables fitting the hypothesis.
2) Add one of the possible confounders to the model, to analyze the relationship between that variable and survival (adjusting for all hypothesis based variables).
3) Repeat 2) for each possible confounder, one at a time.
4) Include the confounder with the strongest significant relationship to survival to the model.
5) Repeat steps 2-4, with the "basic" model now including the confounder added in step 4). Stop when no more significant confounders are found.

UPPSALA
UNIVERSITET

P-value and information criteria approaches

Different ways of deciding which variable is most related to survival:

1) **p-value approach**: choose the variable with the lowest *p*-value.

2) **Information criterion approach**: choose the variable which yields the model with the lowest information criteria value. Information criteria are based on the likelihood function, their value increase when added variables are unnecessary.

Can be combined: add significant variables with not increasing information criteria values.

UPPSALA UNIVERSITET

Common information criteria

Akaike information criterion (AIC)

$$AIC = -2\log L + 2k$$

k = no. of parameters in model

Bayesian information criterion (BIC)

$$BIC = -2\log L + k\log n$$

Akaike information criterion corrected (AIC$_C$)

$$AIC_C = -2\log L + 2k + \frac{2k(k+1)}{n-k-1}$$

$$AIC_C = AIC + 2k(k+1)/(n\text{-}k\text{-}1)$$

UPPSALA
UNIVERSITET

Which to choose?

AIC/AIC$_C$ preferred over BIC, shown by Burnham and Anderson*.

They also recommend AIC$_C$ over AIC, especially for small *n* or large *k*.

*Burnham, K. P.; Anderson, D. R. (2002), Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach (2nd ed.), Springer-Verlag

UPPSALA
UNIVERSITET

Example: infection in burns

In section 1.6 a study is described to evaluate a protocol change in disinfectant practice in a large midwestern university medical center.

Control of infection is the primary concern for the 154 patients entered into the burn unit with varying degrees of burns.

The outcome variable is the time until infection from admission to the unit. Censoring variables are discharge from the hospital without an infection or death without an infection.

84 patients were in a group which had a body-cleansing method (disinfectant: chlorhexidine) and 70 patients received the routine bathing care method (disinfectant: povidone-iodine).

Example: infection in burns

Variables:

Trt = treatment (0=routine bathing, 1=body cleansing)

TimeStaph = Time to staphylococcus infection (days)

Staph = Staphylococcus indicator (1=infection, 0=no inf.)

Possible confounders*:*

Area (percentage burned, % of total surface area)

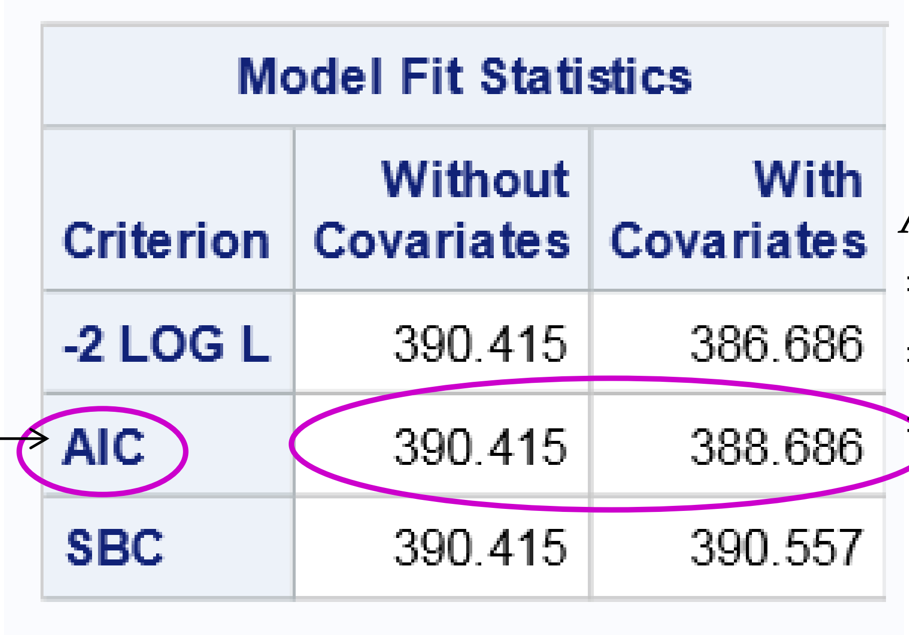
BurnSite (head, buttock, trunk, etc. – 6 indicators)

BurnType (1=chemical, 2=scald, 3=electric, 4=flame)

Example: infection in burns

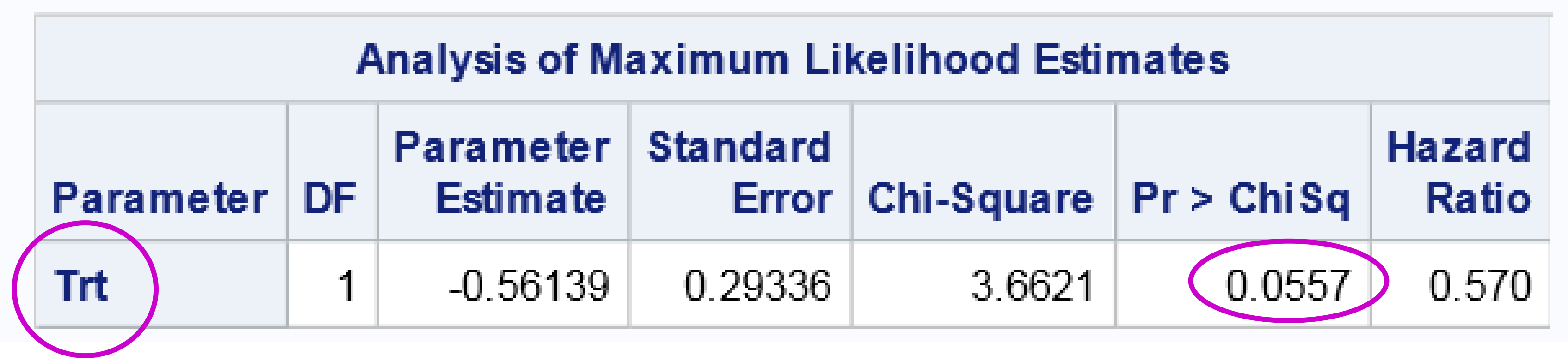
Akaike information criterion

Decreases as variables are added to the model. If it increases, the added variable is unnecessary.



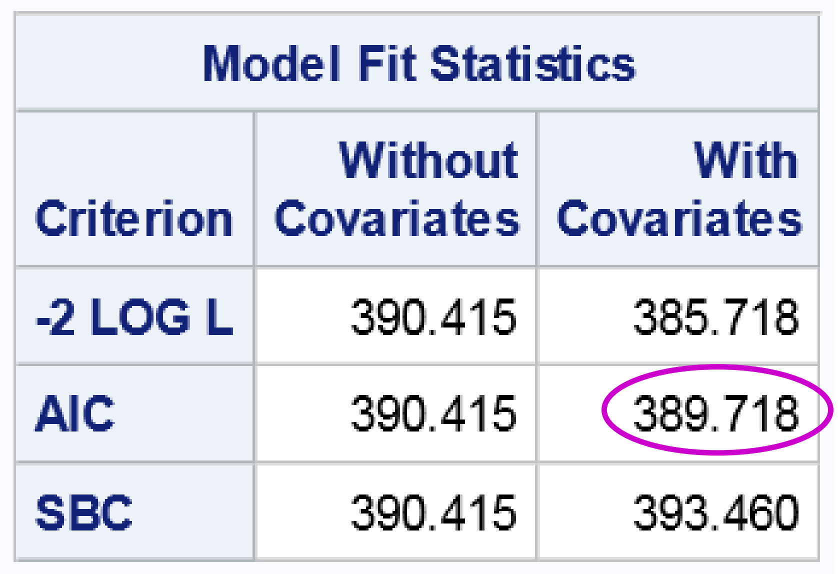
Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	390.415	386.686
AIC	390.415	388.686
SBC	390.415	390.557

$$AIC_C =$$
$$= AIC + 2k(k+1)/(n-k-1) =$$
$$= 388.686+2\cdot 1\cdot 2/(154-1-1)=$$
$$= 388.7123$$



Analysis of Maximum Likelihood Estimates						
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
Trt	1	-0.56139	0.29336	3.6621	0.0557	0.570

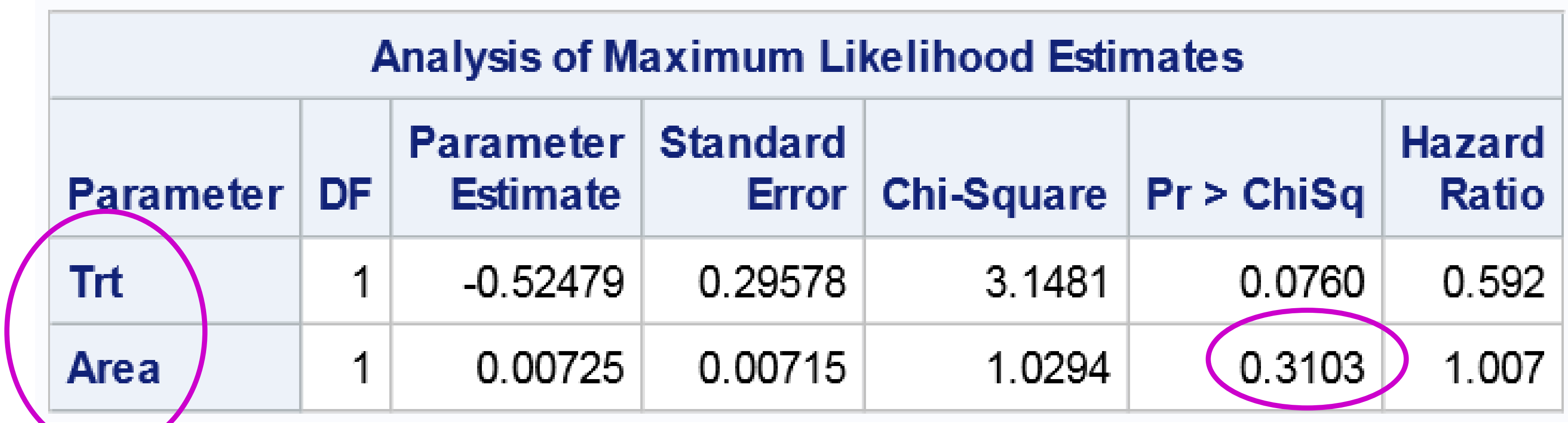
Example: infection in burns



Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	390.415	385.718
AIC	390.415	389.718
SBC	390.415	393.460

$$AIC_C = AIC + 2k(k+1)/(n\text{-}k\text{-}1) =$$
$$= 389.718+2\cdot 2\cdot 3/(154\text{-}2\text{-}1)=$$
$$= 389.7975$$

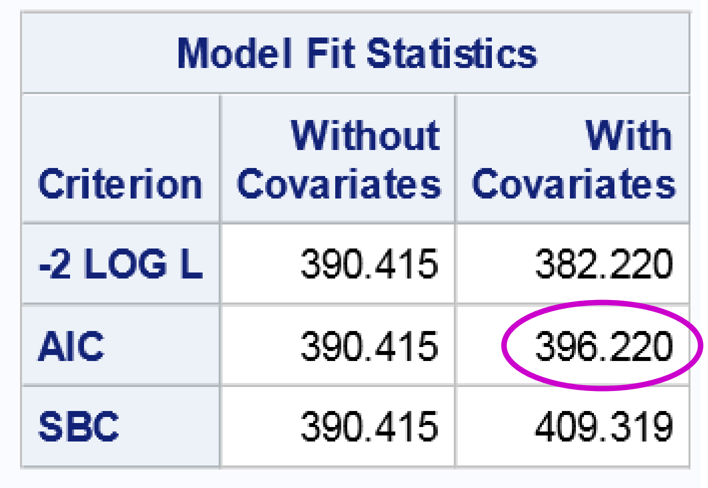
AIC_C increases compared to the model with trt only (388.7123)



Analysis of Maximum Likelihood Estimates						
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
Trt	1	-0.52479	0.29578	3.1481	0.0760	0.592
Area	1	0.00725	0.00715	1.0294	0.3103	1.007

Percentage of area burned not significant

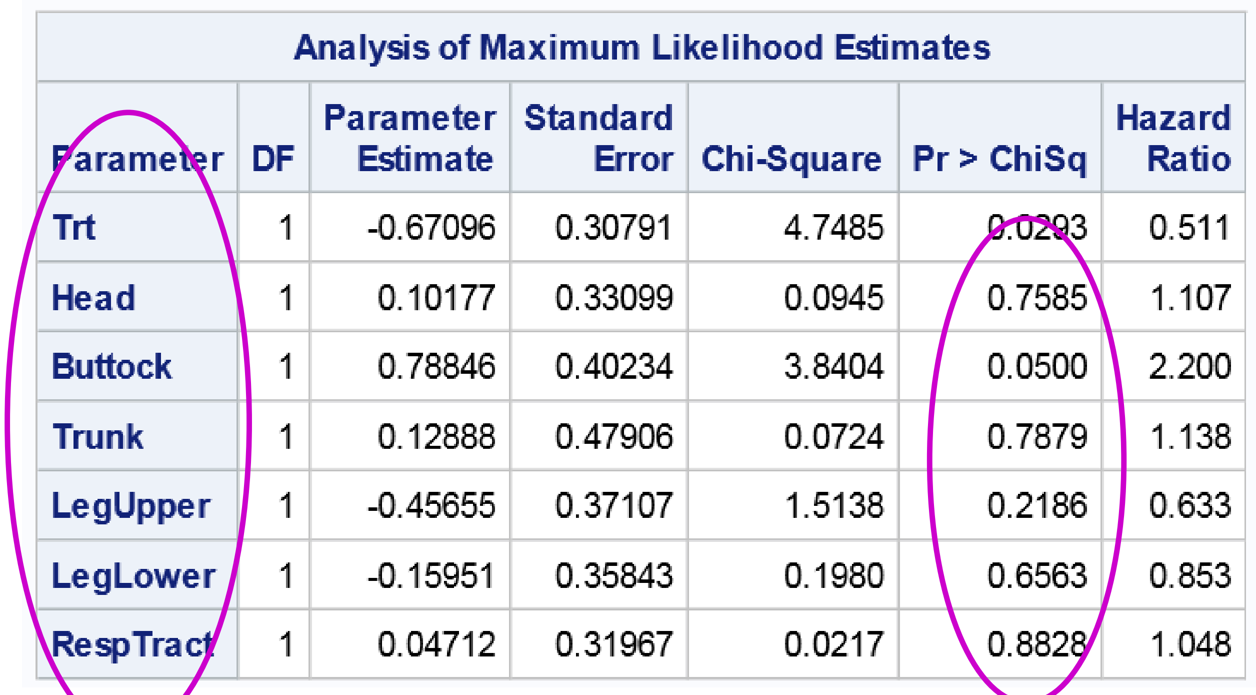
Example: infection in burns



Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	390.415	382.220
AIC	390.415	396.220
SBC	390.415	409.319

$$AIC_C = AIC + 2k(k+1)/(n\text{-}k\text{-}1) =$$
$$= 396.220+2\cdot 7\cdot 8/(154\text{-}7\text{-}1)=$$
$$= 396.9871$$

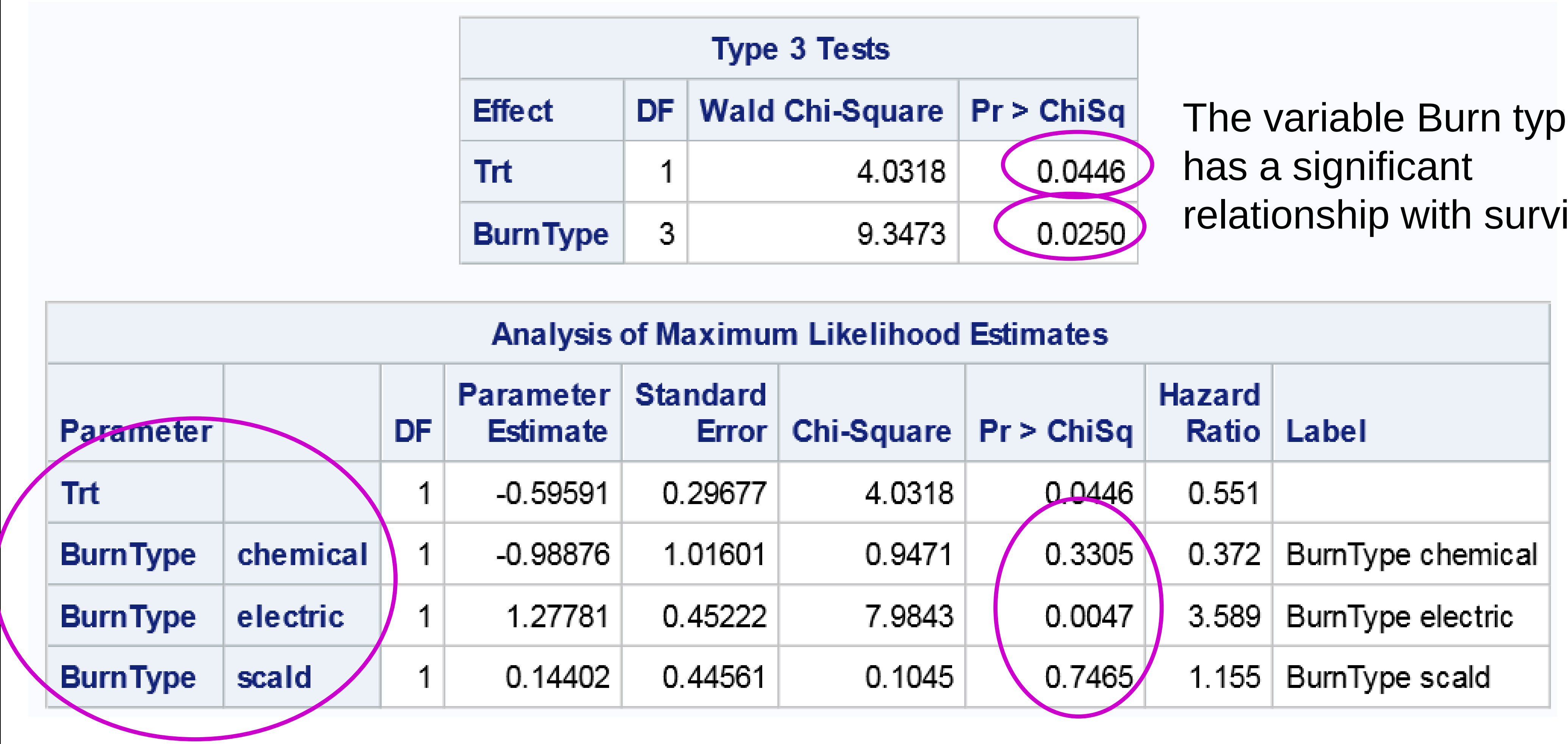
AIC_C increases compared to the model with trt only (388.7123)



Analysis of Maximum Likelihood Estimates						
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
Trt	1	-0.67096	0.30791	4.7485	0.0293	0.511
Head	1	0.10177	0.33099	0.0945	0.7585	1.107
Buttock	1	0.78846	0.40234	3.8404	0.0500	2.200
Trunk	1	0.12888	0.47906	0.0724	0.7879	1.138
LegUpper	1	-0.45655	0.37107	1.5138	0.2186	0.633
LegLower	1	-0.15951	0.35843	0.1980	0.6563	0.853
RespTract	1	0.04712	0.31967	0.0217	0.8828	1.048

Buttock is the only burn site which has a significant relationship with survival

Example: infection in burns



Type 3 Tests			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
Trt	1	4.0318	0.0446
BurnType	3	9.3473	0.0250

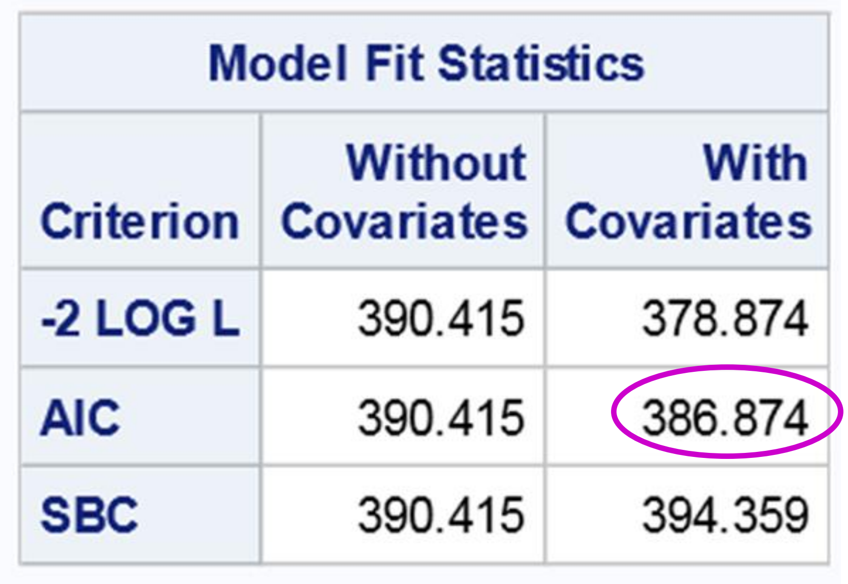
The variable Burn type has a significant relationship with survival

Analysis of Maximum Likelihood Estimates

Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
Trt		1	-0.59591	0.29677	4.0318	0.0446	0.551	
BurnType	**chemical**	1	-0.98876	1.01601	0.9471	0.3305	0.372	BurnType chemical
BurnType	**electric**	1	1.27781	0.45222	7.9843	0.0047	3.589	BurnType electric
BurnType	**scald**	1	0.14402	0.44561	0.1045	0.7465	1.155	BurnType scald

Electric burn is the only burn type which has a significant relationship with survival

Example: infection in burns



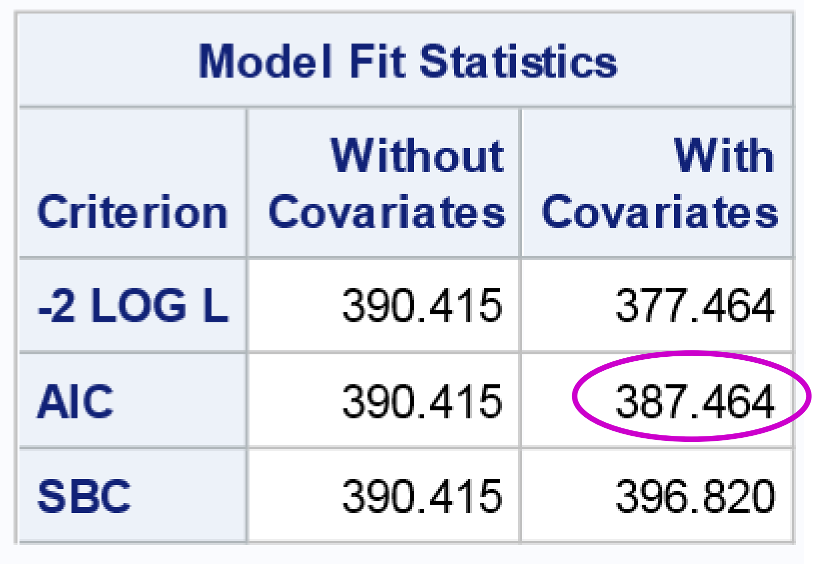
Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	390.415	378.874
AIC	390.415	386.874
SBC	390.415	394.359

$$AIC_C = AIC + 2k(k+1)/(n-k-1) =$$
$$= 386.874+2\cdot 4\cdot 5/(154-4-1)=$$
$$= 387.1425$$

AIC$_C$ decreases compared to the model with trt only (388.7123)

This is a better model than using treatment as a single covariate

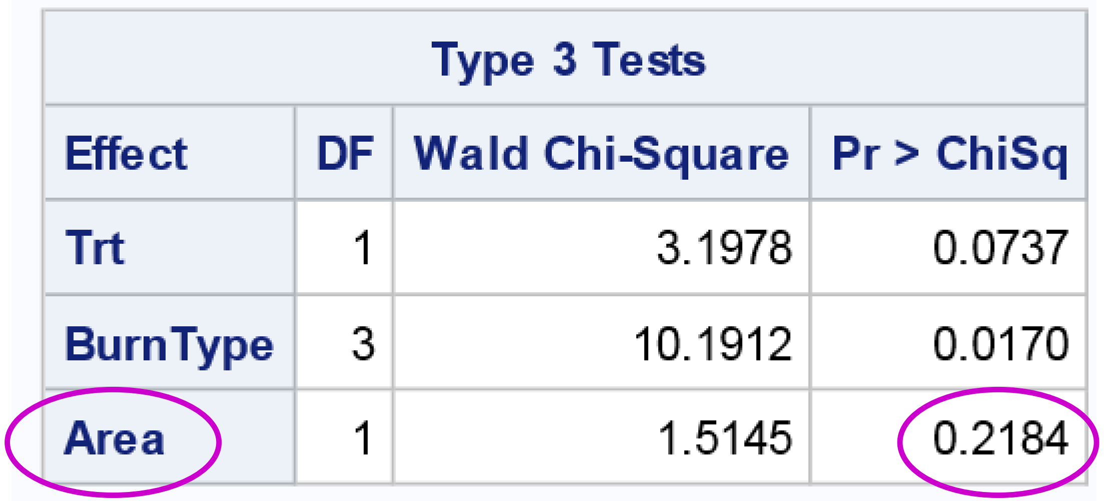
Example: infection in burns



Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	390.415	377.464
AIC	390.415	387.464
SBC	390.415	396.820

$$AIC_C = AIC + 2k(k+1)/(n-k-1) =$$
$$= 387.464+2 \cdot 5 \cdot 6/(154-5-1)=$$
$$= 387.8694$$

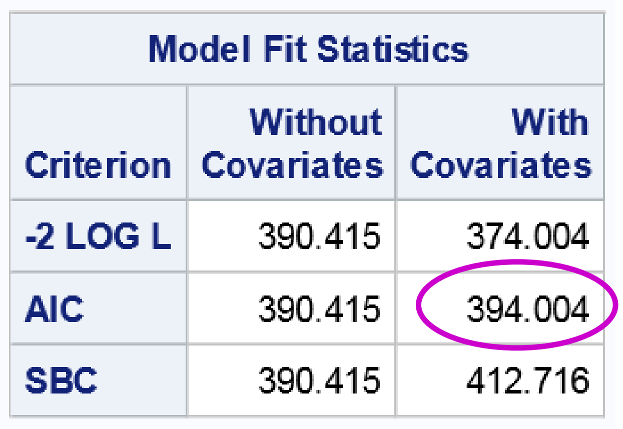
AIC_C increases compared to the model with trt and burn type (387.1425)



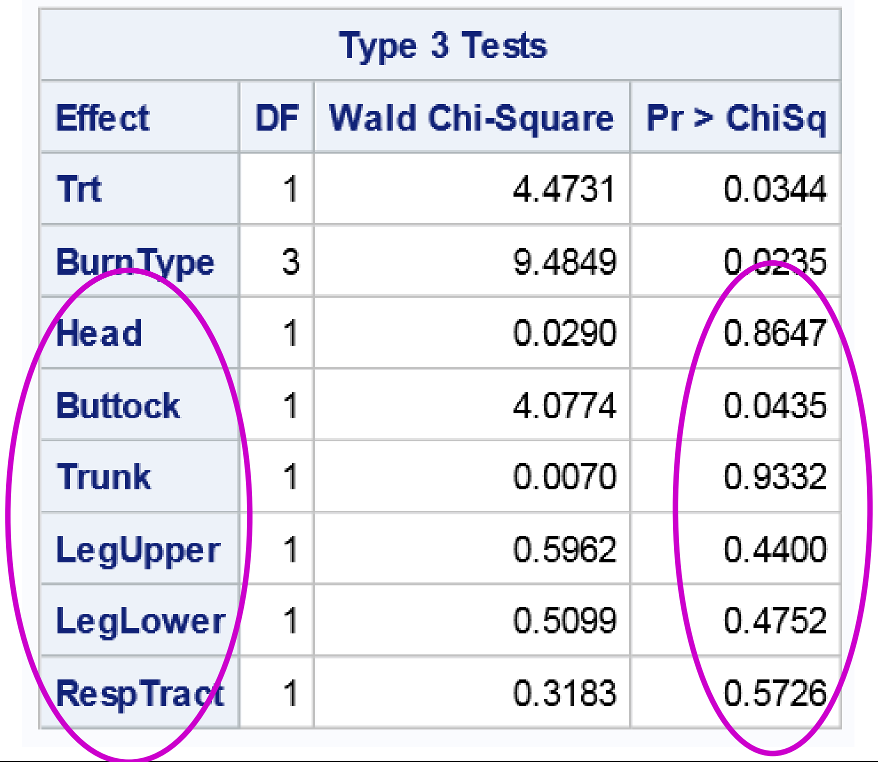
Type 3 Tests			
Effect	DF	Wald Chi-Square	Pr > ChiSq
Trt	1	3.1978	0.0737
BurnType	3	10.1912	0.0170
Area	1	1.5145	0.2184

Percentage of area burned does not have a significant relationship with survival

Example: infection in burns



Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	390.415	374.004
AIC	390.415	394.004
SBC	390.415	412.716



Type 3 Tests			
Effect	DF	Wald Chi-Square	Pr > ChiSq
Trt	1	4.4731	0.0344
BurnType	3	9.4849	0.0235
Head	1	0.0290	0.8647
Buttock	1	4.0774	0.0435
Trunk	1	0.0070	0.9332
LegUpper	1	0.5962	0.4400
LegLower	1	0.5099	0.4752
RespTract	1	0.3183	0.5726

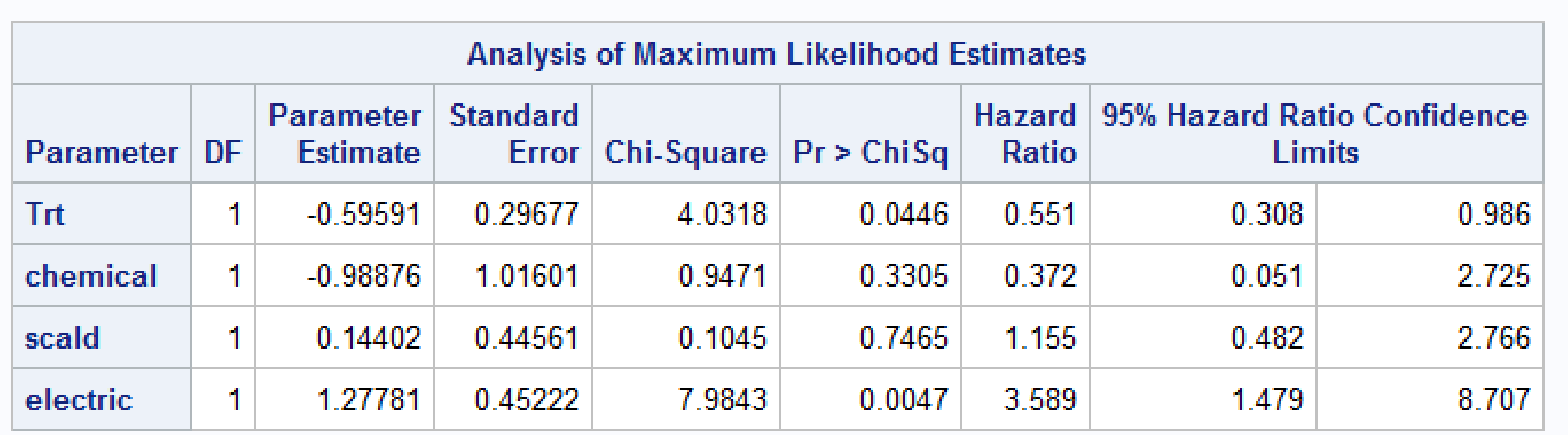
$$AIC_C = AIC + 2k(k+1)/(n-k-1) =$$
$$= 394.004+2\cdot 10\cdot 11/(154-10-1)=$$
$$= 395.5425$$

AIC$_C$ increases compared to the model with trt and burn type (387.1425)

Buttock is again the only burn site which has a significant relationship with survival

Example: infection in burns

The “best” model



Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	
Trt	1	-0.59591	0.29677	4.0318	0.0446	0.551	0.308	0.986
chemical	1	-0.98876	1.01601	0.9471	0.3305	0.372	0.051	2.725
scald	1	0.14402	0.44561	0.1045	0.7465	1.155	0.482	2.766
electric	1	1.27781	0.45222	7.9843	0.0047	3.589	1.479	8.707

Modelling without prior hypothesis – forward selection approach

1) Fit the model including one of the possible explanatory variables.
2) Repeat 1) for each explanatory variable, one at a time.
3) Choose the variable with the strongest significant relationship to survival.
4) Repeat steps 1-3, with the “basic” model now including the variable chosen in step 3). Stop when no more significant explanatory variables are found.

Backward selection approach

1) Fit the model including all possible explanatory variables (covariates).
2) Remove the least significant covariate from the model.
3) Repeat 2) for each covariate, until no insignificant covariates are left.

Stepwise selection approach

The stepwise selection approach combines forward selection and backward selection, adding and deleting variables in an iterative manner.

Forward, backward, and stepwise selection are all available in SAS, based on the *p*-value approach.

Best subset selection approach

A specified number of best models are found containing one, two, or three variables, and so on, up to the single model containing all of the explanatory variables.

The criterion used to determine the "best" subset is based on the global score chi-square statistic (the higher the value, the “better” the model – for that number of explanatory variables).

All possible model selection approach

The “all possible model” selection methodology is based on a combination of stepwise regression, Akaike information criteria, and the best subset selection.

All possible models, from the null model to the full model including all the explanatory variables are determined.

The models will be ordered by minimizing the AIC value at every step.

http://www2.sas.com/proceedings/forum2008/375-2008.pdf

Which approach to choose?

In different literature there are different suggestions on how to build statistical models.

Some consensuses:

1) Avoid blindfolded use of automatic selection procedures
2) Scientific knowledge (e.g. medical) plays an important role.

UPPSALA UNIVERSITET

Program L5

- **Regression for survival data**
  - Cox's proportional hazards regression
    - Partial maximum likelihood
    - Interpretation of estimated coefficients
    - Continuous vs categorical covariates
    - Ties
    - Local tests
    - Model building
    - Time-dependent covariates

Time-dependent covariates

Explanatory variables (covariates) that don’t have fixed values (recorded at the start of the study) are said to be **time-dependent**.

The values of time-dependent covariates may change during the course of the study.

If a covariate Z is time-dependent we denote it $Z(t)$.

UPPSALA UNIVERSITET

Example: Ex-cons

X = time to crime relapse for ex-convicts (months)

The risk of committing another crime decreases with time from prison release.

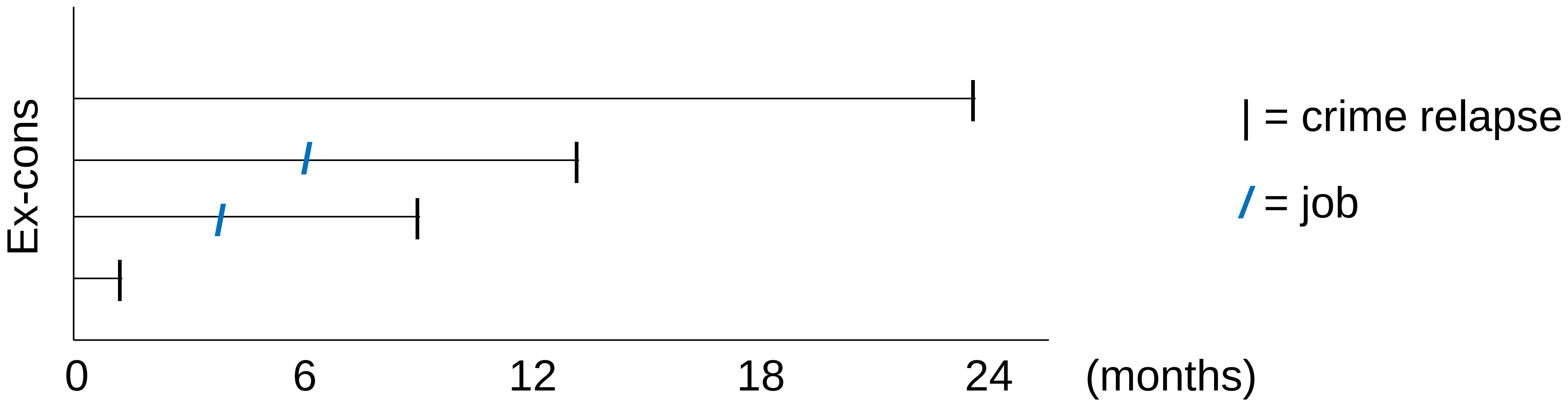
The risk of crime relapse also decreases if the ex-con gets a job (not known at the time of prison release).

Example: Ex-cons

t_i	d_i	Job	$Z(t)$			
(months)		*(months)*	t = 1	8	13	24
1	1	-	0	-	-	-
8	1	4	0	1	-	-
13	1	6	0	1	1	-
24	1	-	0	0	0	0

$Z(t)$ changes with time

Example: Ex-cons



The “partial likelihood” with time-dependent covariates

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{D} \underset{\uparrow}{L_i} = \prod_{i=1}^{D} \frac{\exp\left(\sum_{h=1}^{p} \beta_h Z_{(i)h}(t_i)\right)}{\sum_{j \in R(t_i)} \exp\left(\sum_{h=1}^{p} \beta_h Z_{jh}(t_i)\right)}$$

Partial likelihood for
the *i*th event time

Risk set
at time t_i

Individual hazards with time-dependent covariates

The hazard for an individual experiencing the event at time t_1:

$$h(t_1 \mid Z(t_i) = h_0(t)e^{\beta Z(\underset{\uparrow}{t_i})}$$

Event
time t_i

$$= h_0(t)e^{\beta}$$ if $Z(t)$=1 at this time for this individual

$$= h_0(t)$$ if $Z(t)$=0 at this time for this individual

Example: Ex-cons

t_i	d_i	Job	$Z(t)$			
$(months)$		$(months)$	t = 1	8	13	24
1	1	-	0	-	-	-
8	1	4	0	1	-	-
13	1	6	0	1	1	-
24	1	-	0	0	0	0

$$L_1 = \frac{e^{\beta \cdot 0}}{e^{\beta \cdot 0} + e^{\beta \cdot 0} + e^{\beta \cdot 0} + e^{\beta \cdot 0}} = \frac{1}{4}$$

= probability (likelihood) that no. 1 experiences the event, compared to all the individuals at risk

Example: Ex-cons

t_i	d_i	Job	$Z(t)$			
$(months)$		$(months)$	$t=1$	8	13	24
1	1	-	0	-	-	-
8	1	4	0	1	-	-
13	1	6	0	1	1	-
24	1	-	0	0	0	0

$$L_2 = \frac{e^{\beta \cdot 1}}{e^{\beta \cdot 1} + e^{\beta \cdot 1} + e^{\beta \cdot 0}}$$

$$L_3 = \frac{e^{\beta \cdot 1}}{e^{\beta \cdot 1} + e^{\beta \cdot 0}}$$

$$L_4 = \frac{e^{\beta \cdot 0}}{e^{\beta \cdot 0}} = 1$$

Time-dependent covariates

Read more (Studium module Articles):

Therneau et al. (2018), Using Time Dependent Covariates and Time Dependent Coefficients in the Cox model.