

Überlebenszeitanalyse mit Hilfe generalisierter linearer und additiver Modelle

Ulrich Gehrman

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG)
Dillenburger Str. 27 , 51105 Köln



Übersicht

- Motivation
- Piecewise Exponential Model
- Poissonmodell mit Offset
- Anwendungsbeispiel Multiple Sklerose
- Fazit

Motivation

Drei gängige Typen von Überlebenszeit-Modellen:

- Parametrisch (Exponential-, Weibullverteilung, ...)
 - + Leichte Interpretierbarkeit
 - Modellvoraussetzungen oft nicht erfüllt
- Nicht-parametrisch (Kaplan-Meier, Log-Rank-Test)
 - + Keine Modellvoraussetzungen
 - Nicht ohne weiteres reduzierbar auf eine Maßzahl
 - Einfluss von Kovariablen nicht modellierbar
- Semi-parametrisch (Cox-Modell)
 - + Hazard Ratio zwischen Gruppen
 - + Einfluss von Kovariablen modellierbar
 - Keine direkte Schätzung der Überlebenszeit-Funktion (dadurch keine Einbettung in GLMs)

Motivation

Die hier vorgestellten Methoden bilden einen Kompromiss zwischen parametrischen und nicht-parametrischen Methoden:

- Stückweises Exponentialmodell (Piecewise Exponential Model, PEM)
→ Überlebenszeit-Funktion und Hazardrate direkt schätzbar
- Die Stückelung reduziert die Modellvoraussetzungen.
- Einbettung in generalisierte lineare und additive Modelle
→ Flexible Möglichkeiten der Schätzung der Kovariablen-Einflüsse (Selektion, Größe und Form der Einflüsse)



Piecewise Exponential Model

Daten:

n : Anzahl der Beobachtungen (Patienten, Individuen)

i : Nr. des Individuums; $i \in \{1, \dots, n\}$

y_i : Zensierungsvariable:

1 falls Tod bzw. Ereignis eingetreten

0 sonst

t_i : Überlebenszeit bzw. Zeit bis zur Zensierung

p : Anzahl der Kovariablen

x_i : p -dimensionaler Kovariablenvektor des i -ten Individuums

Piecewise Exponential Model

- Die Zeitachse wird in ein Gitter aus k Intervallen eingeteilt:

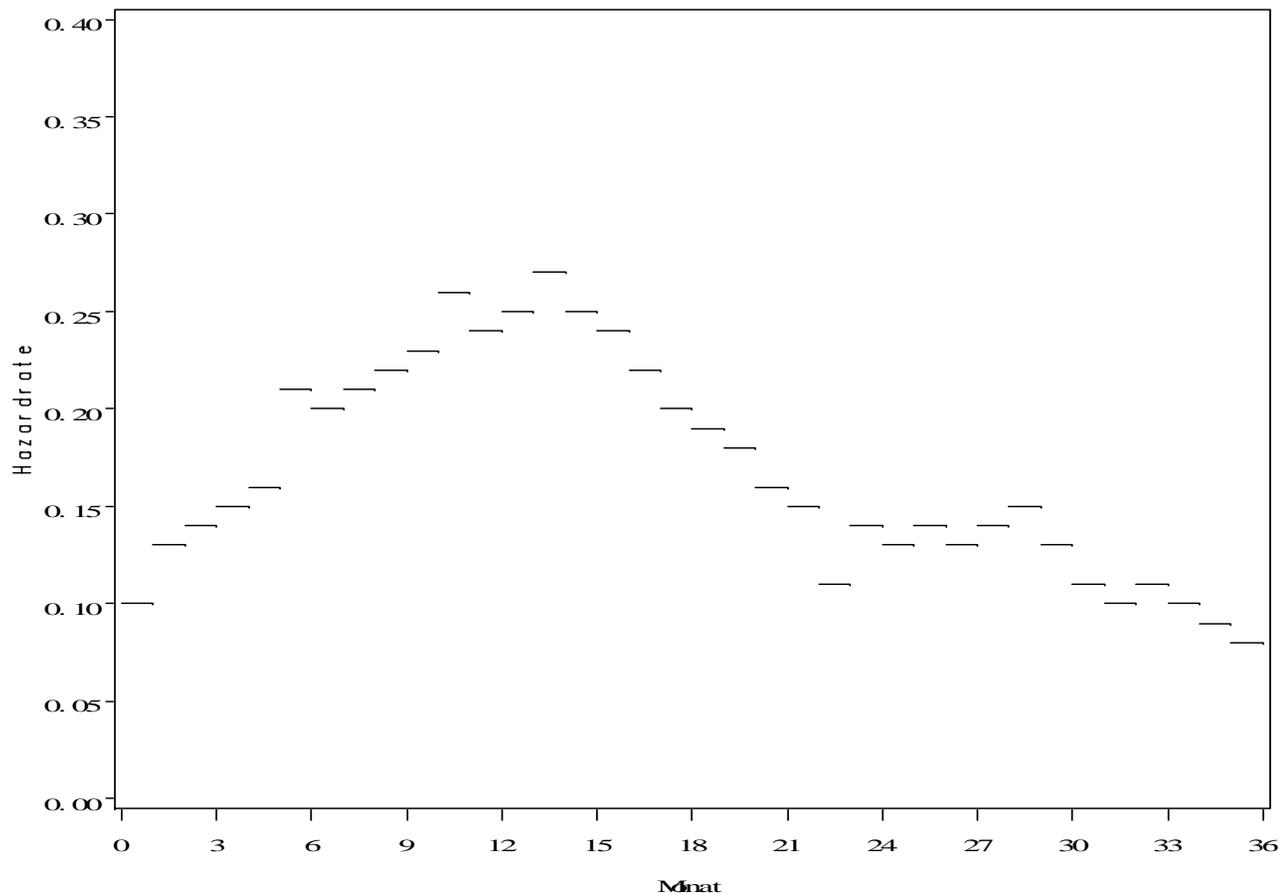
$$I_1=[0,a_1], I_2=(a_1,a_2], \dots, I_s=(a_{s-1},a_s], \dots, I_k=(a_{k-1},\infty)$$

- Die Hazardrate ist konstant innerhalb eines Intervalls, kann aber zwischen den Intervallen beliebig springen:

$$\lambda(t|x_i) = \lambda_{is} \quad (s = 1, \dots, k)$$

Piecewise Exponential Model

Skizze: Hazardraten eines Individuums im PEM



Piecewise Exponential Model

Ziel: Einbettung des PEM in ein generalisiertes lineares Modell (GLM)

Weg: Expansion des Original-Datensatzes:
Eine Datenzeile pro Patient und Zeitintervall

Beispiel:

- Originaldaten:

i	y_i	t_i [Tage]	$x_{i1} =$ Alter	$x_{i2} =$ Ge- schlecht
1	1	20	40	W
2	0	120	35	W
3	1	75	50	M

Piecewise Exponential Model

- Datensatz nach Expansion (Intervalllänge 30 Tage):

i	$S = \text{Inter-vall-Nr.}$	y_{is}	$\Delta_{is} = \text{Inter-vall-Länge}$	$x_{i1} = \text{Alter}$	$x_{i2} = \text{Ge-schlecht}$
1	1	1	20	40	W
2	1	0	30	35	W
2	2	0	30	35	W
2	3	0	30	35	W
2	4	0	30	35	W
3	1	0	30	50	M
3	2	0	30	50	M
3	3	1	15	50	M

→ Poissonmodell mit Offset anwendbar



Poissonmodell mit Offset

- Exponentialverteilung auf den Δ_{is}
→ Poissonverteilung der y_{is} :

$$y_{is} \sim \text{Po}(\mu_{is})$$

$$E(y_{is}) = \mu_{is} = \Delta_{is} * \exp(x_i'\beta + s\gamma)$$

$$\log(\mu_{is}) = \log(\Delta_{is}) + x_i'\beta + s\gamma$$

- GLM: Poissonmodell mit log-Link und Offset
- Offset: $\log(\Delta_{is})$
- Zeit s als gewöhnliche Kovariable

Poissonmodell mit Offset

- Standardsoftware anwendbar (SAS, S-Plus ...)
- Erweiterung auf nonparametrische Splines möglich (generalisierte additive Modelle, GAM)
- Auch fraktionale Polynome möglich
- Einbeziehung zufälliger Effekte möglich, z.B. Studieneffekte

Anwendungsbeispiel MS

- Multiple Sklerose, 2 Krankheitsphasen:
 - schubförmig wiederkehrend (mit Rückbildung)
 - progrediente (nachhaltige) Verschlechterung
- Zielgröße: Zeit bis zum Eintritt in die progrediente Phase

Sylvia Lawry Centre for Multiple Sclerosis Research

(München, gegründet 2001 unter Leitung von Prof. Neiss)

- Datenbank aus Studien der Industrie und Universitäten
- Placebo-Daten
- (ein) Ziel: Bildung virtueller Placebo-Gruppen
- Betreuung der hier vorgestellten Analyse: Prof. Fahrmeir

Anwendungsbeispiel MS

Daten von Ende 2002:

- 857 Patienten
- 26 Studien
- Studiendauer meist bis zu 2 Jahren (Maximum 5 Jahre)

Krankheits-Score:

- EDSS (Expanded Disability Status Scale)
- Skala von 0 bis 10
- Bimodale Verteilung (Maxima bei 2-3 und 6)
- Im oberen Bereich starke Gewichtung der Gehstrecke

Anwendungsbeispiel MS

Mögliche Einflussgrößen:

Alter bei Krankheitsbeginn

Alter bei Studieneintritt

Geschlecht

Krankheitsdauer bei Studieneintritt

EDSS bei Studieneintritt

Anzahl Krankheitsschübe im Jahr vor Studieneintritt

Definition der Zielgröße:

Ereignis = „Nachhaltige Verschlechterung“

Definiert über Verlaufsdaten der EDSS

Anwendungsbeispiel MS

Vorgehensweise:

- Schätzen von Splines im GAM:
 - Verschiedene Kovariablen-Kombinationen
 - Verschiedene Spline-Grade
- Selektion nach dem AIC-Kriterium
- Parametrisierung des GAM zu einem GLM, soweit durch die Form des Splines nahegelegt (hier: stückweise lineare Parametrisierung)

Vorteil dieser Parametrisierung:

- AIC kann reduziert werden, da weniger Parameter geschätzt werden müssen.
- Parametrisierte Form leichter vermittelbar, da durch eine einfache Formel beschreibbar.

Anwendungsbeispiel MS

Ergebnisse:

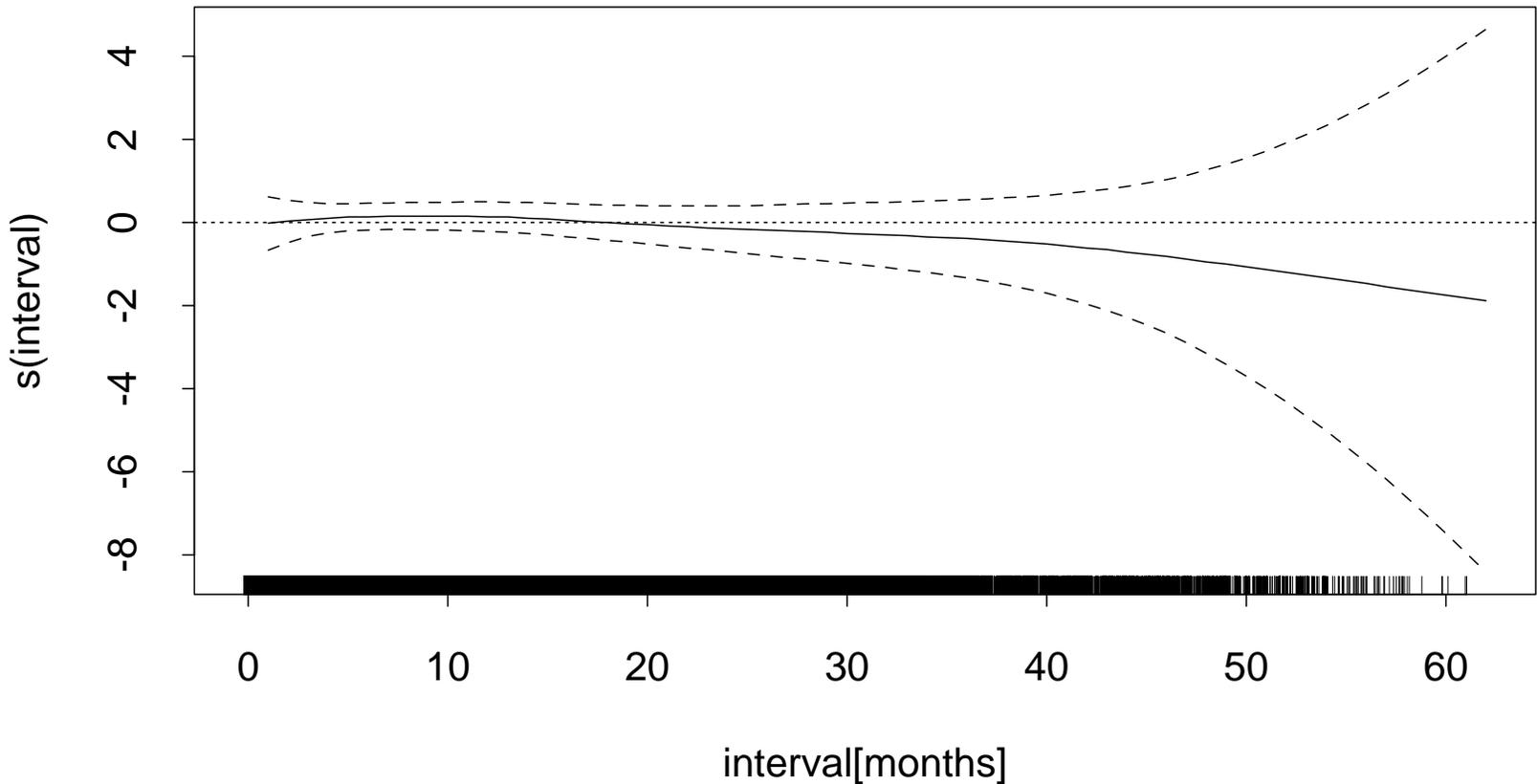
- Zunächst: Kein signifikanter Zeiteinfluss feststellbar
- Daher: Übergang zu einem reinen Exponentialmodell (d. h. keine Aufteilung in Zeitintervalle)
- Signifikante Einflüsse:

EDSS bei Studieneintritt

Anzahl der Schübe im Jahr vor Studieneintritt

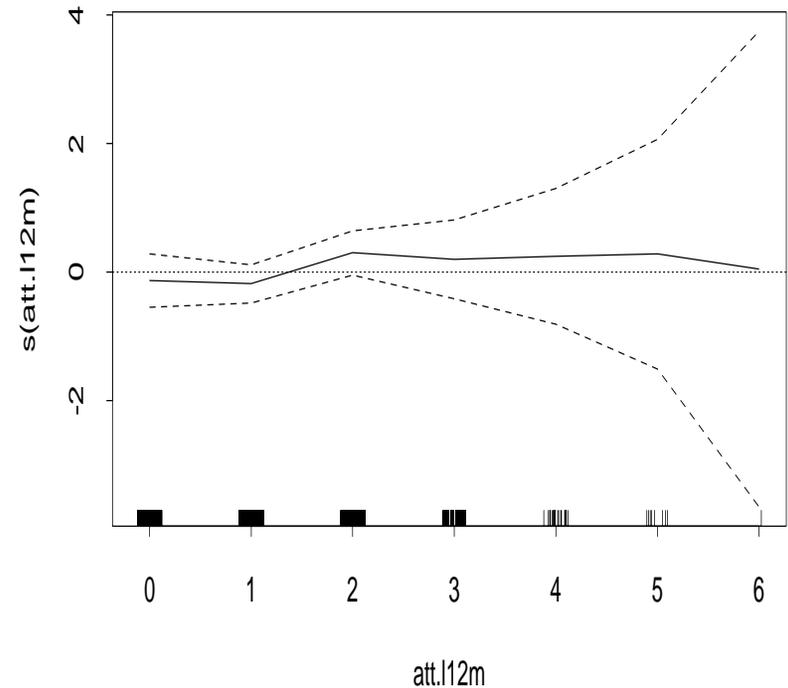
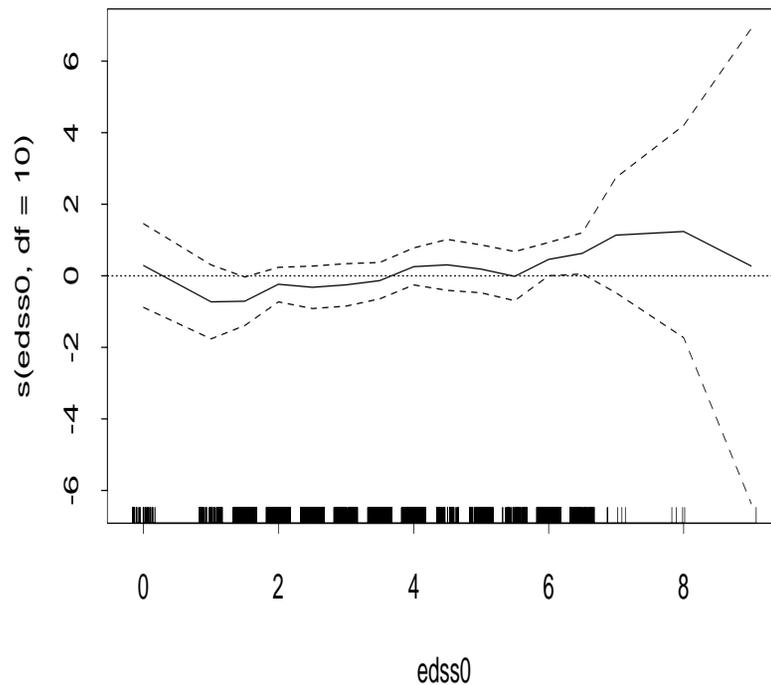
Anwendungsbeispiel MS

Spline des Zeiteinflusses:



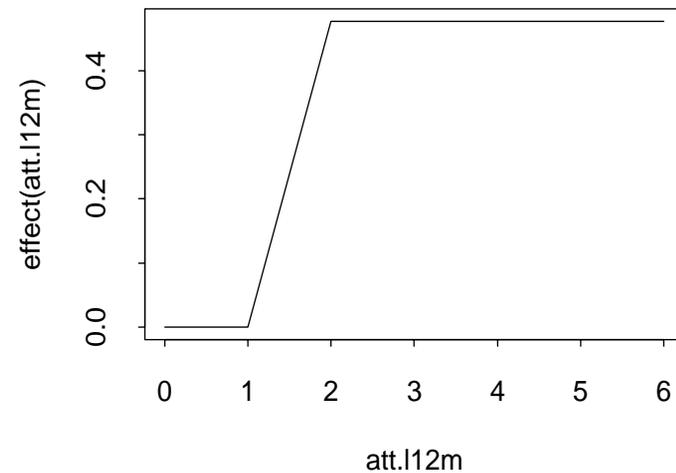
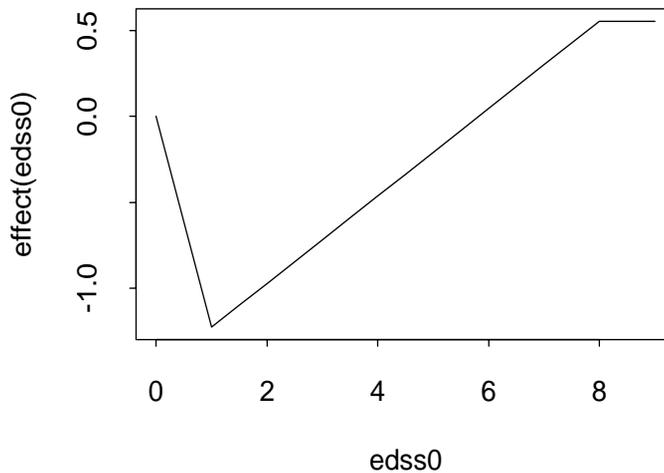
Anwendungsbeispiel MS

Splines der *EDSS* und Anzahl der Schübe vor Studieneintritt:



Anwendungsbeispiel MS

Parametrisierte Einflüsse der EDSS und *Anzahl Schübe* vor Studieneintritt:



Anwendungsbeispiel MS

Modelle mit zufälligen Effekten:

- *Patienteneffekte:*
Ebenfalls kein signifikanter Zeiteinfluss,
daher reines Exponentialmodell wie oben.
- *Studieneffekte:*
Leicht fallender Zeiteinfluss auf die Hazardrate,
sonst keine signifikanten Einflussgrößen
- Bereinigungseffekt:
Labile Patienten fallen frühzeitig aus der Beobachtung heraus.
- Für die Feststellung der tatsächlichen, längerfristigen Effekte
waren die Studien mit meist 2 Jahren zu kurz.



Fazit

- Das PEM ist ein Kompromiss aus parametrischen (Exponentialverteilung) und nichtparametrischen (stückweise) Ansätzen.
 - Direkte Schätzung der Hazardrate und Überlebenszeit-Funktion möglich.
 - Einbettung in GLM.
 - Erweiterung auf GAM bietet flexible grafische Möglichkeiten der Modellierung.
- Geeignetes Werkzeug für die explorative Analyse.

Literatur

1. Brezger A, Kneib T, Lang S. BayesX Version 0.9 Documentation. München: Institut für Statistik der Ludwig-Maximilians-Universität, 2002.
2. Fahrmeir L, Tutz G. Multivariate Statistical Modelling Based on Generalized Linear Models. 2nd ed. New York: Springer, 2001: 385-395.
3. Gehrman U, Hellriegel B, Neiss A and Fahrmeir L. Analysis of the time to sustained progression in Multiple Sclerosis using generalised linear and additive models. Sonderforschungsbereich 386 der Ludwig-Maximilians-Universität München, 2003.
4. SAS OnlineDoc Version 8. SAS Institute Inc., 2000.
5. Tutz G. Die Analyse kategorialer Daten. München: Oldenbourg, 2000: 255-256.
6. Venables WN, Ripley BD. Modern Applied Statistics with S-PLUS. 2nd ed. New York: Springer, 1997.