CRS-MANNHEIM
Clinical Research Services

A supplement to the SAS survival guide – nonparametric regression

Karl Ernst Siegler
CRS-Mannheim GmbH
Grünstadt, Germany

PhUSE SDE
17 April 2013

Overview

- Introduction to survival time analysis and the concept of censoring.
- The nonparametric additive hazard model.
- Application to real life data.
- Conclusions.

Introduction to Survival Time Analysis - In the Beginning

- Application to demographic objectives (Lexis 1875).
- Analysis of life cycles in engineering sciences (Weibull 1939).

- The methods do not translate straightforward to clinical trials.
- E.g. life times of light bulbs:
  The complete life span from first use to failure is observed.

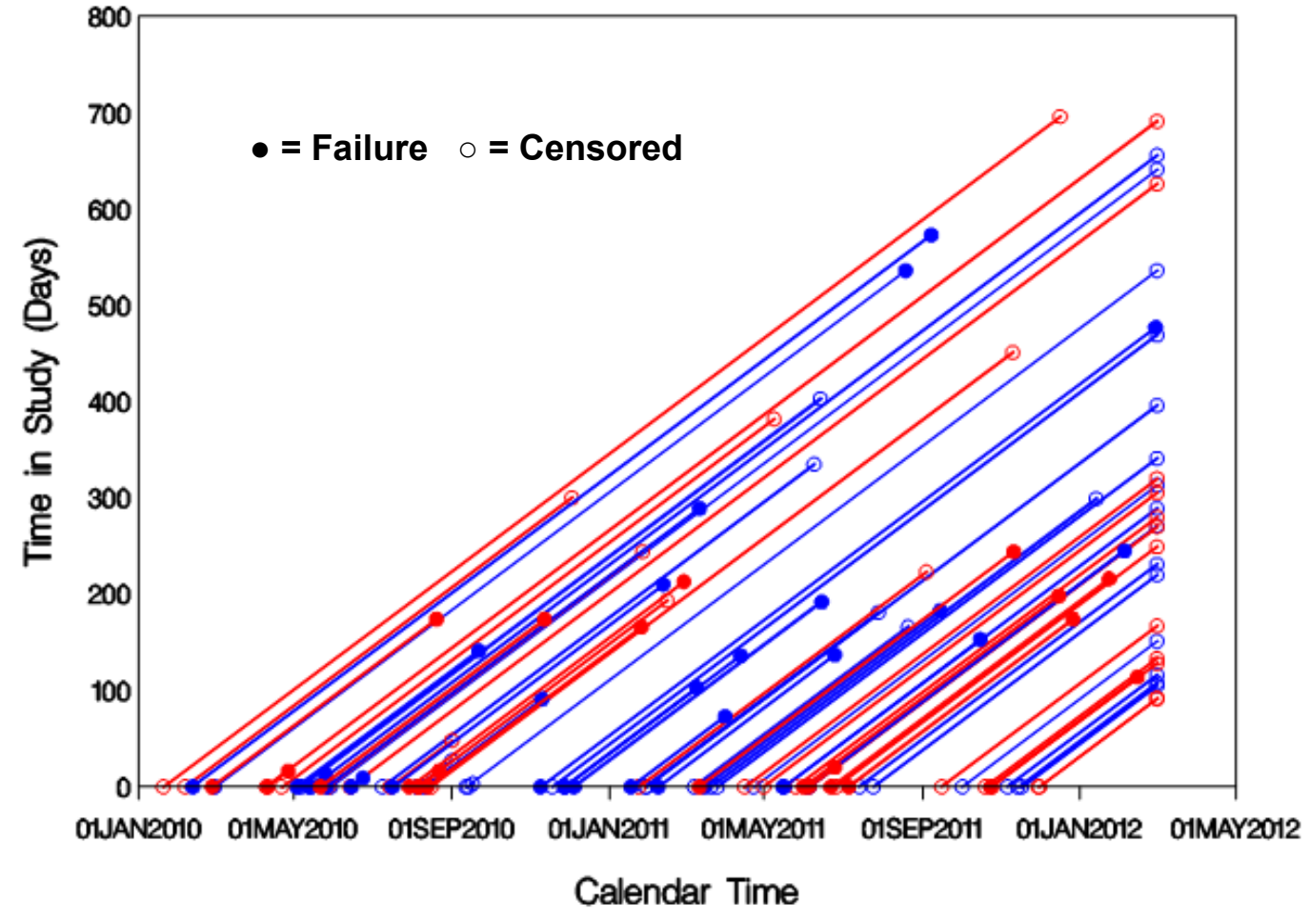
- In clinical trials:
  - Date of first occurrence of a disease in a patient might be unknown.
  - Date of death might be unknown.
  - Death related to other reasons might be observed.

Introduction to Survival Time Analysis - The Concept of Censoring

(Right) censored observations:

- **Some patients are alive at the end of the observation period. Some patients die due to circumstances not linked to the study.**
- **In both cases a certain “survival time" is known.**
- **Knowledge is incomplete, it covers the time under observation.**
- **Disregarding censoring information leads to under-estimation of survival probabilities.**

Introduction to Survival Time Analysis - Lexis Diagram



Introduction to Survival Time Analysis - Application in Clinical trials

- Estimation of the survival probability functions (Kaplan, Meier 1958).
- Statistical comparison of survival probability functions with Log-Rank Test (Peto, Peto 1972).

Introduction to Survival Time Analysis - Data Considerations

Kaplan-Meier Curves with Log-Rank Test

- **Observed survival time (in days, weeks or years).**
- **Censoring information (“censored", “complete Observation until death").**
- **Information about treatment arms.**

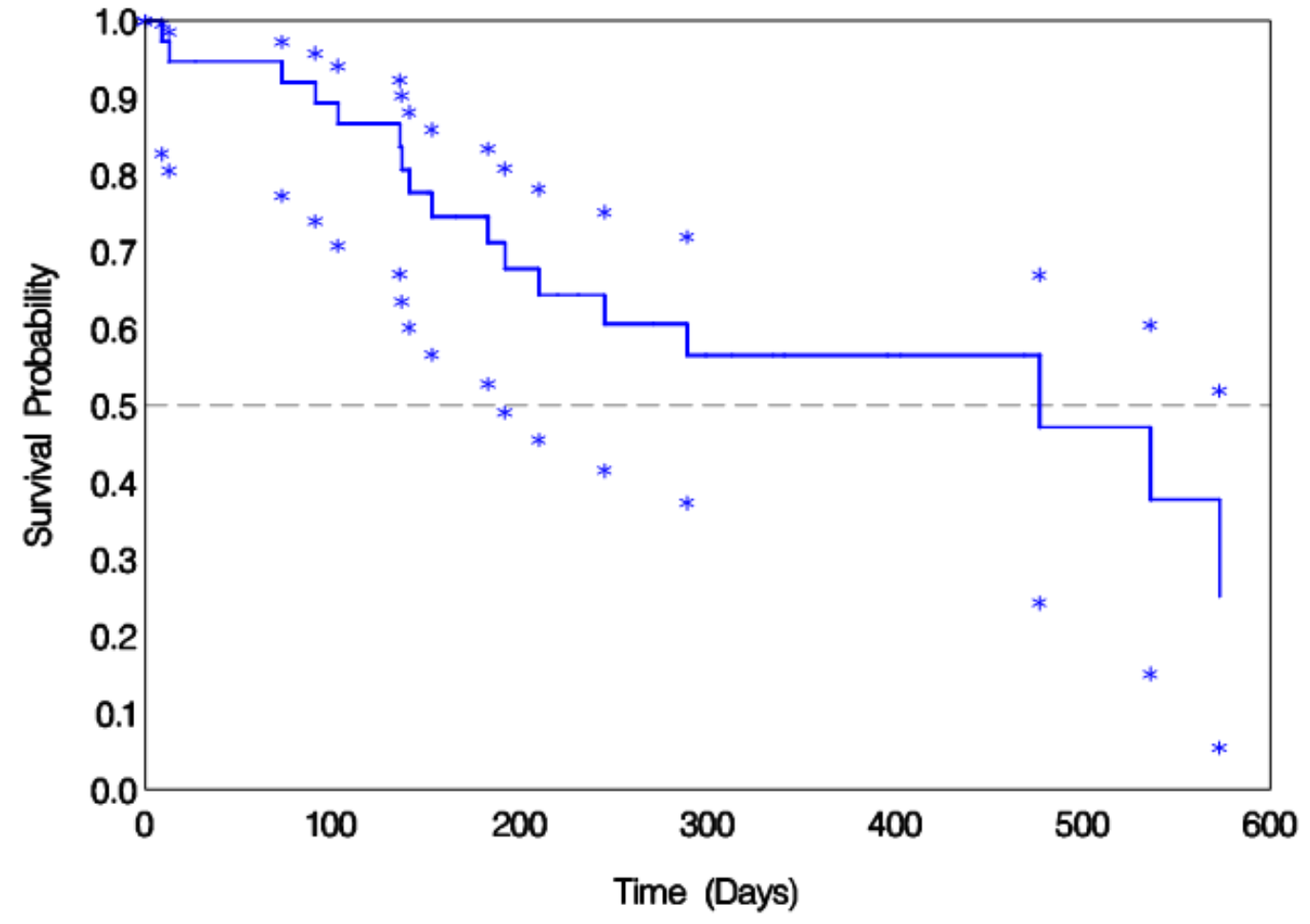
Patnr	Censored	Survival Time (Days)	Treatment
001	No	255	Test
002	No	307	Standard
003	Yes	269	Standard
004	Yes	17	Test

Introduction to Survival Time Analysis - Survival Time Analysis with SAS® Software

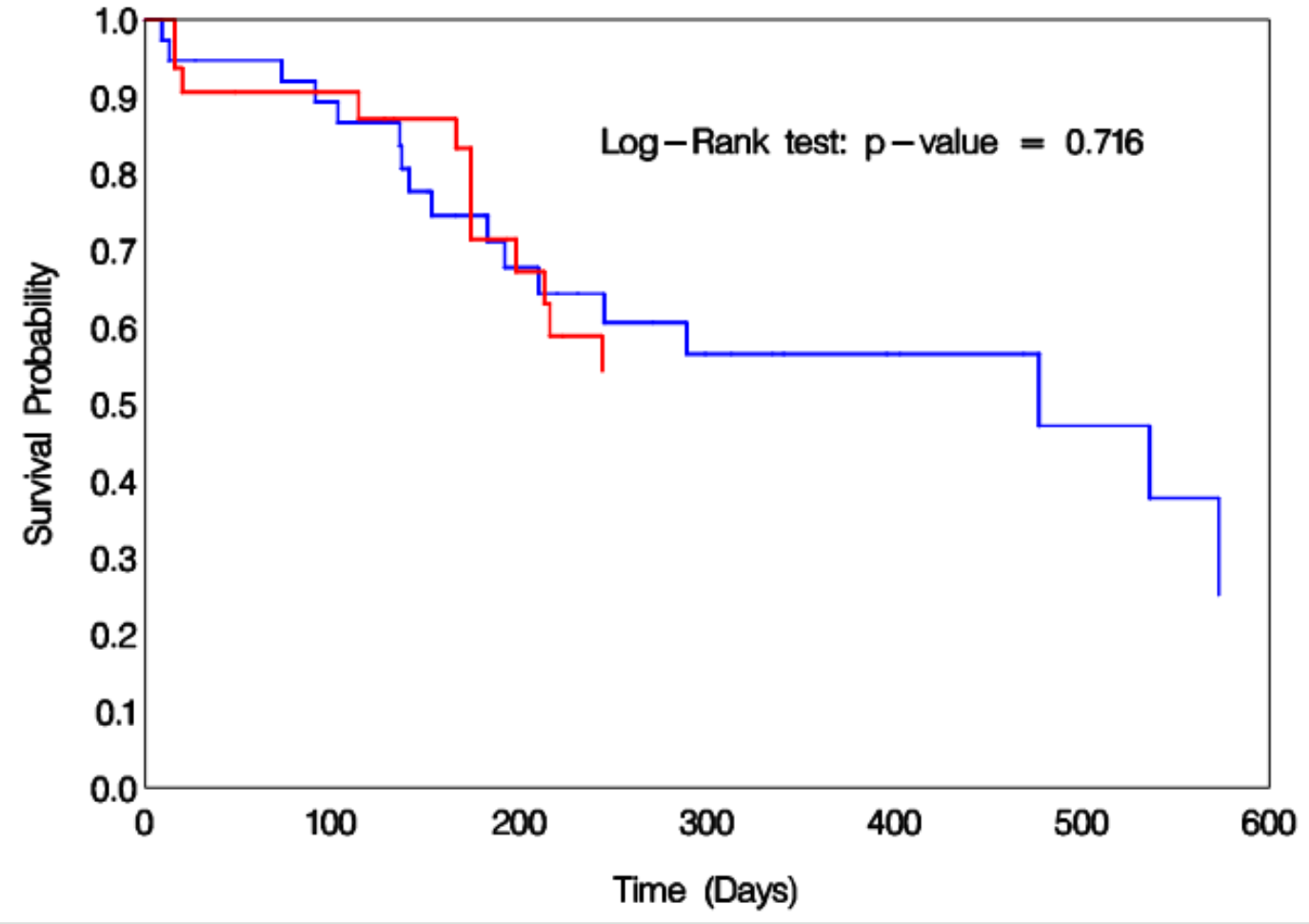
- **proc lifereg - Parametric models (e.g. Weibull Distribution) for failure time data with or without censored observations.**

- **proc lifetest - Estimation of survival probability functions and comparisons (Kaplan-Meier Curves, Log-Rank Tests and many more applications).**

Introduction to Survival Time Analysis
- Kaplan-Meier Curve with point-wise 95% CIs



Introduction to Survival Time Analysis
- Comparison of two Kaplan-Meier Curves



Introduction to Survival Time Analysis - Regression Models

- Estimating the influence of several covariates with Cox Proportional Hazards Model (Cox 1972), regression methods.

- Hazard function (or rate) describes the instantaneous probability of death as a function of time.

Introduction to Survival Time Analysis - Data Considerations

Hazard Regression

- **Observed survival time (in days, weeks or years).**
- **Censoring information (“Censored", “Complete Observation until Death").**
- **Information about treatment arms.**
- **Covariates as further variables, e.g. heart rate, tumor staging etc.**

Patnr	Censored	Survival Time (Days)	Treatment	Heart Rate (bpm)	Tumor Staging
001	No	255	Test	88	II
002	No	307	Standard	93	III
003	Yes	269	Standard	95	IV
004	Yes	17	Test	88	II

Introduction to Survival Time Analysis - Regression Analysis with SAS® Software

- **proc phreg - Regression analysis, Cox Proportional Hazards Model, allows for time varying covariates with the counting process style of input.**

- **proc surveyphreg - Regression analysis based on the Cox Proportional Hazards Model for more complex designs.**

Introduction to Survival Time Analysis - Modern Mathematics

- Growing interest in counting processes and martingale theory in the 1980s and 1990s.
- Another regression model for the estimation of covariate influence on survival appeared: Nonparametric Additive Hazard Model.
- Referred to as "Aalen-model“ (O.O. Aalen 1980, 1989 and 1993).
- Regression models in survival analysis: Hazard functions are modeled, not survival probability functions.
- Results of hazard regression:
  For each covariate a regression parameter $ß_i$ is estimated.
  Statistical Test with Hypothesis H0: $ß_i=0$.

The Nonparametric Additive Hazard Model - Linear Hazard Model

$$\alpha_i(t) = \gamma_0(t) + \gamma_1(t) \cdot Z_{i1}(t) + \cdots + \gamma_p(t) \cdot Z_{ip}(t)$$

α_i	**Hazard function**
$\gamma_j(t)$	**Regression functions in contrast to regression coefficients in Cox-Model**
$Z_{ij}(t)$	**Covariate matrix contains covariate information might be time varying**

The Nonparametric Additive Hazard Model - Cumulative Regression Functions (CRF)

$$\Gamma(t) = \int_0^t \gamma(s)ds$$

- **Theory of counting processes:**
  The cumulative (or integrated) regression function is estimated.

- **A sequence of cumulative sums over time.**

The Nonparametric Additive Hazard Model - Nonparametric Estimation of CRF

$$\hat{\Gamma}(t) = \sum_{T_k \le t} \left[\left(Y(T_k)^T Y(T_k) \right)^{-1} Y(T_k)^T \cdot I_k \right]$$

- Structure of estimator is reasonable simple: $(X'X)^{-1}(X'y)$.
- T_k = Observed survival times.
- $Y(T_k)$ = Modified covariate matrix.
- I_k = Vector, indicating the current survival time. (k-th Value = 1, all other values zero).
- This is used for informative graphics: Aalen-Plots.

The Nonparametric Additive Hazard Model - Design-Matrix

$$Y(T_i) = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \\ 1 & Z_{i1}(T_i) & \cdots & Z_{ip}(T_i) \\ \vdots & \vdots & & \vdots \\ 1 & Z_{n1}(T_i) & \cdots & Z_{np}(T_i) \end{pmatrix}$$

- $Y(T_k)$ = Modified covariate matrix.
- Rows up to current survival time are set to zero.
- Allows incorporation of time varying covariates.

The Nonparametric Additive Hazard Model - Confidence Intervals and Statistical Tests

$$\hat{\Omega}(t) = \sum_{T_k \le t} Y^{-}(T_k) \cdot diag(I_k) \cdot Y^{-}(T_k)'$$

- Estimation of confidence intervals follows the same principles.
- Notations as before.
- Point-wise 95% confidence intervals in Aalen-Plots.

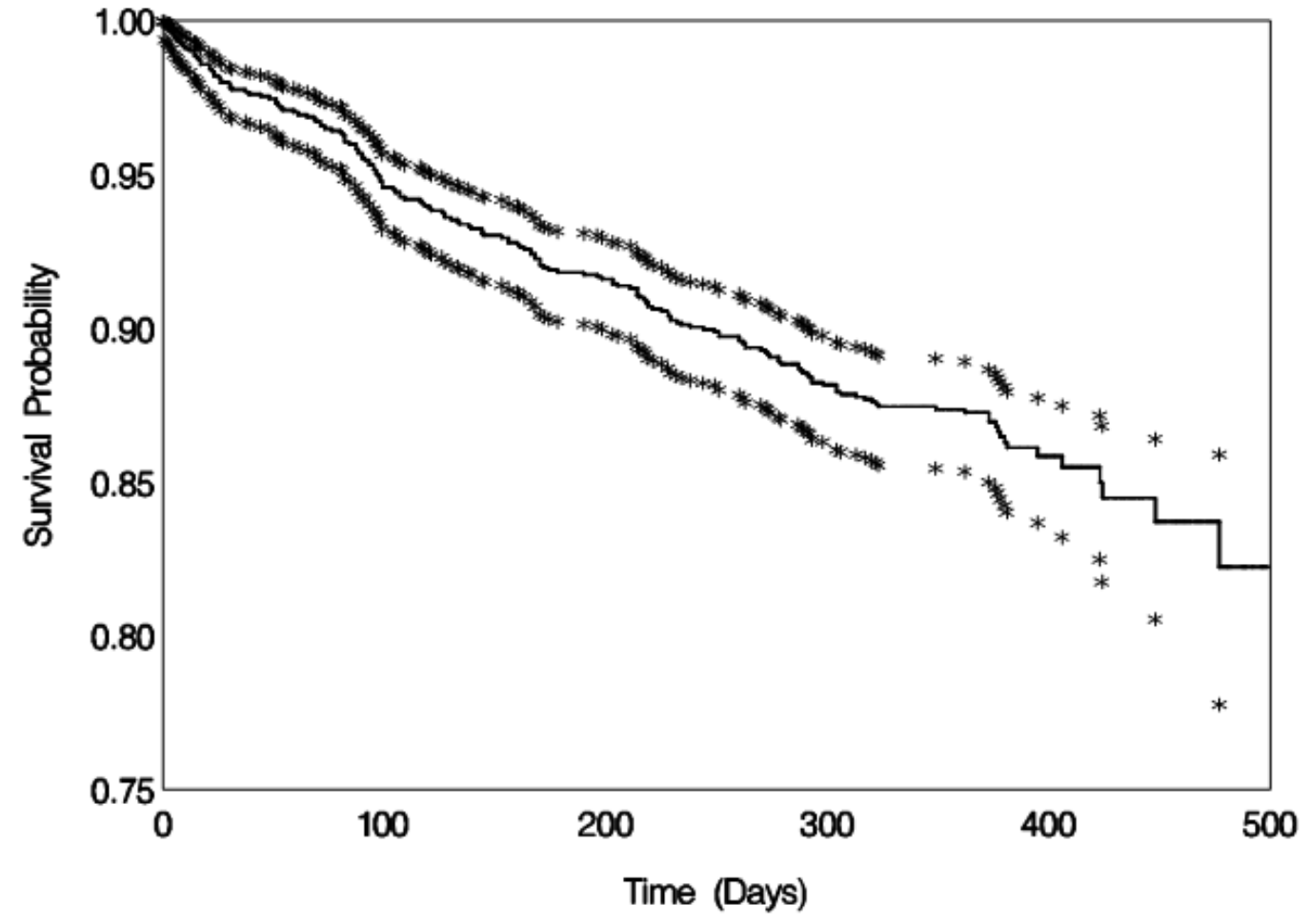
$$\hat{H}(t) = \sum_{T_k \le t} L(T_k) \cdot Y^{-}(T_k) \cdot I_k$$

- Estimation of test statistics follows the same principles.
- Notations as before.
- $L(T_k)$ is a weight process, e.g. number still alive at T_k.

Heart Failure Data
- Evidence Based Treatment of Heart Failure

- Evidence Based Treatment of Heart Failure (EVITA-HF).
- Prospective multi-center survey.
- 13 hospitals in Germany.
- Observation period started 2009.
- Now ca. 2800 patients.
- Follow-up completed ca. 1420 patients.
- Eight covariates analyzed.
- Observations with missing values in covariates omitted.
- 1347 observations left with 187 events (=deaths) and 1160 censored.
- Survival times (censored or not) from 1 day to 811 days (2.25 years).

Heart Failure Data
- Kaplan-Meier Curve with point-wise 95% CIs



Heart Failure Data
- Covariates

1. Gender (male / female).
2. Age at entry in years.
3. LVEF = Left Ventricular Ejection Fraction,
   volume of blood pumped out of the heart with each beat.
4. NYHA classification with values from I to IV,
   measure for physical performance “New York Heart Association”.
5. ICM = Ischemic Cardiomyopathy.
   Heart failure by poor oxygen supply of the heart
6. CMP = Cardiomyopathy or “Heart Muscle Disease“,
   heart failure by deterioration.
7. MI = Previous Myocardial Infarction.
8. Renal Failure, concomitant disease with a large prognostic value.

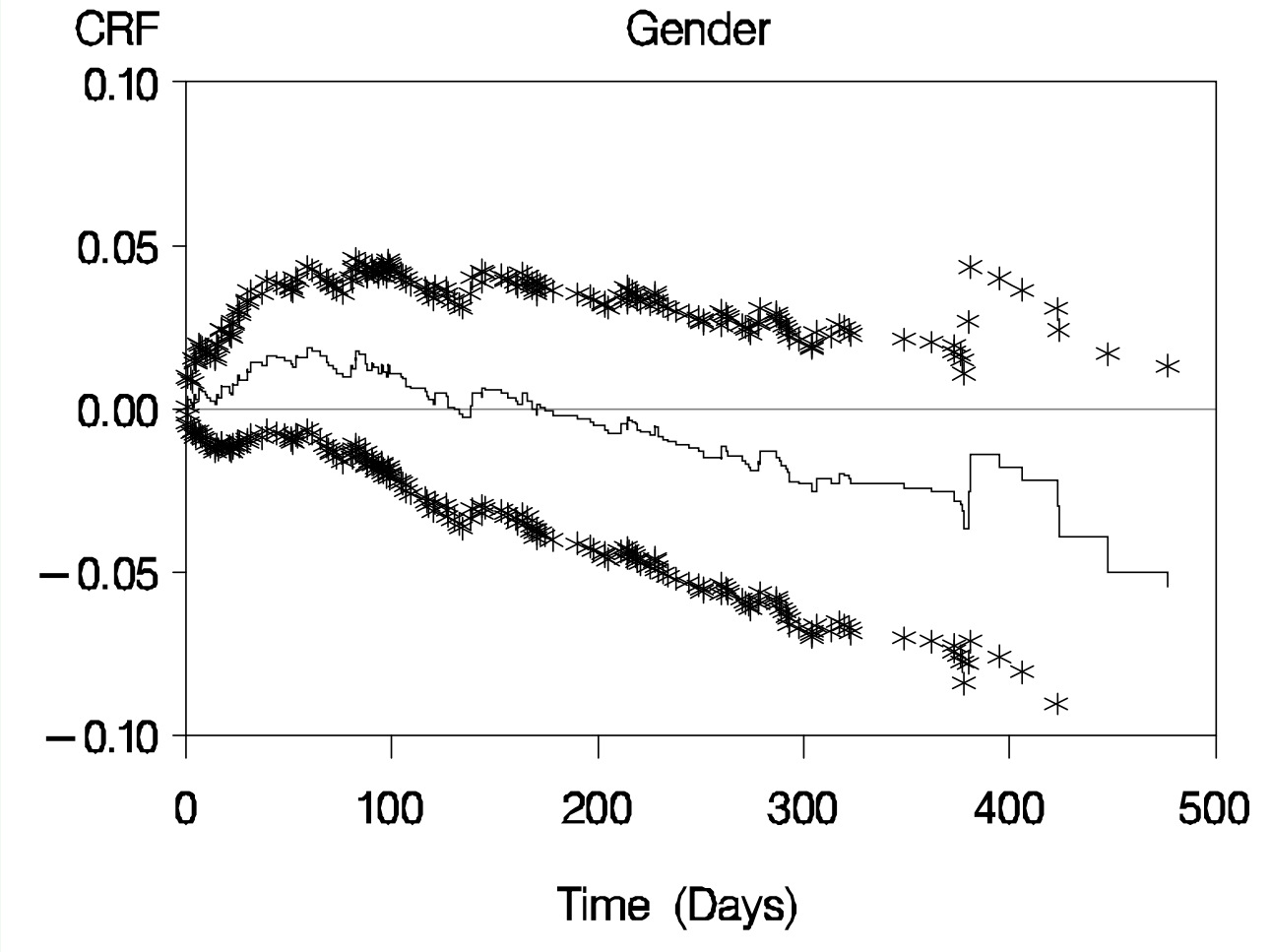
Heart Failure Data
- Results of Cox- and Aalen-Model

	p-Values for influence of covariates	
	Cox-Model	Aalen-Model
Gender	0.359	0.277
Age	0.007	0.005
LVEF	<0.001	<0.001
NYHA	0.005	0.010
ICM	0.304	0.174
CMP	0.194	0.146
MI	0.137	0.102
Renal Failure	<0.001	<0.001

Heart Failure Data
- Aalen-Plots: Covariate with no Influence on Hazard

- Covariate "Gender“
- p-Value = 0.277
- 95% confidence intervals include zero at each time point
- No influence on hazard rate

Heart Failure Data
- Aalen-Plots: Covariate with no Influence on Hazard

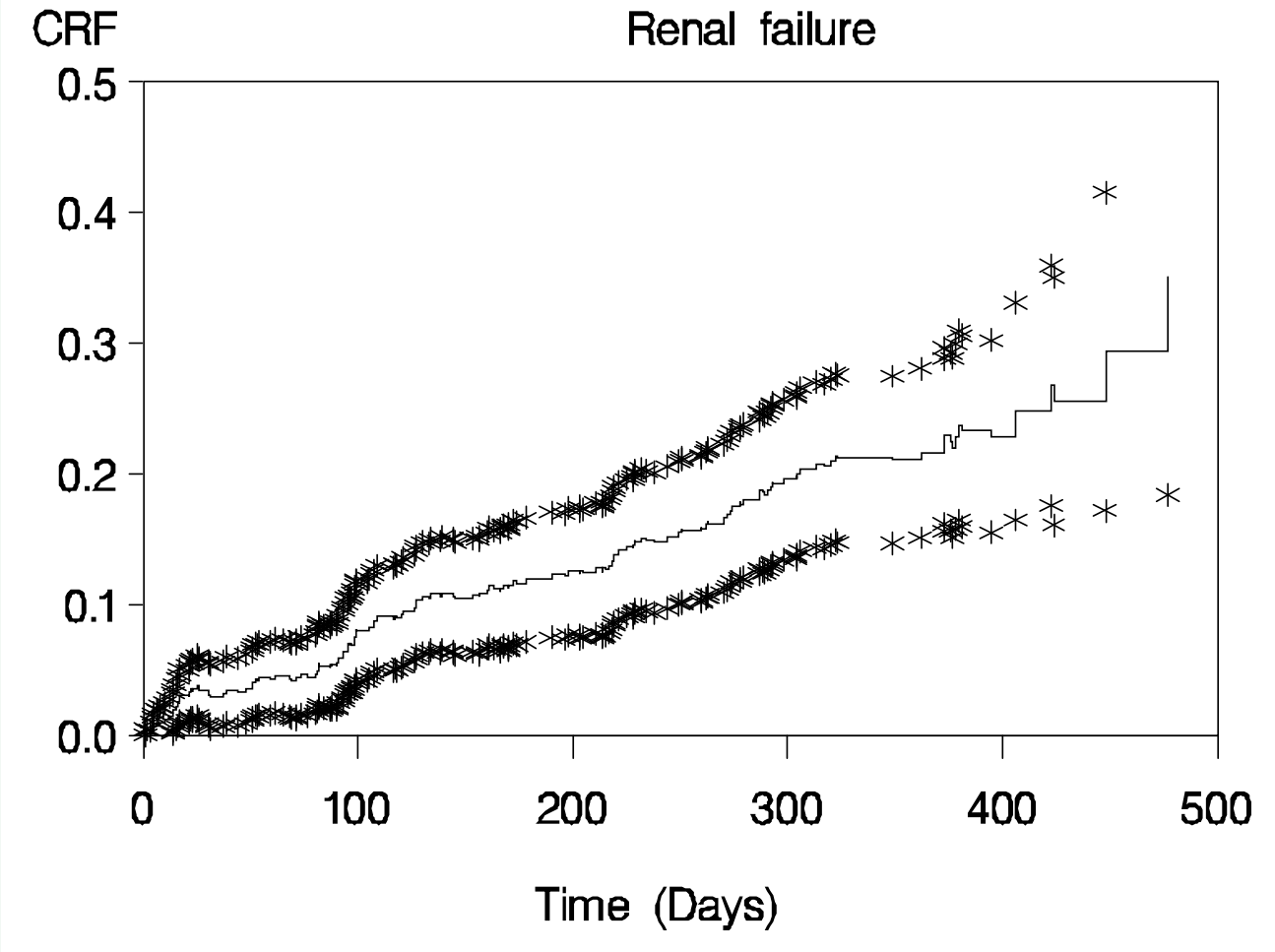


Heart Failure Data
- Aalen-Plots: Covariate with Persistent Influence

- Covariate “Renal Failure“.
- p-Value < 0.001.
- 95% confidence intervals do not include zero at any time point.
- Persistent influence on hazard rate.

- Direction of slope depends on coding:
  Ascending: Higher covariate values = higher risk.

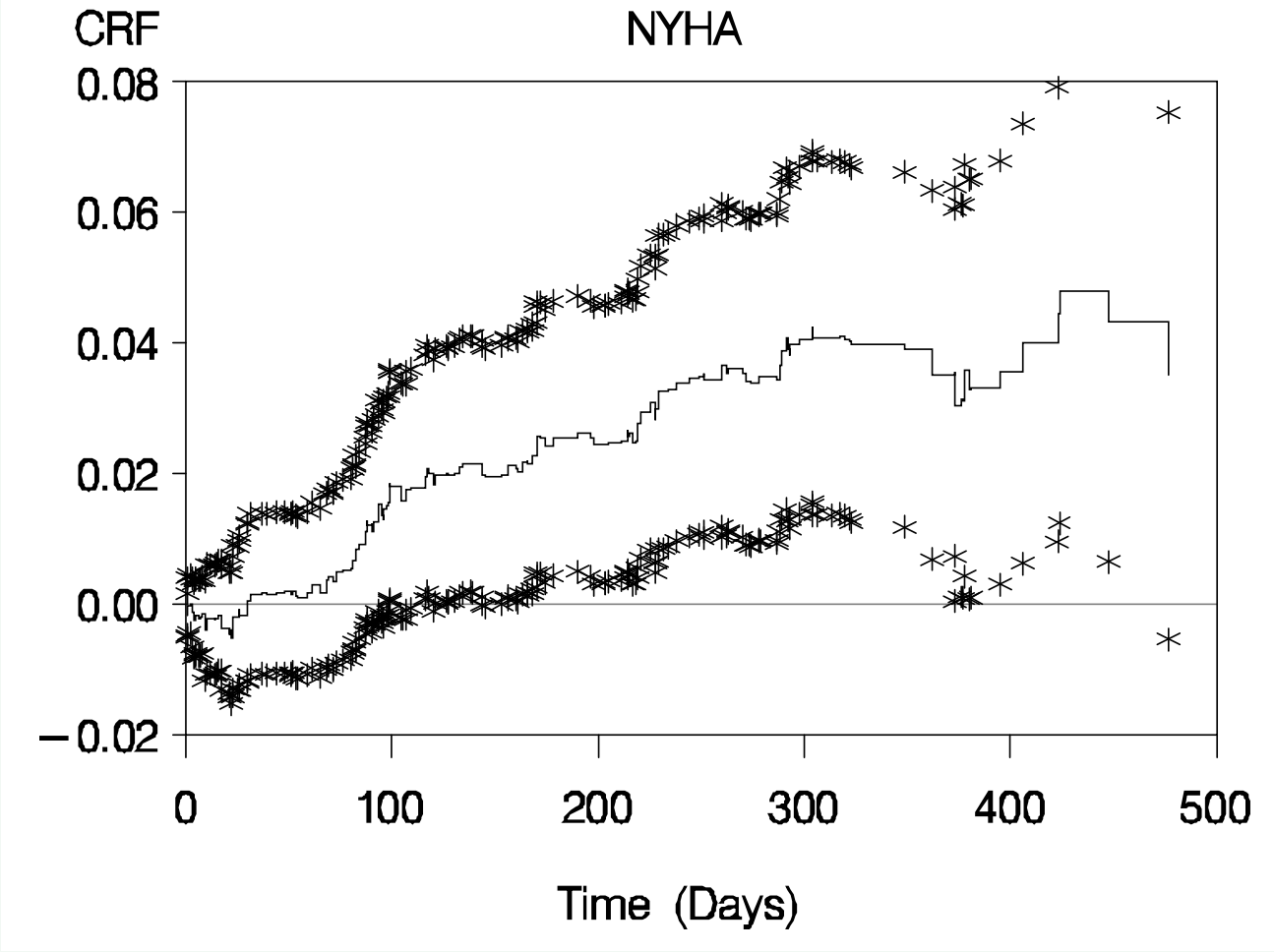
Heart Failure Data
- Aalen-Plots: Covariate with Persistent Influence



Heart Failure Data
- Aalen-Plot: Covariate with Time-Varying Influence

- Covariate "NYHA-Class“.
- p-Value = 0.010.
- 95% confidence intervals include zero up to 100 days,
  but they do not include zero after 100 days.
- No Influence on hazard in first 100 days
  afterwards patients with larger NYHA-classes have a higher risk.

Heart Failure Data
- Aalen-Plot: Covariate with Time-Varying Influence



Conclusion

- Model is described extensively in the statistical literature.
- Calculations are easy to do, no SAS procedures available.
- Experience in clinical trials is limited.
- Interpretation of Cumulated Regression Function is not easy.
- "Aalen-plots" can give additional insight in results of any survival analysis.
- Allows for all kinds of censoring schemes.
- Permits time varying covariates.

- Not all possibilities explored!