

A supplement to the SAS survival guide – nonparametric regression

**Karl Ernst Siegler
CRS-Mannheim GmbH
Grünstadt, Germany**

**PhUSE 2012
SP01**

Overview

- **Introduction to Survival Time Analysis and the Concept of Censoring**
- **The Nonparametric Additive Hazard Model**
- **Application to Real Life Data**
- **Conclusion**

History of 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**

History of Survival Time Analysis

- The Concept of Censoring

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**

- **Type of Censoring is denoted as Right Censoring**
Individual Time Axis is cut off at the right-hand Side
- **Left Censoring: Start of Observation Period is unknown**
e.g. Starting Point of a certain Disease
- **Interval Censoring with intermittent Observation Periods**

History of 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)**
- **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**

History of 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 β_i is estimated
Statistical Test with Hypothesis $H_0: \beta_i=0$

Introduction to Survival Time Analysis

- Data Considerations

Kaplan-Meier Curves

- **Observed Survival Time (in Days, Weeks or Years)**
- **Censoring Information ("Censored", "Complete Observation until Death")**

Comparison of Kaplan-Meier Curves with Log-Rank Test

- **Information about Treatment Arms**

Hazard Regression

- **Covariates as further Variables, e.g. Blood Pressure, Tumor Staging etc.**

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)**
- **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**

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

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

- 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

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

- $Y(T_k)$ = 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

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

- Estimation of Confidence Intervals follows the same Principles
- Notations as before
- Point-wise 95% Confidence in Aalen-Plots

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

- 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 - 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

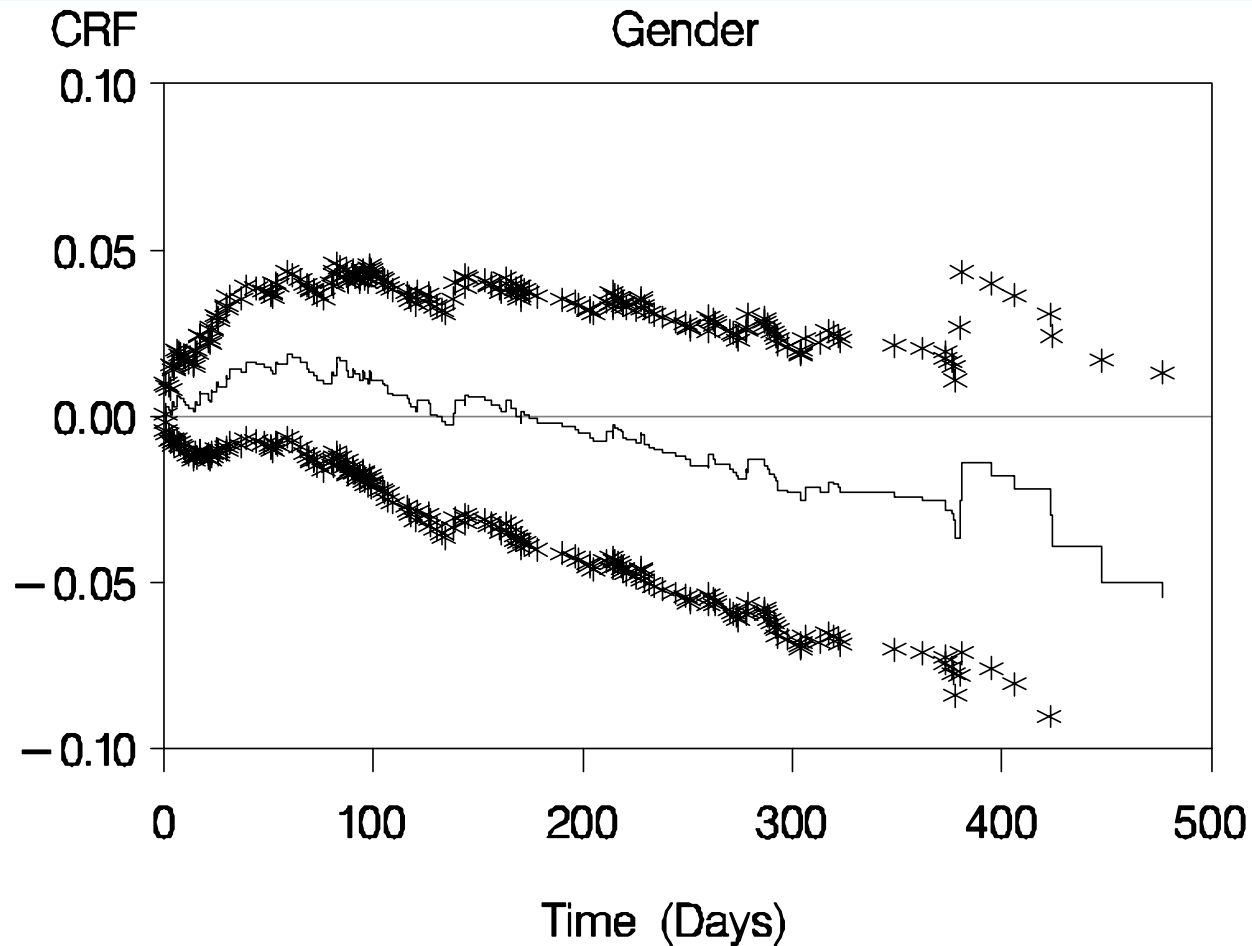
Heart Failure Data

- 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



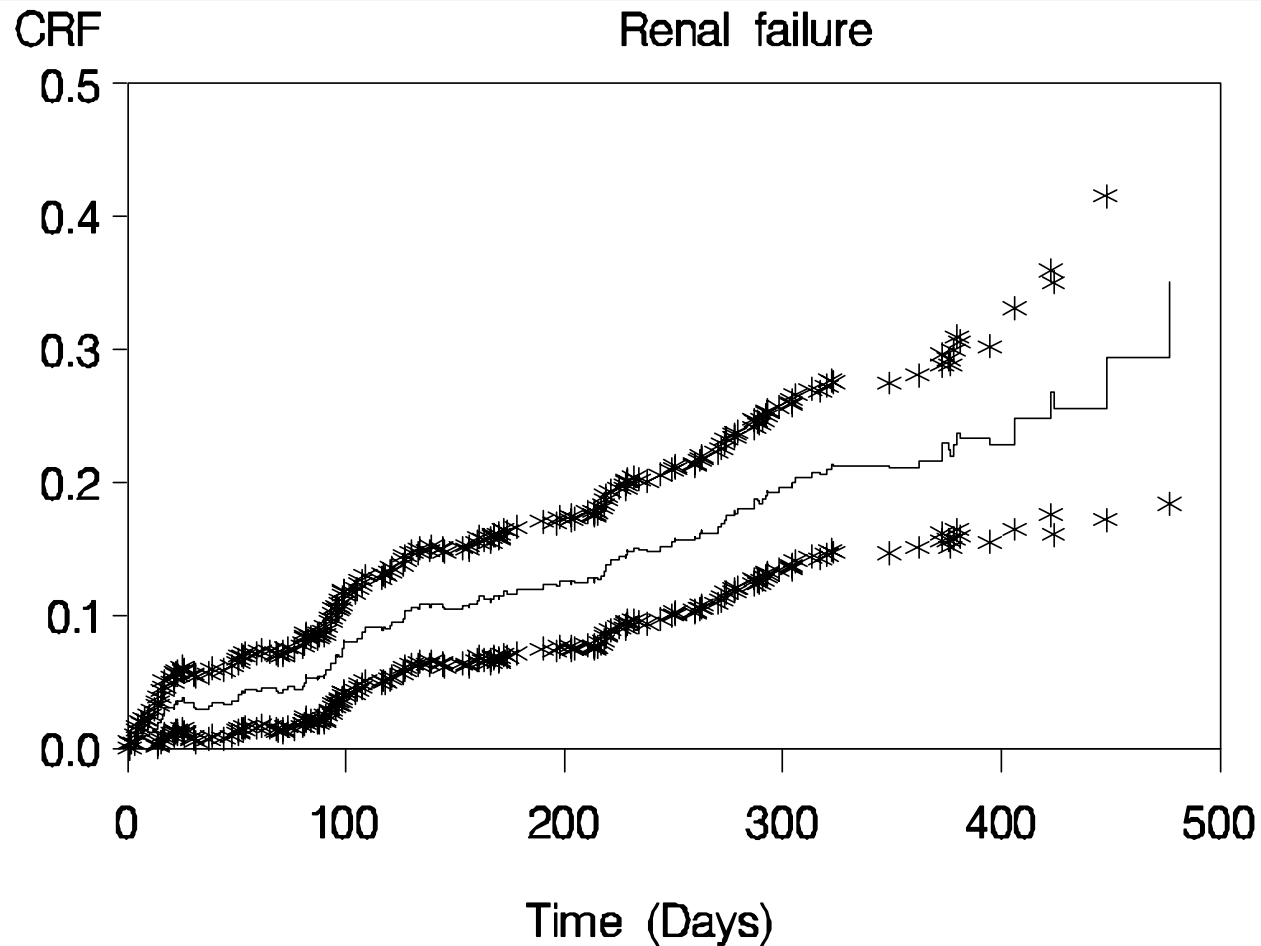
Heart Failure Data

- 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



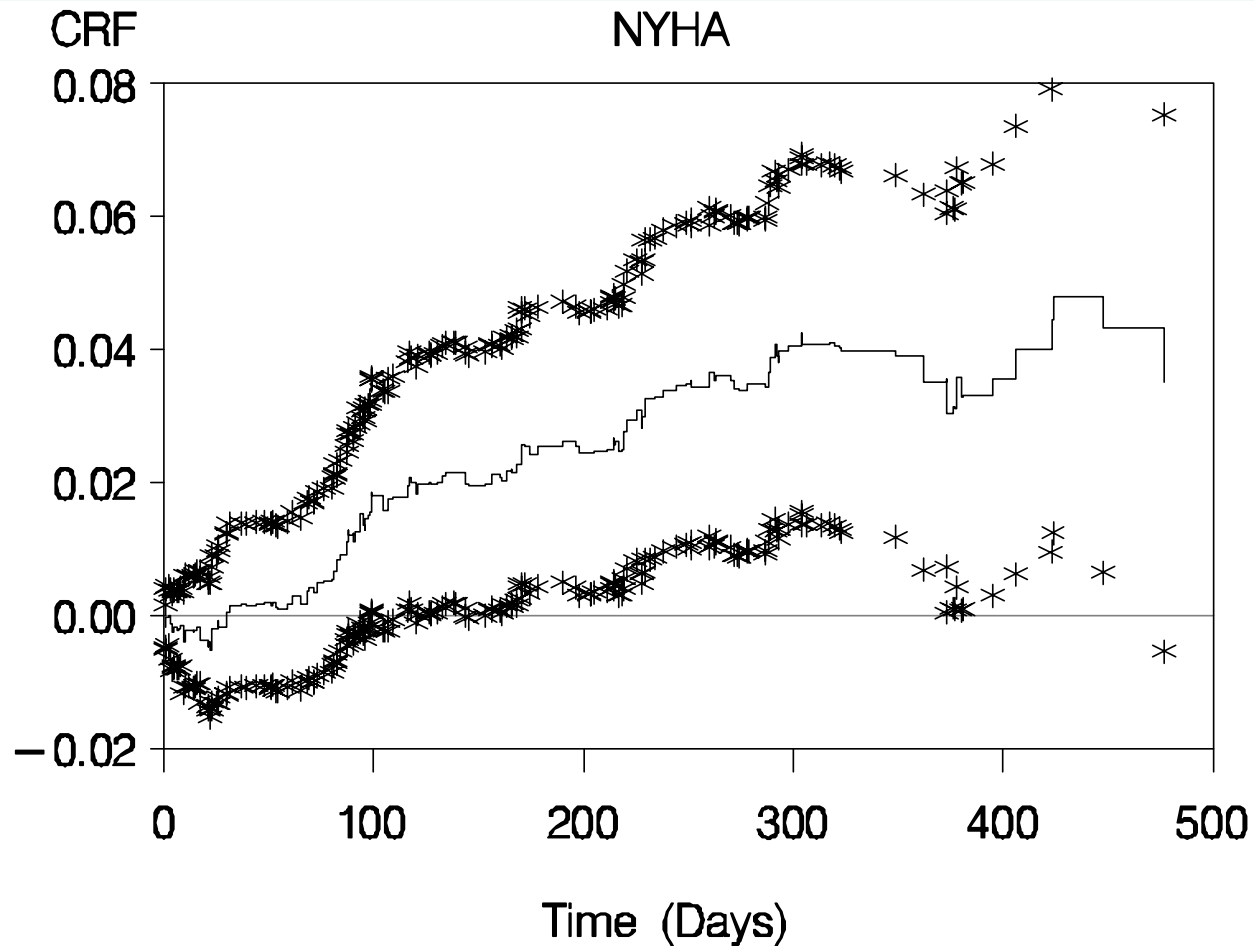
Heart Failure Data

- 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 Potentials were explored: Several published Approaches for describing a Change Point in Hazard Rates. These Methods might be translated to the Cumulative Regression Functions**

References

Comprehensive textbooks:

- Klein, J.P., Moeschberger, M.L. (1997) Survival analysis: Techniques for Censored and Truncated Data. Springer, New York.
- Andersen, P.K., Borgan, Ø., Gill, R.D., Keiding, N. (1993) Statistical Models based on Counting Processes. Springer, New York.

Thank you for your Attention!