Ein relatives Überlebenszeitmodell für korrelierte Beobachtungen

Oliver Kuß

Institut für Medizinische Epidemiologie, Biometrie und Informatik, Medizinische Fakultät, Martin-Luther-Universität Halle-Wittenberg, Halle (Saale)

Graduiertenkolleg Dortmund, 8.7.2008

Übersicht

- Relatives Überleben
- Motivation
- Ein relatives Überlebenszeitmodell für korrelierte Beobachtungen
- Ergebnisse
- Diskussion

Relatives Überleben I

Eine zentrale Fragestellung der Epidemiologie: Messung von krankheitsspezifischer Mortalität.

Wichtig für Patienten, Kliniker und das öffentliche Gesundheitssystem.

Dafür reichen Daten aus randomisierten Studien nicht aus (selektierte Populationen), man braucht Daten aus allgemeinen Registern.

Relatives Überleben II

Drei Möglichkeiten:

- Rohe/beobachtete Überlebenszeitanalyse

  Problem: Nicht alle Patienten sind am relevanten Tumor verstorben

- Todesursachen-spezifische Überlebenszeitanalyse

  Problem: Angaben auf den Totenscheinen müssen korrekt und eindeutig sein

- **Lösung:** Relative Überlebenszeitanalyse

Relatives Überleben III: Definition

Für eine Gruppe von Patienten:

$$\text{Relatives Überleben} = \frac{\text{Beobachtetes Überleben}}{\text{Erwartetes Überleben}}$$

mit dem „erwarteten Überleben“ aus veröffentlichten alters-, geschlechts- und zeitspezifischen Sterbetafeln.

Interpretation: Das Relative Überleben beschreibt das Überleben in einer hypothetischen Bevölkerung, in der die Krankheit von Interesse die einzige Todesursache ist.

Relatives Überleben IV: Eigenschaften

Vorteile:

- Es wird keine Information zur Todesursache benötigt.
- Heilung (im statistischen Sinne) kann beschrieben werden.
- Vergleich von verschiedenen Registern/Regionen/Ländern wird erleichtert.

Relatives Überleben V: Eigenschaften

Nachteile:

- „Gute“ Information über die Sterblichkeit der Allgemeinbevölkerung ist notwendig.
- Die Registerpatienten müssen dieser Allgemeinbevölkerung entstammen.

Relatives Überleben VI: Erwartetes Überleben

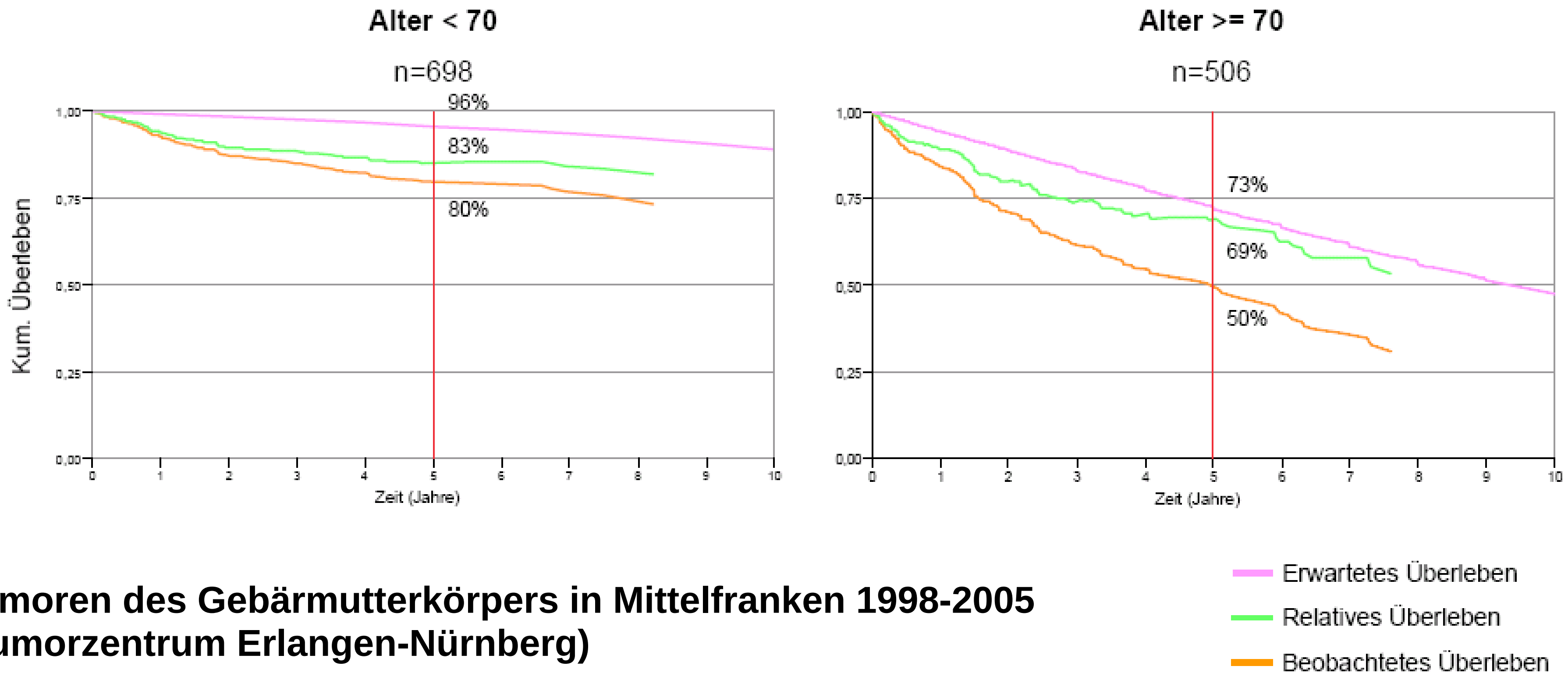
Idee: Jedem Registerpatienten wird ein „Zwilling“ gleichen Alters und gleichen Geschlechts an die Seite gestellt.

Für diese Gruppe von „Zwillingen“ wird aus den Sterbetafeln das erwartete Überleben berechnet.

Die Methoden zur Berechnung unterscheiden sich nur dadurch, wie lange der „Zwilling“ unter Risiko steht:

Ederer I: unendlich, **Ederer II:** genauso lange wie der Registerpatient, **Hakulinen:** höchstens bis zum Ende der Studie.

# Relatives Überleben nach Alter bei Diagnosestellung
n=1204



Tumoren des Gebärmutterkörpers in Mittelfranken 1998-2005
(Tumorzentrum Erlangen-Nürnberg)

Relatives Überleben VII: Regressionsmodelle

Über die reine Beschreibung des relativen Überlebens hinausgehend, sind Regressionsmodelle vorgeschlagen worden, um den Einfluss von Risikofaktoren auf das relative Überleben zu messen (Hakulinen/Tenkanen, 1987; Estève et al., 1990)

Diese sind alle *additiv* in den Hazards:

$$\lambda_{beob} = \lambda_{Pop} + \lambda_{Exzess}$$

mit λ_{beob} = beobachteter Hazard; λ_{Pop} = Populationshazard; λ_{Exzess} = Übersterblichkeits-Hazard, $\lambda_{Exzess} = \exp(X\beta)$

Vergleich mit Cox-Modell: $\lambda_{beob} = \lambda_0 \exp(X\beta)$ (*multiplikatives* Modell); λ_0 = Baseline-Hazard

Relatives Überleben VIII: Das Estève Modell als GLM

Das Estève Modell kann als GLM mit binärer Zielgröße, Poisson likelihood, einem speziellen Offset, und einer individualisierten Link-Funktion geschrieben werden (Dickmann et al., 2004).

Notation:

Gegeben $i = 1, \ldots, N$ Patienten, jeder in $j = 1, \ldots, J_i$ (i.d.R.) jährlichen Intervallen beobachtet.

δ_{ij} beschreibt das Zielereignis im ij-ten Intervall ($\delta_{ij} = 1$: Tod, $\delta_{ij} = 0$: Überleben).

r_{ij} ist die Zeit unter Risiko (in %) und $e^*_{ij} = (\lambda_{Pop} * r_{ij})$ der gewichtete Populationshazard im ij-ten Intervall.

Relatives Überleben IX: Das Estève Modell als GLM II

Modellgleichung:

$$\ln(\mu_{ij} - e^*_{ij}) = \ln(r_{ij}) + x_i\beta.$$

- Proportionale Hazards für die Risikofaktoren.
- Konstante Hazards in den ij-Intervallen.
- Durch die J_i Beobachtungen pro Patient wird keine Korrelation induziert.

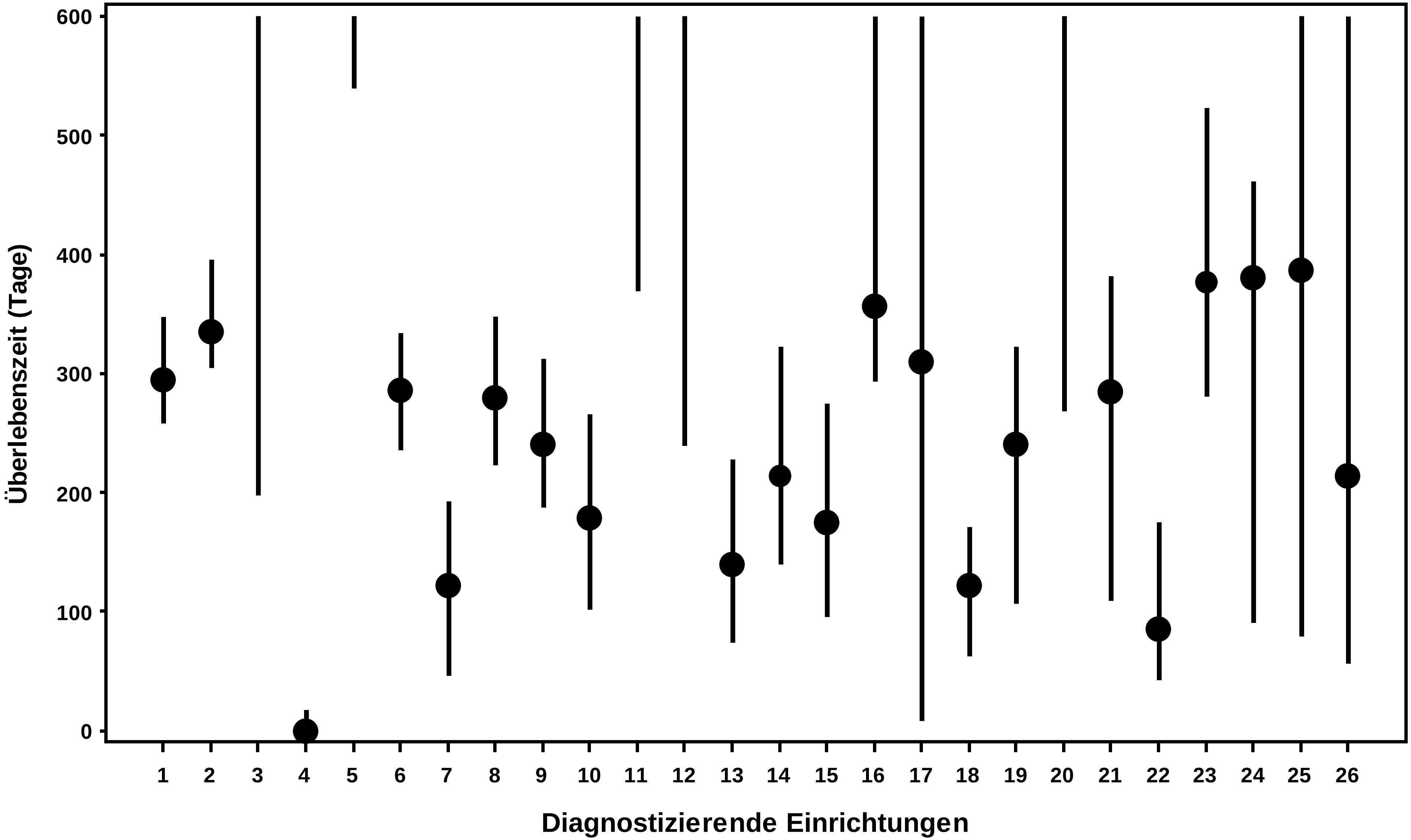
Motivation I: Die HALLUCA-Studie

HALLUCA-(= Halle Lung Carcinoma)-Studie: Epidemiologische Studie zur Versorgung von Lungenkrebspatienten im südliche Sachsen-Anhalt.

Standardisierte Erhebung von allen von Lungenkrebspatienten von 4/1996 bis 9/1999, Follow-up bis 9/2000.

N=1696 Lungenkrebspatienten, 1349 Patienten (79.5%) waren bis zum Ende des Follow-Up verstorben, mediane Ü-zeit 284 Tage (= 9,3 Monate).

Sterbetafeln vom Statistischen Landesamt Sachsen-Anhalt.


Überlebenszeit (Tage)
600
500
400
300
200
100
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
Diagnostizierende Einrichtungen

Ein relatives Überlebenszeitmodell für korrelierte Beobachtungen I

Um für die beobachtete Heteogenität zu adjustieren, verallgemeinere Dickmans Modell durch Einfügung eines zufälligen Effektes für die diagnostizierenden Einrichtungen in den linearen Prädiktor (und erhalte ein GLMM).

Konkret, mit δ_{hij} als Indikator für das Zielereignis in Einrichtung h ($h = 1, \ldots, H$), modelliere

$$\ln(\mu_{hij} - e^*_{ij}) = \ln(r_{ij}) + x_i\beta + u_h$$

Der zufällige Intercept u_h wird als normal-verteilt mit Varianz σ_h^2, $u_h \sim N(0, \sigma_h^2)$ angenommen.

Ein relatives Überlebenszeitmodell für korrelierte Beobachtungen II

Parameterschätzung in diesem Modell (wie in allen echten GLMM) ist kompliziert , da die Likelihood-Funktion aus H nicht geschlossenlösbaren Integralen besteht.

Benutze nummerische (SAS PROC NLMIXED) oder stochastische Integration (WinBUGS) zur Parameterschätzung.

Zusätzliche Komplikation: Individualisierte Link-Funktion

Ein relatives Überlebenszeitmodell für korrelierte Beobachtungen III: SAS PROC NLMIXED

```
proc nlmixed data=... ;
     parms int=-1 b_stage2=0.5 b_stage3=0.7 ...  sd2=1;

     Xbeta = int + b_stage2*stage2 + b_stage3*stage3 + ... + u_h;

     Mu    = exp(Xbeta+log_r_ij) + e_ij;

     loglike = delta_ij*log(Mu) - Mu;
     model delta_ij ~ general(loglike);
     random u ~ normal(0,sd2_h) subject=DiagnosticUnit;
run;
```

Ein relatives Überlebenszeitmodell für korrelierte Beobachtungen IV: WinBUGS

```
model; {
    for (i in 1:N){
        Xbeta[i] <- int + b_stage2*stage2[i] + b_stage3*stage3[i] + ...
                        + u_zentrum[zentrum[i]];

        logrisk[i]<- log(risk[i]);
        log(mue[i]) <- logrisk[i] + Xbeta[i]+ exp(e[i]);
        survival_status[i] ~ dpois(mue[i]);
                   }
    for (h in 1:H){
         u_zentrum[h]~ dnorm( 0.0000, tau_zentrum);
                  }
    tau_zentrum ~ dgamma(0.001,0.001);
    var_zentrum <- 1 / tau_zentrum;

   # priors
     int~ dnorm( 0.0,1.0E-6) b_stage2~ dnorm( 0.0,1.0E-6) ...
         }
```

Ergebnisse I: Feste Effekte (Auswahl)

Kovariate	Kategorie	Estève Modell β (SE)	PROC NLMIXED β (SE)
Geschlecht	Weiblich	-0,139 (0,074)	-0,145 (0,076)
Alter	$57- < 63$ Jahre	0,029 (0,088)	0,025 (0,091)
	$63- < 67$ Jahre	0,123 (0,092)	0,123 (0,095)
	$67- < 73$ Jahre	0,171 (0,088)	0,161 (0,090)
	$>=$ 73 Jahre	0,268 (0,093)	0,279 (0,096)
Histologischer	SCLC	0,128 (0,069)	0,123 (0,071)
Typ	Fehlend	-0,139 (0,116)	-0,126 (0,119)

Ergebnisse II: Zufällige Effekte

Parameter	Estève Modell	PROC NLMIXED
σ_h^2	–	0,053 (0,037)
-2LogL	6725,4	6721,7
BIC	6860,4	6846,0

p(LR-test: $\sigma_h^2 = 0$, $\chi^2_{obs} = 2.888$): 0.045

Diskussion I

- Ein relatives Überlebenszeitmodell für korrelierte Beobachtungen kann als Einbettung des Estève-Modells in die Klasse der GLMM leicht definiert werden.

- In unserem Datensatz waren Schätzer für die festen Effekte im Estève-Modell (ohne Berücksichtigung der Korrelation) und im Modell mit zufälligen Effekten ähnlich, auch wenn das neue Modell das bessere Modell war.

- Eine Reihe von Erweiterungen ist denkbar: Negativ-binomiale Likelihood, mehr hierarchische Ebenen, Kovariaten mit zufälligen Effekten, nicht-proportionale Hazards, nicht-normal-verteilte zufällige Effekte ...

Diskussion II

- Zur Parameterschätzung kann SAS PROC NLMIXED (oder WinBUGS, neben anderen) herangezogen werden.

- Vorteile PROC NLMIXED: Datenmanagement, Laufzeit

- Vorteile WinBUGS: Erlaubt mehrere hierarchische Ebenen

- Modelle in mehreren Paketen hinzuschreiben, ist immer eine gute Idee.

References

1. Dickman PW, Sloggett A, Hills M, Hakulinen T. Regression Models for Relative Survival. Stat Med 2004; 23:51-64.

2. Estève J, Benhamou E, Croasdale M, Raymond L. Relative survival and the estimation of net survival: Elements for further discussion. Stat Med 1990; 9:529-538.

3. Hakulinen T, Tenkanen L. Regression analysis of relative survival rates. Appl Stat 1987; 36:309-317.

4. Kuss O, Blankenburg T, Haerting J. A relative survival model for clustered responses. Biom J 2008; 50:408-418.