

# Introduction to Survival Analysis

Thomas Hielscher  
Division of Biostatistics

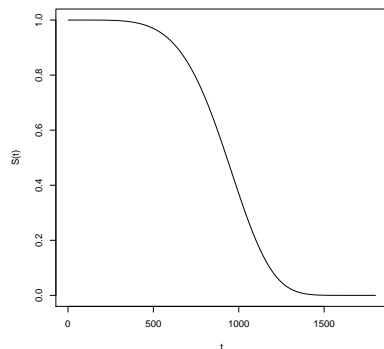
Hazard Function and Cox Regression

- 1 Hazard function
- 2 Cox PH Regression
- 3 Model diagnostics
- 4 Stratification
- 5 Interaction
- 6 SigmaPlot
- 7 Competing Risk

# Survival function

Recall:

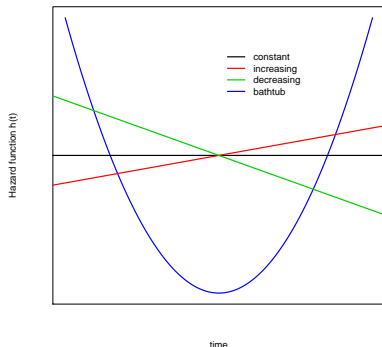
- $T$  denotes the survival time,  $T \geq 0$ , right-skewed
- $F(t)$ : distribution function of the survival times
- Survival function  $S(t) = \Pr(T > t) = 1 - F(t)$ ,  $t \geq 0$



# Hazard function

Hazard function  $h(t)$  is the instantaneous rate at which events/failures occur at  $t$ , given no event occurred until  $t$ .

If  $t$  continuous:  $h(t) = f(t)/S(t)$ ,  $f = F'$

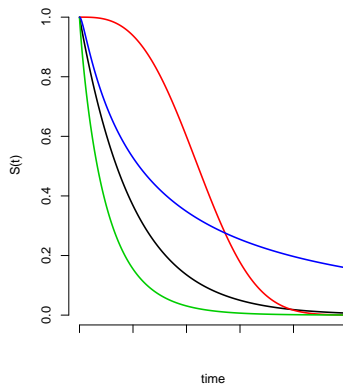
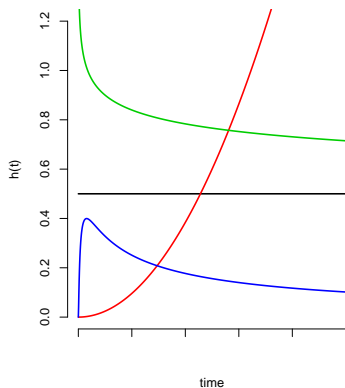


## Examples:

- increasing: natural aging
- decreasing: early onset of failure (e.g. transplant)
- bathtub: life span (early risk after birth)

Alias: conditional failure rate, force of mortality, hazard rate

# Hazard vs. Survival function



Based on exponential (black), weibull (red/green) and lognormal (blue) distribution

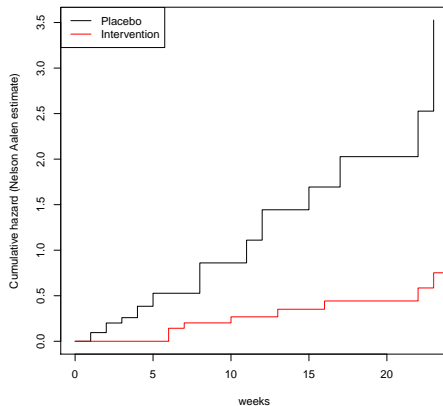
# Hazard function

- Is not a probability ( $h(t) \geq 0$ , but can be larger than 1)
- Can be interpreted as expected number of events per subject per time unit
- Example:  $h(t) = 0.5$ , time unit is months, on average 0.5 events will occur at  $t$  per subject at risk and per month, assuming constant hazard during that month
- If survival times can be described by exponential distribution, hazard is constant:  

$$h(t) = \frac{f(t)}{S(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$$
- cumulative hazard  $H(t) = -\log[S(t)]$

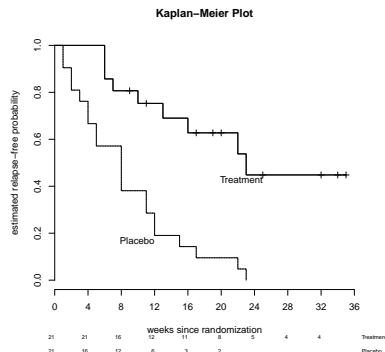
# Non-parametric estimation of cumulative hazard function

Nelson Aalen estimator  $\widehat{H}(t) = \sum_{t_i \leq t} \frac{d_i}{n_i}$



# Example data

Data from Freireich et al. 'The effect of 6-mercaptopurine on the duration of steroid-induced remissions in acute leukemia' Blood 21, 699-716, 1963.



- logrank test p-value < 0.001
- Quantify difference in risk of event?
- Adjust for confounders?
- Assess impact of continuous covariates?



# Cox Proportional Hazard Regression (1)

**Objective:** examine the relationship between survival and one or more explanatory variables (covariates)

$X_1, \dots, X_k$  are  $k$  covariates, measured for all  $n$  subjects

- Modeling survival linear-like using (log) hazard function  $h(t)$ :  

$$\log h(t, X) = \alpha(t) + \beta_1 X_1 + \dots + \beta_k X_k$$
- Equivalent to  

$$h(t, X) = h_0(t) \exp\{\beta_1 X_1 + \dots + \beta_k X_k\}, \alpha(t) = \log h_0(t)$$
- $h_0(t)$  is baseline hazard, assumed to be the same for all subjects
- Hazard depends on time and covariates, but effect of covariates time-invariant

## Cox Proportional Hazard Regression (2)

- Proportional: hazard rates of two individuals with different covariate values are proportional → relative effects
- Use relative hazards or hazard ratio (HR) to compare subjects with covariate values  $x_i = (x_{1i}, \dots, x_{ki})$ :

$$\frac{h(t|X = x_1)}{h(t|X = x_2)} = \frac{h_0(t) \exp\{\beta_1 x_{11} + \dots\}}{h_0(t) \exp\{\beta_1 x_{21} + \dots\}} \quad (1)$$

$$= \frac{\exp\{\beta_1 x_{11} + \dots\}}{\exp\{\beta_1 x_{21} + \dots\}} \quad (2)$$

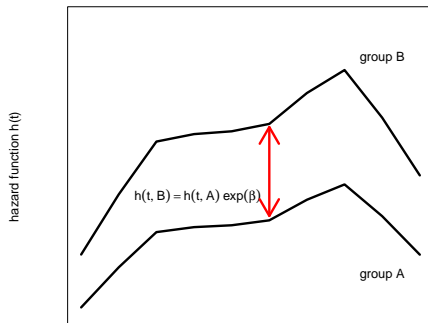
$$= \exp(\beta_1 (x_{11} - x_{21}) + \dots) \quad (3)$$

- $h(t|X = x_1) = h(t|X = x_2) \times \exp(\beta_1 (x_{11} - x_{21}) + \dots)$
- $h_0(t)$  cancels out → semiparametric model
- Ratio does not depend on time, i.e. constant over time

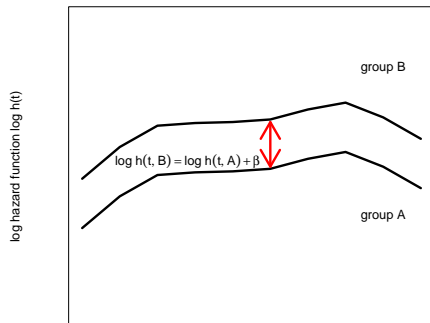
# Illustration Proportional Hazards

One binary covariate  $X$  ( $k=1$ ), group 1 vs group 0:

$$\text{relative hazard} = \frac{h(t|group=1)}{h(t|group=0)} = \exp(\beta(1 - 0)) = \exp(\beta), \beta > 0$$



time



time

# Interpretation of Hazard Ratio

What does a hazard ratio of 2 in terms of overall survival mean?

- interpretation of relative risks/ risk ratio
- In general: Increasing  $X_j$  by one unit multiplies the estimated hazard by  $\exp(\beta_j)$
- Binary covariate (group B vs group A): risk of death is 2 times that in group B compared to group A, failure rate is 100% higher at any time point
- Continuous covariate (age per 10 years): risk of death increases by factor 2 for a 50-year old compared to a 40-year old (60 vs 50, etc...)
- If  $T$  is exponentially distributed, hazard ratio (B vs A) corresponds to ratio of median survival times (A vs B), i.e. median survival in A is 100% longer

## Example data

Covariates  $X$ : treatment group placebo (B) vs treatment (A),  
log-transformed white blood count, sex

		$\hat{\beta}$	$HR = \exp(\hat{\beta})$	$se(\hat{\beta})$	$z$	p-value
group	B vs A	1.50	4.50	0.46	3.26	0.001
logWBC		1.68	5.38	0.34	5.00	< 0.0001
gender	m vs f	0.31	1.37	0.45	0.69	0.49

- $H_0 : \beta_1 = 0$  (no effect of treatment)
- $z = \hat{\beta}/se(\hat{\beta})$  (Wald test statistic)
- $z$  asymptotically standard-normal distributed
- Hazard ratio:  $\exp(\hat{\beta})=4.5$ , 4.5-fold increased risk in placebo group
- 95% confidence interval:  $\exp(\hat{\beta} + / - 1.96 \times se)$   
→ [1.8 ; 11.1]

# Assumptions to check

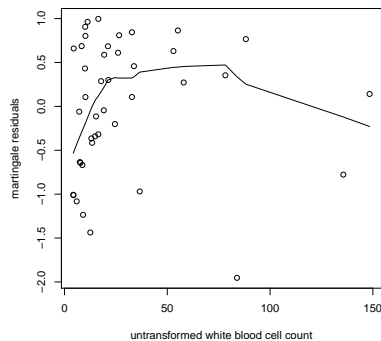
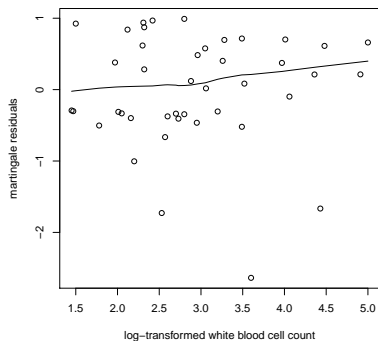
- Proportional hazards
- Linear relationship between (continuous) covariates and log hazard  $h(t)$
- Influential observations
- Collinearity of covariates

Note: Nonlinear functional form and interaction of covariates can affect PH assumption

# Linearity

Linearity with log hazard.

Martingale residuals for subject  $i$  at time  $t_i$ :  $M_i = \delta_i - H(t_i, x_i, \beta_i)$



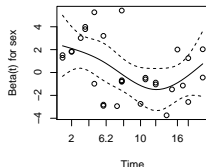
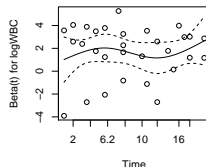
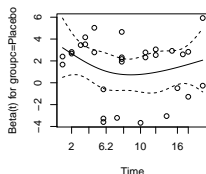
# Proportional hazards assumption (1)

- Assumption is violated, if covariate effect not constant over time
- Graphically:  $\log(-\log(S(t)))$  vs  $\log(\text{time}) \rightarrow$  parallel lines
- Scaled Schoenfeld residuals based on Cox model (Therneau and Grambsch, 2000):  
separate residual for each individual for each covariate
- Plot residuals against time in order to detect non-constant effects
- Test PH assumption based on correlation between scaled Schoenfeld residuals and time



# Proportional hazards assumption (2)

## Example data



	rho	chisq	p
group	-0.10	0.34	0.56
logWBC	0.06	0.16	0.69
sex	-0.37	4.08	0.04
GLOBAL		4.23	0.24

- no correlation, thus no time-varying effect for group
- violation for sex

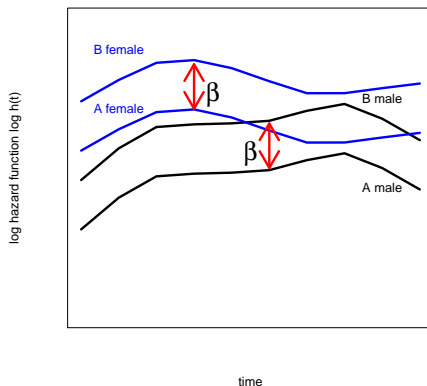
## Proportional hazards assumption (3)

What if PH assumption fails for covariate  $Z$ ?

- Test/p-value for covariate  $Z$  is not needed, but adjustment
  - Stratification (categorical, grouped continuous)
- Test/p-value for covariate  $Z$  is needed
  - If p-value is very small: 'principle of conservatism'
  - Include additional time-varying/time-dependent effect  $Z \times \log(\text{time})$ :  

$$h(t, X, Z) = h_0(t) \exp\{\beta_1 Z + \beta_2 Z \times \log(t) + \dots + \beta_k X_k\}$$
  - Local test  $H_0 : \beta_1 = 0$  and  $\beta_2 = 0$

# Stratification



- Divide data into strata based on values of covariate  $Z$  (e.g. male vs female)
- Each stratum can have a different baseline hazard  

$$h(t, X, Z = j) = h_{0j}(t) \exp\{\beta_1 X_1 + \dots + \beta_k X_k\},$$

$$j = 1, \dots, k$$
- Coefficients are assumed to be equal and constant across strata
- 'Weighted' hazard ratio across strata
- Issue: number of strata

## Example data with stratification

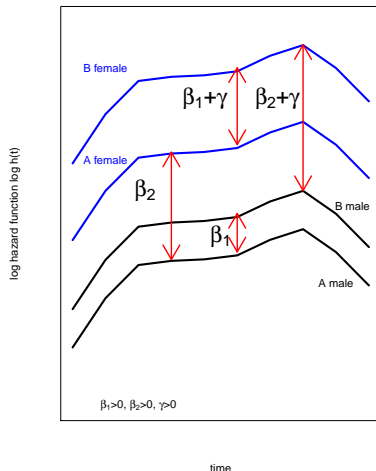
Covariates  $X$ : treatment group placebo (B) vs treatment (A),  
log-transformed white blood count

Strata: gender

		$\hat{\beta}$	$\exp(\hat{\beta})$	$\text{se}(\hat{\beta})$	z	p-value
group	B vs A	1.00	2.71	0.47	2.11	0.0351
logWBC		1.45	4.28	0.34	4.22	< 0.0001

- Smaller hazard ratio for treatment:  $\exp(\hat{\beta})=2.7$  after adjustment
- No estimate for gender

# Interaction



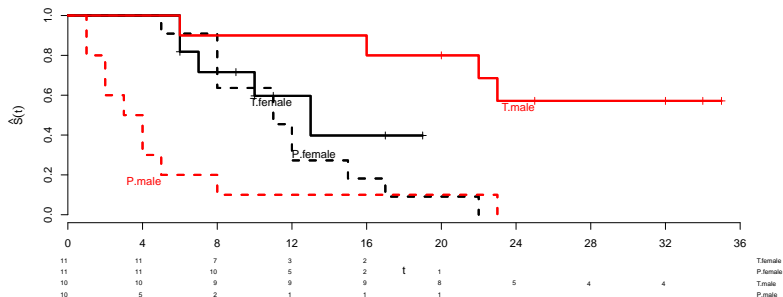
- Detection of subgroup effects, predictive factors
- Interaction between two covariates  $X_1$  and  $X_2$  occurs, if effect of  $X_1$  on survival outcome depends on  $X_2$
- $$h(t, X) = h_0(t) \exp\{\beta_1 X_1 + \beta_2 X_2 + \gamma Z\},$$
  

$$Z = X_1 \times X_2$$
- Quantitative interaction: effect of  $X_1$  has same direction, different magnitude for different values of  $X_2$
- Qualitative interaction: effect of  $X_1$  has opposite directions for different values of  $X_2$

# Example data with interaction

		$\hat{\beta}$	$\exp(\hat{\beta})$	$se(\hat{\beta})$	z	p-value
group	B vs A	0.62	1.87	0.54	1.15	0.25
sex	m vs f	-1.11	0.33	0.71	-1.57	0.12
group * sex		1.96	7.08	0.81	2.40	0.02

Kaplan-Meier Plot – treatment vs sex interaction



# Analysis with SigmaPlot

- Cox PH model including local and global tests (Breslow ties method)
- Cox PH model including stratification
- Check PH assumption not possible
- Interaction term inclusion not directly possible, need to be defined in dataset
- More flexible alternatives: R, SAS, SPSS
- Not doable with GraphPad Prism

# How to analyze with SigmaPlot?

SigmaPlot - Notebook2

Home W Create Graph G Graph Page A Analysis R Report T TopBox

Graph Worksheet Report Notebook Insert into Create Graph Output Notebook Manager Property Browser Gallery pane Report Page Dat... Full Screen Title Status Bar Tabs Cascade Close All Window List

Notebook Manager Data 1

Notebook

- All Open Notebooks
- Notebook2.JNB
  - Section 1
  - Data 1

Summary Info

Created 09.11.2011 14:4...

Modified 03.12.2012 08:4...

Author hielsche

Description Worksheet

	1-time	2-status	3-sex	4-logWBC	5-group	6	7	8	9	10	11	12
13	11.0000	0.0000	0.0000	2.6000	Treatment							
14	10.0000	0.0000	0.0000	2.7000	Treatment							
15	10.0000	1.0000	0.0000	2.9600	Treatment							
16	9.0000	0.0000	0.0000	2.8000	Treatment							
17	7.0000	1.0000	0.0000	4.4300	Treatment							
18	6.0000	0.0000	0.0000	3.2000	Treatment							
19	6.0000	1.0000	0.0000	2.3100	Treatment							
20	6.0000	1.0000	1.0000	4.0600	Treatment							
21	6.0000	1.0000	0.0000	3.2800	Treatment							
22	23.0000	1.0000	1.0000	1.9700	Placebo							
23	22.0000	1.0000	0.0000	2.7300	Placebo							
24	17.0000	1.0000	0.0000	2.9500	Placebo							
25	15.0000	1.0000	0.0000	2.3000	Placebo							
26	12.0000	1.0000	0.0000	1.5000	Placebo							
27	12.0000	1.0000	0.0000	3.0600	Placebo							
28	11.0000	1.0000	0.0000	3.4900	Placebo							
29	11.0000	1.0000	0.0000	2.1200	Placebo							
30	8.0000	1.0000	0.0000	3.5200	Placebo							
31	8.0000	1.0000	0.0000	3.0500	Placebo							
32	8.0000	1.0000	0.0000	2.3200	Placebo							
33	8.0000	1.0000	1.0000	3.2600	Placebo							
34	5.0000	1.0000	1.0000	3.4900	Placebo							
35	5.0000	1.0000	0.0000	3.9700	Placebo							
36	4.0000	1.0000	1.0000	4.3600	Placebo							
37	4.0000	1.0000	1.0000	2.4200	Placebo							
38	3.0000	1.0000	1.0000	4.0100	Placebo							
39	2.0000	1.0000	1.0000	4.9100	Placebo							
40	2.0000	1.0000	1.0000	4.4800	Placebo							
41	1.0000	1.0000	1.0000	2.8000	Placebo							
42	1.0000	1.0000	1.0000	5.0000	Placebo							
43												
44												
45												
46												



# How to analyze with SigmaPlot?

SigmaPlot - Notebook2

Home | Worksheet | Create Graph | Graph Page | Analysis | Report | TopBox

Advisor | Tests | Options | Statistics | Transform | Nonlinear Regression | Graph Analysis

Cox PH Model | Sample size | Power | Quick Transform... | Regression Wizard... | Dynamic Fit Wizard... | Global Curve Fit Wizard... | Histogram... | Normalized Ternary Data... | Smoothers... | Linear Regression... | Plot Equation...

Notebook Manager | Data 1

Notebook

- All Open Notebooks
- Notebook2.JNB
  - Section 1
    - Data 1

Summary Info

Created	09.11.2011 14:4...
Modified	03.12.2012 08:4...
Author	hielsche
Description	Worksheet

	1-time	2-status	3-sex	4-logWBC	5-group	6	7	8	9	10	11	12
13	11.0000	0.0000	0.0000	2.6000	Treatment							
14	10.0000	0.0000	0.0000	2.7000	Treatment							
15	10.0000	1.0000	0.0000	2.9600	Treatment							
16	9.0000	0.0000	0.0000	2.8000	Treatment							
17	7.0000	1.0000	0.0000	4.4300	Treatment							
18	6.0000	0.0000	0.0000	3.2000	Treatment							
19	6.0000	1.0000	0.0000	2.3100	Treatment							
20	6.0000	1.0000	1.0000	4.0600	Treatment							
21	6.0000	1.0000	0.0000	3.2800	Treatment							
22	23.0000	1.0000	1.0000	1.9700	Placebo							
23	22.0000	1.0000	0.0000	2.7300	Placebo							
24	17.0000	1.0000	0.0000	2.9500	Placebo							
25	15.0000	1.0000	0.0000	2.3000	Placebo							
26	12.0000	1.0000	0.0000	1.5000	Placebo							
27	12.0000	1.0000	0.0000	3.0600	Placebo							
28	11.0000	1.0000	0.0000	3.4900	Placebo							
29	11.0000	1.0000	0.0000	2.1200	Placebo							
30	8.0000	1.0000	0.0000	3.5200	Placebo							
31	8.0000	1.0000	0.0000	3.0500	Placebo							
32	8.0000	1.0000	0.0000	2.3200	Placebo							
33	8.0000	1.0000	1.0000	3.2600	Placebo							
34	5.0000	1.0000	1.0000	3.4900	Placebo							
35	5.0000	1.0000	0.0000	3.9700	Placebo							
36	4.0000	1.0000	1.0000	4.3600	Placebo							
37	4.0000	1.0000	1.0000	2.4200	Placebo							
38	3.0000	1.0000	1.0000	4.0100	Placebo							
39	2.0000	1.0000	1.0000	4.9100	Placebo							
40	2.0000	1.0000	1.0000	4.4800	Placebo							
41	1.0000	1.0000	1.0000	2.8000	Placebo							
42	1.0000	1.0000	1.0000	5.0000	Placebo							
43												
44												
45												
46												

# How to analyze with SigmaPlot?

The screenshot displays the SigmaPlot software interface. The main window shows a data table with columns: 1-time, 2-status, 3-sex, 4-logWBC, 5-group, 6, 7, 8, 9, 10, 11, 12. The data rows show various time points, status values (0.0000, 1.0000), sex values (0.0000, 1.0000), logWBC values, and group labels (Treatment, Placebo). A dialog box titled 'Cox PH Model - Select Data' is open, showing a list of columns to select for the model. The 'Data for Covariate' section is set to '4-logWBC'. The 'Selected Columns' list includes '1-time 1', '2-Status for time, status, and covariates (risk factors)', '3-sex', '4-logWBC', and '5-group'. The 'Next' button is highlighted.

	1-time	2-status	3-sex	4-logWBC	5-group
13	11.0000	0.0000	0.0000	2.6000	Treatment
14	10.0000	0.0000	0.0000	2.7000	Treatment
15	10.0000	1.0000	0.0000	2.9600	Treatment
16	9.0000	0.0000	0.0000	2.8000	Treatment
17	7.0000	1.0000	0.0000	4.4300	Treatment
18	6.0000	0.0000	0.0000	3.2000	Treatment
19	6.0000	1.0000	0.0000	2.3100	Treatment
20	6.0000	1.0000	1.0000	4.0600	Treatment
21	6.0000	1.0000	0.0000	3.2800	Treatment
22	23.0000	1.0000	1.0000	1.9700	Placebo
23	22.0000	1.0000	0.0000	2.7300	Placebo
24	17.0000	1.0000	0.0000	2.9500	Placebo
25	15.0000	1.0000	0.0000	2.3000	Placebo
26	12.0000	1.0000	0.0000	1.5000	Placebo
27	12.0000	1.0000	0.0000	3.0600	Placebo
28	11.0000	1.0000	0.0000	3.4900	Placebo
29	11.0000	1.0000	0.0000	2.1200	Placebo
30	8.0000	1.0000	0.0000	3.5200	Placebo
31	8.0000	1.0000	0.0000	3.0500	Placebo
32	8.0000	1.0000	0.0000	2.3200	Placebo
33	8.0000	1.0000	1.0000	3.2600	Placebo
34	5.0000	1.0000	1.0000	3.4900	Placebo
35	5.0000	1.0000	0.0000	3.9700	Placebo
36	4.0000	1.0000	1.0000	4.3600	Placebo
37	4.0000	1.0000	1.0000	2.4200	Placebo
38	3.0000	1.0000	1.0000	4.0100	Placebo
39	2.0000	1.0000	1.0000	4.9100	Placebo
40	2.0000	1.0000	1.0000	4.4800	Placebo
41	1.0000	1.0000	1.0000	2.8000	Placebo
42	1.0000	1.0000	1.0000	5.0000	Placebo

# How to analyze with SigmaPlot?

The screenshot displays the SigmaPlot software interface. The main window shows a report titled "Cox PH Model 3 Report". The report content is as follows:

**Cox Regression - Proportional Hazards Model** Montag, Dezember 03, 2012, 09:05:31

**Data source:** Data 1 in Notebook2

**Event labels**      **Censor labels**  
 1                      0

**Time unit:** None

Cases	Missing	Events	Censored	% Censored
42	0	30	12	29

**Categorical Covariates:**  
**Covariates**      **Reference Group**  
 sex                      1  
 group                      Placebo

**Regression Results:**  
 Breslow method of ties.  
 Variable selection method = Complete  
 Maximum log likelihood = -72.109  
 Computed in 5 iterations using a tolerance of 1E-008.

**Testing Global Null Hypothesis:**

	df	Statistic	P Value
Likelihood Ratio	3	43.752	<0.001
Global Chi-Square	3	44.571	<0.001

The Likelihood Ratio test and the Global Chi-Square test (also called Score test) usually give compatible results. If the significance results differ, use the p-value from the Likelihood Ratio Test.

Significance Level = 0.050  
 One or more coefficients in the hazard model are significantly different than zero ( $P = <0.001$ ). It follows that at least one covariate is significantly associated with survival time.

**Summary Info**

Field	Value
Created	03.12.2012 09:0...
Modified	03.12.2012 09:0...
Author	hielsche
Description	Report

# How to analyze with SigmaPlot?

The screenshot displays the SigmaPlot software interface with the 'Cox PH Model 3 Report' open. The report includes the following sections:

**One or more coefficients in the hazard model are significantly different than zero ( $P = <0.001$ ). It follows that at least one covariate is significantly associated with survival time.**

**Model Estimates:**

Covariate	Coefficient	StdErr	Wald Chi-Square	P Value
sex_0	-0.263	0.449	0.343	0.558
logWBC	1.594	0.330	23.321	<0.001
group_Treatment	-1.391	0.457	9.277	0.002

Significance Level = 0.050  
The hazard rate is significantly affected by:  
logWBC  
group\_Treatment

**Confidence Intervals:**

Covariate	Coefficient	95%Conf-L	95%Conf-U
sex_0	-0.263	-1.144	0.618
logWBC	1.594	0.947	2.240
group_Treatment	-1.391	-2.286	-0.496

**Covariate Hazard Ratio 95%Conf-L 95%Conf-U**

sex_0	0.769	0.319	1.855
logWBC	4.922	2.578	9.397
group_Treatment	0.249	0.102	0.609

The hazard ratio for a covariate is the proportional change in the hazard rate due to a unit change in the value of the covariate. If the hazard ratio refers to a group from a categorical covariate, then it also equals the ratio of the hazard rate of the group to the hazard rate of the reference group.

Since the value 1.0 lies in the confidence interval for the group sex\_0, there is no significant difference between the hazard rates of sex\_0 and the reference group sex\_1.

**Survival Table:**  
The covariate-adjusted survival probabilities and the cumulative hazard function are computed using the median value of each covariate. The Kaplan-Meier probabilities are provided for comparison.

# Competing Risk

Recap: Fundamental assumption of independence of event and censoring time distribution:

- Hazard of censored individuals = hazard of individuals remaining at risk
- risk for individual is not to change by censoring
- Violation leads to bias in Kaplan-Meier estimates

# Examples

## Violations

- loss to follow-up: if more healthy individuals drop out, survival curve will overestimate event rate
- loss to follow-up: toxicity-related drop-out, bias between (treatment-)groups
- competing risk: more than one possible cause of failure, competing risk event precludes occurrence of event of interest or fundamentally alters the probability for this event
- Death due to cancer may be of interest, and death due to other causes (surgical mortality, old age) are competing risks.
- One could be interested in time to relapse with death due to any cause being a competing risk.

# Approaches

- 1 Cause-specific hazards (CSH)
- 2 Cumulative Incidence Function (CIF)/ sub-distribution hazards (SH)
  - there is no longer a one-to-one relation between hazard and survival for a certain type of event
  - both approaches have pros/cons
  - difference in interpretation/generalization

# Cause-specific hazard

CSH: instantaneous rate of occurrence of event of cause  $k$  amongst the patients still event-free

$$h_k(t) = f_k(t)/S(t)$$

- cumulative CSHs  $H_k(t)$  can be estimated and displayed
- overall hazard  $h(t)$  is sum of all CSHs
- the survival function  $S(t) = 1 - \sum_{j=1}^k F_j(t) = \exp(-\sum_{j=1}^k H_j(t))$ ,  
This survival function is the probability of not having an event of any cause at time  $t$ .
- $S_k(t) = \exp(-H_k(t))$  can be estimated but is not interpretable as marginal survival function



# Cause-Specific Hazard (CSH) - inference

- Failures due to competing event(s) are censored
- non-parametric estimation of cumulative CSH: Nelson-Aalen estimator based on cause  $k$ :

$$\widehat{H}_k(t) = \sum_{t_i \leq t} \frac{d_i I(\text{cause}=k)}{n_i}$$

- non-parametric group comparison: log-rank test
- cause-specific Cox model
- BUT: KM estimator is biased

# Cumulative Incidence Function (CIF)

Probability of occurrence of a given event  $k$  by time  $t$ :

CIF for failure type  $k$ :  $CIF_k = F_k(t) = P(T \leq t, \text{cause} = k)$

- if  $k=1$ ,  $CIF = 1 - S(t)$ , i.e.  $1 - KM$  in estimation
- CIF is a crude probability, all risks act on the population, CIF depends on all CSHs
- $CIF_k$  always  $< 1$  if  $k > 1$ , hence not a proper probability distribution
- $CIF_k(t) \leq 1 - S_k(t)$ , KM estimates biased, sum across all causes can exceed 1
- use CIF for graphical display in competing risk data

# Cumulative Incidence Function (CIF)

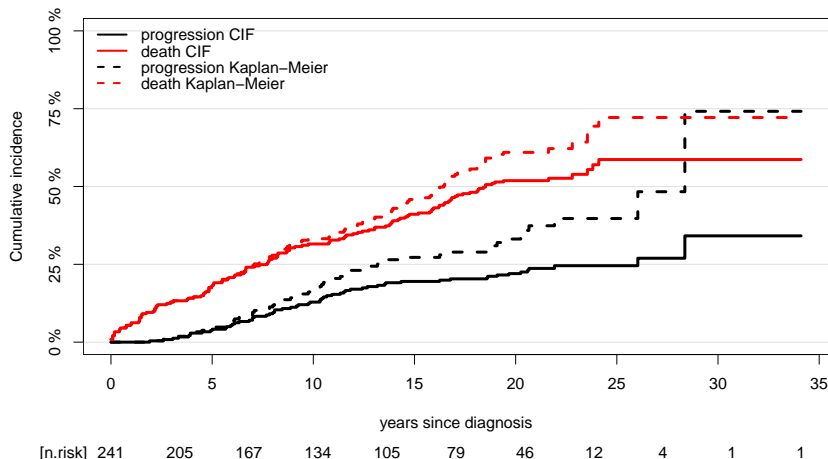
- Aalen-Johansen estimator (non-parametric), like a cause-specific Nelson-Aalen estimator weighted by the overall survival function (KM estimates):

$$\widehat{CIF}_k(t) = \sum_{t_i \leq t} \frac{d_i I(\text{cause}=k)}{n_i} * KM_{t-1}$$

- non-parametric group comparison: Gray's logrank type test compares CIFs

# MGUS Example Data (Kyle et al)

CIF vs. 1-KM



# Subdistribution hazards

subdistribution hazard (SH)  $i_k(t) = f_k(t)/(1 - F_k(t))$

- SH can be estimated from CIF
- Fine and Gray model with competing events being modelled
- Subjects with competing event are kept in risk set

# Interpretation

- CIF is affected by/accounts for hazard of competing event: Gray's test correctly detects whether there is a difference in cumulative incidence between groups for a given event, whether that difference is caused by a difference in hazards between the groups for the event itself or by a difference in hazards for the competing events.
- Gray's test will indicate, for example, that a treatment is beneficial with respect to one event type, when in fact the treatment has simply increased the incidence of a competing event.
- the log-rank test correctly detects differences in cause-specific hazards, and, unless there is strong dependence between failure times, is largely unaffected by between-group differences in hazards for other competing events

# References

Clark, Bradburn, Love and Altman, 2003. Survival Analysis Part III: Survival Analysis Part II: Multivariate data analysis – an introduction to concepts and methods. British Journal of Cancer 89, 431-436.

Altman and Royston, 2006, The cost of dichotomizing continuous variables, BMJ, 332(7549): 1080.

SigmaPlot Statistics User Guide (PDF Manuals)

# How to get support

- The biostatistics division C060 provides statistical support for all scientific activities of the DKFZ from in vitro and animal studies to human subject.
- Request statistical support via email to [biostatistics-consulting@dkfz.de](mailto:biostatistics-consulting@dkfz.de)