



LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

INSTITUT FÜR STATISTIK  
SONDERFORSCHUNGSBEREICH 386



Aydemir, Aydemir, Dirschedl:

## Survivalanalysen mit Berücksichtigung der zeitlichen Kovariablenentwicklung in klinischen Studien

Sonderforschungsbereich 386, Paper 44 (1996)

Online unter: <http://epub.ub.uni-muenchen.de/>

Projektpartner



# Survivalanalysen mit Berücksichtigung der zeitlichen Kovariablenentwicklung in klinischen Studien

Sibel Aydemir<sup>1</sup>, Ülker Aydemir<sup>2</sup>, Peter Dirschedl<sup>1</sup>

## Abstract

Data from the Stanford Heart Transplantation Study and our own study on brain tumor include time-dependent covariates like transplantation, which may switch only once, and others changing their value several times during follow-up. But classical analyses never used this additional information.

In a comparative study we applied the time-dependent Cox model, pooled Cox regression and the linear counting process by Aalen to these data sets.

All methods do show similar results when they are carried out in their 'fixed' version, i.e. using baseline information only, or when covariates are being treated as time-dependent.

But the estimated effects do differ remarkably between fixed and time-dependent approaches, thus leading to different interpretations of risks.

---

<sup>1</sup> Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie der Ludwig-Maximilians-Universität München

<sup>2</sup> Münchner Forschungsverbund Öffentliche Gesundheit-Public Health, Ludwig-Maximilians-Universität München

# 1 Einführung

In der multiplen Überlebenszeitanalyse hat sich das von D. Cox entwickelte semiparametrische Regressionsmodell, trotz der kritisierten restriktiven Annahme der "proportional hazards", weitgehend durchgesetzt. Im sogenannten Cox-Proportional-Hazards-Regressionsmodell wird die Hazardfunktion in Abhängigkeit von den potentiellen Einflußgrößen modelliert, die zum Baselinezeitpunkt  $t=0$  (meist Untersuchung zur Aufnahme in eine Studie) ermittelt werden.

In klinischen Studien werden gewöhnlich in festgelegten Abständen follow-up Untersuchungen durchgeführt, die die Entwicklung relevanter Kovariablen über die Zeit dokumentieren. Von diesen Kovariablen liegen folglich nicht nur Baselinewerte, sondern auch Informationen im Verlauf vor. Daneben gibt es Kovariablen, die zum Zeitpunkt der Baseline überhaupt nicht vorliegen bzw. vorliegen können und erst im Laufe der Studienzeit definiert werden, wie zum Beispiel die Durchführung einer Transplantation oder das Auftreten von Blutungen. Kovariablen, die eine derartige zeitliche Struktur aufweisen, werden als zeitveränderliche Kovariablen bezeichnet. Standardanalysen der Überlebenszeit vernachlässigen die zeitliche Struktur der Kovariablen gänzlich, obwohl die Beachtung der Zeitveränderlichkeit zu wichtigen neuen Erkenntnissen führen kann.

Das Teilprojekt B3 des Sonderforschungsbereiches 386 beschäftigt sich mit Survivalmodellen, die die zeitliche Struktur von Kovariablen mitberücksichtigen. Zunächst wird in Kapitel 2 der Begriff der Zeitveränderlichkeit von Kovariablen erläutert, insbesondere werden hier die Bedeutung zeitveränderlicher Kovariablen in klinischen Studien und der Erkenntnisgewinn gegenüber einer gewöhnlichen Baselineanalyse aufgezeigt. Das dritte Kapitel beschreibt anschließend drei verschiedene Ansätze zur Überlebenszeitanalyse unter Berücksichtigung zeitveränderlicher Kovariablen, nämlich das time-dependent Cox-Modell, einen gepoolten Cox-Regressionsansatz und das lineare Modell von Aalen, das bereits in Discussion Paper 25 des SFB 386 (S. Aydemir, Ü. Aydemir, P. Dirschedl, 1996) ausführlich behandelt wurde. Während im time-dependent Cox-

Modell und im gepoolten Cox-Ansatz konstante Effekte für die Kovariablen angenommen werden, ermöglicht das lineare Modell von Aalen die zeitliche Entwicklung der Effekte mitzuschätzen. Die Gegenüberstellung und Diskussion der vorgestellten zeitveränderlichen Ansätze sowie der Vergleich zu fixen Baselineanalysen erfolgt in Kapitel 4 am Beispiel der Stanford Heart Transplantation Data (Crowley & Hu, 1977, Kalbfleisch & Prentice, 1980, Arjas, 1986) und an den Daten einer Studie zur Therapie hochmaligner Hirntumoren.

## **2 Zeitveränderliche Kovariablen - Definition und ihre Bedeutung in klinischen Studien**

Kovariablen, die zu Beginn einer Studie bzw. zum sogenannten Baselinezeitpunkt  $t=0$  ermittelt werden und im weiteren Verlauf der Beobachtung unverändert bleiben, werden als fixe Kovariablen bezeichnet. Derartige fixe Kovariablen sind beispielsweise das Geschlecht, die Tumorlokalisierung oder die Operationsart. Im Gegensatz zu fixen Kovariablen spricht man von zeitveränderlichen Kovariablen, falls entweder Kovariablenwerte sich im Beobachtungsverlauf ändern (updated covariates) oder falls eine Kovariable zum Baselinezeitpunkt noch nicht existent ist und erst im Laufe der Studie definiert wird. Eine Klassifikation zeitveränderlicher Kovariablen ist in Tab.1 zu finden.

Das Paradebeispiel einer Kovariable, die zum Baselinezeitpunkt noch nicht definiert ist, ist die Transplantation. Es handelt sich hierbei um den einfachsten Fall der Zeitveränderlichkeit mit einer einmaligen Zustandsänderung im Zeitverlauf. Da für eine adäquate Analyse berücksichtigt werden muß, daß zum einen Patienten mit schlechtem Gesundheitszustand die Wartezeit bis zur Transplantation häufig nicht überleben und zum anderen bei transplantierten Patienten der Transplantationszeitpunkt variabel ist, führt auch diese einfache zeitveränderliche Kovariable noch

zu Problemen. Das Ignorieren der Zeitveränderlichkeit der Transplantation in den Analysen, derart, daß die Transplantationsvariable als fix betrachtet wird und somit der Zustand der Transplantation auf den Baselinezeitpunkt zurückprojiziert wird, würde zweifellos zu falschen Aussagen führen, da in der Gruppe der Transplantierten frühe Todesfälle nicht enthalten sein können. Weitere Beispiele für Kovariable, die erst nach Studienbeginn definiert werden und sich im Laufe der Beobachtungszeit ändern können, sind in klinischen Studien die Dosis bzw. die Dosisintensität.

Eine zweite Kategorie der Zeitveränderlichkeit bilden Kovariable, deren Werte zu Beginn einer Studie ermittelt werden, sich aber im Laufe der Studienzeit ändern können (*updated covariates*). Mehrfache Zustandsänderungen im Laufe der Studienzeit liegen beispielsweise bei wiederholtem Auftreten epileptischer Anfälle oder bei der wiederholten Messung der Veränderung des klinischen Status vor. Desweiteren wird eine metrische Kovariablenentwicklung bei wiederholter Messung von metrischen Kovariablen, wie z. B. Laborwerten, beobachtet.

Wie unschwer zu erkennen ist, sind zeitveränderliche Kovariable in klinischen Studien weitaus häufiger anzutreffen als fixe Kovariable. Die zeitliche Folge der Kovariablenwerte ist in vielen Studien durch follow-up Untersuchungen gut dokumentiert. Trotz der verfügbaren Information bezüglich der Kovariablenentwicklung wird -vor allem im Falle von *updated covariates*- die Analyse der Überlebenszeit vorwiegend mit dem üblichen Cox-proportional-hazards-Regressionsmodell vorgenommen, in das die Kovariablen eben nur mit ihren zum Baselinezeitpunkt ermittelten Werten eingehen. In diesem Ansatz werden alle Kovariablen als fix betrachtet, so daß die zeitliche Entwicklung der Kovariablen, also ein Großteil der verfügbaren Information, unberücksichtigt bleibt. Baselineanalysen haben konstruktionsgemäß zum Ziel, den Einfluß von Kovariablenausprägungen, die zu Studienbeginn vorliegen, auf ein späteres Ereignis wie z.B. Tod oder Rezidiv zu ermitteln. Oft beträgt die Zeitspanne zwischen dem Baselinezeitpunkt und dem Ereigniszeitpunkt in klinischen Studien aber viele Jahre, so daß sich hieraus unmittelbar die Frage nach der Bedeutung und der Aussagekraft einer Überlebenszeitanalyse mit Baselinewerten der Kovariablen stellt. Liegt der Baselinezeit-

punkt weit entfernt vom Ereigniszeitpunkt, so findet in Baselineanalysen ein sogenannter "zeitlicher Verfall" der Kovariableneffekte statt (Altman et al., 1994).

**Tab.1:** Typen von zeitveränderlichen Kovariablen

<b>Zeitveränderliche Kovariable</b>	
<p>Variablen, die zum Baselinezeitpunkt noch nicht definiert sind, im Laufe der Beobachtungszeit belegt werden und sich im weiteren Verlauf auch ändern können</p>	<p>Variablen, die zum Baselinezeitpunkt vorliegen und sich im Laufe der Beobachtungszeit ändern (updated covariates)</p>
<p><u>Beispiele :</u></p> <ul style="list-style-type: none"> <li>• Transplantation (einmaliger Wechsel)</li> <li>• Dosis, Dosisintensität (metrisch im Verlauf)</li> </ul>	<p><u>Beispiele:</u></p> <ul style="list-style-type: none"> <li>• Therapiewechsel (ein- oder mehrmaliger Wechsel)</li> <li>• Veränderung d. klinischen Status, TNM-Status (wiederholte Messung von nominalen und ordinalen Merkmalen)</li> <li>• Laborwerte (wiederholte Messung von metrischen Kovariablen)</li> </ul>

### **3 Survivalmodelle unter Berücksichtigung von zeitveränderlichen Kovariablen**

Es ist sinnvoll, sich zunächst Gedanken über die Notwendigkeit der Berücksichtigung veränderlicher Kovariablenstrukturen zu machen. So ist die Beachtung der zeitlichen Entwicklung von Kovariablen gewiß nicht immer notwendig und in manchen Fällen ist trotz Vorliegen zeitveränderlicher Risikofaktoren eine Beschränkung der Analyse auf die Baselinewerte der Kovariablen sogar vorzuziehen. Beispielsweise sind in Therapiestudien, in denen primär die Ermittlung des Effektes der Therapie von Interesse ist, Analysen mit Baselinewerten der Risikofaktoren und der Therapie als Einflußgrößen durchaus ausreichend. Die Effekte der Risikofaktoren werden in diesem Fall nur im Hinblick auf eine unverzerrte (adjustierte) Schätzung des Therapieeffektes mitgeschätzt. Ein Einbezug der zeitlichen Struktur der Risikofaktoren würde hier den Wirkmechanismus der Therapie absorbieren, so daß ein auffälliger Therapieeffekt möglicherweise nicht erkannt werden würde. Falls jedoch nicht der Therapieeffekt, sondern die Schätzung der Effekte der Kovariablen selbst im Vordergrund stehen, sollte die zeitliche Entwicklung von Kovariablen in Überlebenszeitanalysen nicht unberücksichtigt bleiben. Die Schätzung der Kovariableneffekte auf der Basis der zeitlichen Kovariablenstruktur und die daraus resultierenden Aussagen bezüglich der Risikofaktoren sind vor allem in Studien, in denen kein auffälliger Therapieeffekt vorhanden ist, von großer Bedeutung. Denn fundierte Kenntnisse über den Einfluß und die Entwicklung prognostischer Faktoren können entscheidend zur Konzeption und Prüfung neuer Therapien beitragen.

Der Erkenntnisgewinn durch Einbezug der vorliegenden Information bezüglich der Zeitveränderlichkeit sollte allerdings immer in Relation zu dem Aufwand gesehen werden, den Analysen mit zeitveränderlichen Kovariablen, allein schon durch die komplexe Aufbereitung der Daten, verursachen. Die in der Praxis recht seltene Berücksichtigung der zeitlichen Strukturen von Kovariablen in Überlebenszeitanalysen ist neben dem aufwendigen Datenhandling vor allem auf das begrenzte Angebot an entsprechender Analysesoftware zurückzuführen.

Im folgenden werden drei unterschiedliche Ansätze zur Berücksichtigung der Zeitveränderlichkeit bei der Analyse der Überlebenszeit vorgestellt.

Der bekannteste Ansatz in diesem Rahmen ist das auf die Situation von zeitveränderlichen Kovariablen erweiterte Cox-Regressionsmodell, das sogenannte time-dependent Cox-Modell.

Eine weitere Möglichkeit die zeitliche Folge von Kovariablen einzubeziehen, bietet das Poolen der Daten. Ein Vorteil des gepoolten Verfahrens ist, daß die Analyse des gepoolten Datensatzes letztendlich mit dem bekannten, fixen Cox-proportional-hazards-Regressionsmodell durchgeführt werden kann. Beide Modelle berücksichtigen die zeitliche Entwicklung von Kovariablen, die Effekte der Kovariablen werden allerdings über den gesamten Beobachtungszeitraum hinweg konstant geschätzt.

Im Unterschied dazu ermöglicht der dritte in dieser Arbeit vorgestellte Ansatz, der Zählprozeßansatz von Aalen, sowohl den Einbezug von zeitveränderlichen Kovariablen, als auch die Schätzung der zeitlichen Variation der Kovariableneffekte.

### 3.1 Das time-dependent Cox-Regressionsmodell

Sei im folgenden  $h_i(t)$ ,  $i=1,\dots,n$ , die Hazardfunktion des  $i$ -ten Patienten zum Zeitpunkt  $t$ ,  $h_0(t)$  die unspezifizierte Baselinehazardfunktion und  $x_{ij}(t)$  der Wert der  $j$ -ten Kovariable ( $j=1,\dots,p$ ) des  $i$ -ten Patienten zum Zeitpunkt  $t$ . Entfällt in Verbindung mit den Kovariablen der zeitliche Zusatz  $t$ , so entspricht  $x_{ij}$  dem Wert der  $j$ -ten Kovariable des  $i$ -ten Patienten, gemessen zum Baselinezeitpunkt  $t=0$ .

Analog zum gewöhnlichen Cox-proportional-hazards-Regressionsmodell

$$h_i(t) = h_0(t) \exp \left[ \sum_{j=1}^p \beta_j x_{ij} \right] \quad (1)$$

erfolgt die Modellierung der Hazardrate des  $i$ -ten Individuums im time-dependent Ansatz als

$$h_i(t) = h_0(t) \exp \left[ \sum_{j=1}^p \beta_j x_{ij}(t) \right], \quad (2)$$



wobei die Unterschiede zwischen den beiden Ansätzen (1) und (2) zum einen in der berücksichtigten Kovariableninformation und zum anderen in der Interpretation der Kovariableneffekte liegen. Im gewöhnlichen Cox-Regressionsmodell (1), in das lediglich die zum Baselinezeitpunkt  $t=0$  ermittelten Ausprägungen der Kovariablen eingehen, impliziert der Regressionskoeffizient  $\beta_j$  den Effekt der zum Baselinezeitpunkt gemessenen  $j$ -ten Kovariable auf die Hazardrate. Demgegenüber werden im time-dependent Ansatz (2) jeweils die aktuellen Werte der Kovariablen zu den Ereigniszeitpunkten  $T_i$  betrachtet. Der Regressionskoeffizient  $\beta_j, j=1, \dots, p$ , stellt im zeitabhängigen Ansatz den Effekt der  $j$ -ten Kovariable zum Baselinezeitpunkt und zu allen späteren Zeitpunkten dar, d.h. es wird ein konstanter Effekt der Kovariablen über den gesamten Beobachtungszeitraum hinweg angenommen.

#### *Der Partial-Likelihood-Ansatz*

Sei  $\delta_i$  eine Indikatorvariable, die den Wert 1 annimmt, falls  $T_i$  eine Ereigniszeit ist, und 0, falls es sich um eine zensierte Beobachtungszeit handelt. Bezeichne ferner  $R_i$  (risk set) alle Individuen, die zum Zeitpunkt  $T_i$  unter Risiko stehen, d.h. bis dahin noch kein Ereignis erfahren haben. Unter der Annahme, daß keine Bindungen in den Ereigniszeiten vorliegen, erfolgt die Schätzung der unbekanntten Kovariableneffekte  $\beta_j, j=1, \dots, p$ , durch Maximierung der partiellen Log-Likelihood  $LL^{part}$

$$LL^{part} = \sum_{i=1}^n \delta_i \log p(T_i) \quad (3)$$

mit

$$p(T_i) = \left\{ \frac{\exp\left(\sum_{j=1}^p \beta_j x_{ij}(T_i)\right)}{\sum_{k \in R_i} \exp\left(\sum_{j=1}^p \beta_j x_{kj}(T_i)\right)} \right\}. \quad (4)$$

### *Fehlende Kovariablenwerte und Imputationsmethoden*

Das time-dependent Cox-Modell setzt voraus, daß Kovariablenwerte zu allen Ereigniszeitpunkten vorliegen. In praxi ist diese Voraussetzung natürlich nicht immer erfüllt, so daß in manchen Fällen ein adäquates Ersetzen fehlender Werte erforderlich ist. Fehlen Kovariablenwerte im Verlauf, wird gewöhnlich die letzte verfügbare Kovariableninformation fortgeschrieben. Handelt es sich um stetige Kovariablen, können fehlende Werte auch durch lineare Interpolation zwischen dem unmittelbar vorangegangenen und dem direkt nachfolgenden Meßwert der Kovariable ersetzt werden. Größere Schwierigkeiten bereiten fehlende Baselinewerte der Kovariablen. In der Regel werden Individuen mit fehlenden Angaben zum Baselinezeitpunkt aus der Analyse ausgeschlossen (Altman et al., 1994). Falls durch dieses Vorgehen jedoch zu viele Beobachtungen eliminiert werden, empfiehlt es sich, fehlende Baselinewerte beispielsweise durch Stichprobenmittelwerte bzw. Stichprobenmediane der entsprechenden Kovariablen zu ersetzen.

### *Proportionalität der Hazards*

Das zeitveränderliche Cox-Modell geht von konstanten Effekten der Kovariablen im Zeitverlauf aus, im Unterschied zum Cox-Proportional-Hazards-Regressionsmodell mit den Baselinewerten der Kovariablen entfällt hier jedoch die Annahme der Proportionalität der Hazards. Da im zeitveränderlichen Modell die zeitliche Variation der Hazards nicht allein durch die Zeitabhängigkeit der Baselinehazard, sondern auch durch die Veränderlichkeit der Kovariablen verursacht wird, ist es nicht mehr möglich, ein konstantes Hazardverhältnis über die gesamte Beobachtungszeit hinweg voranzusetzen.

### *Auswirkungen der zeitlichen Kovariableninformation auf den Kovariableneffekt*

Wird der Verlauf von "updated covariates" ignoriert, fallen die Effektschätzungen der Baseline-Cox-Analyse erfahrungsgemäß betragsmäßig

schwächer aus als die entsprechenden Kovariableneffekte bei Berücksichtigung der Zeitabhängigkeit im time-dependent Ansatz. Der Grund hierfür liegt in der oftmals langen Zeitspanne zwischen dem Baselinezeitpunkt und dem Ereignis. Je weiter nämlich Baselinewerte vom Ereignis zurückliegen, umso geringer wird ihr Einfluß auf das Überleben. Altman et al. (1994) bezeichnen dieses Phänomen als "zeitlichen Verfall" der Effekte von Baselinewerten ("...time decay of the effect of entry values"). Diese Aussage ist jedoch nicht übertragbar auf die zweite Kategorie zeitveränderlicher Kovariablen, nämlich Kovariablen, die zum Baselinezeitpunkt noch nicht existent sind. Die Durchführung einer Baselineanalyse ist in diesem Fall methodisch falsch, und die Effektschätzungen sind demnach verzerrt (vgl. Seite 4).

### **3.2 Die gepoolte Cox-PH-Regression**

Eine weitere Möglichkeit zur Berücksichtigung der zeitlichen Folge von Kovariablen besteht im Poolen von Beobachtungen (Hughes et al., 1992). Bei diesem Ansatz wird die gesamte Beobachtungszeit in Intervalle unterteilt, wobei für jedes Intervall jeweils der Meßwert der interessierenden Kovariable zu Beginn des Intervalls und der Status des Patienten am Ende des Intervalls festgehalten werden. Die auf diese Weise erhaltenen einzelnen Intervalle der Patienten werden gepoolt, so daß jedes Einzelintervall nun einer neuen Beobachtung entspricht. Die Untersuchungseinheiten der gepoolten Analyse sind somit nicht mehr die ursprünglichen Patienten, sondern die Einzelintervalle der Patienten. Die gepoolten Daten werden anschließend mit dem fixen Cox-proportional-hazards-Regressionsmodell analysiert.

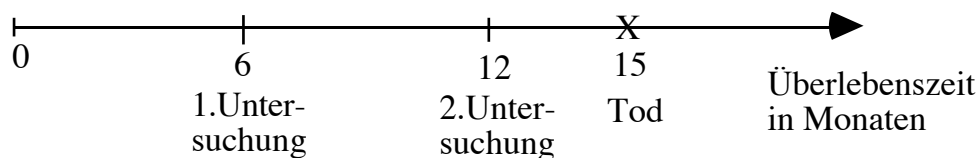
#### *Annahmen des gepoolten Cox-Ansatzes*

Bei diesem gepoolten Verfahren wird aus den ursprünglichen Patientendaten ein bezüglich der Beobachtungseinheiten aufgeblähter und bezüglich

der Beobachtungszeit verkürzter Datenkörper erzeugt. Durch Bildung von Einzelintervallen geht die Historie der Patienten vollständig verloren, so daß im nachhinein eine Zuordnung der Intervalle zu den Patienten nicht mehr möglich ist. Folglich kann nach dem Poolen der Intervalle auch nicht mehr nachvollzogen werden, aus welcher Krankheitsphase eines Patienten die Intervalle stammen, ob aus der anfänglichen oder einer späteren, fortgeschrittenen Krankheitsphase. Die Annahmen des gepoolten Cox-Regressionsmodells sind somit besonders restriktiv. Neben der Annahme, daß die Vergangenheit eines Patienten keinen Einfluß auf seine Prognose besitzt, werden für alle Intervalle sowohl identische Baselinehazardraten als auch identische Risikoverhältnisse, d.h. konstante Effekte über die Zeit, vorausgesetzt. Alle Intervalle werden desweiteren gleichermaßen gewichtet, ohne Beachtung dessen, aus welchen Krankheitsphasen die Intervalle stammen.

#### *Vorgehensweise beim Poolen der Daten*

Das Vorgehen der gepoolten Cox-Regression wird an einem fiktiven Beispiel kurz demonstriert. Angenommen, die Patienten einer Studie werden gemäß dem Studienprotokoll nach einer Eingangsuntersuchung regelmäßig alle 6 Monate nachuntersucht. Die Überlebenszeit eines beliebigen Patienten betrage 15 Monate.



Dieser betrachtete Patient mit der Überlebenszeit von 15 Monaten ist im gepoolten Datenkörper mit folgenden drei Intervallen vertreten, die jeweils die aktuelle Kovariableninformation enthalten. Das erste Intervall

geht in die gepoolten Daten mit den bei der Eingangsuntersuchung ( $t=0$ ) ermittelten Kovariablenwerten und einer zensierten Überlebenszeit von 6 Monaten ein. Durch das zweite Intervall werden die Kovariablenmeßwerte der ersten Nachuntersuchung und ebenfalls eine zensierte Überlebenszeit von 6 Monaten eingebracht. Das dritte Intervall schließlich, in dem der Patient nach 3 Monaten verstirbt, enthält die bei der zweiten Nachuntersuchung ermittelten, aktuellen Kovariablenwerte. Anstelle einer Überlebenszeit von 15 Monaten werden somit zwei zensierte jeweils 6 monatige Beobachtungszeiten und eine Ereigniszeit von 3 Monaten berücksichtigt. Die gepoolte Analyse setzt eine hohe Ereignisrate in den ursprünglichen Daten voraus, da infolge des Poolens der Zeitintervalle die Ereignisrate entsprechend reduziert wird.

In klinischen Studien ist die Einteilung der Beobachtungszeit in Intervalle in der Regel nicht frei wählbar, sondern meist durch den im Studienprotokoll bereits festgelegten Untersuchungsablauf vorgegeben. Auch sind in klinischen Studien einheitliche Intervalllängen kaum realisierbar, da mit Fortschreiten der Beobachtungszeit eines Patienten Untersuchungen oftmals seltener durchgeführt werden.

Die Idee des Poolens zur Berücksichtigung der zeitlichen Struktur von Kovariablen ist eng verwandt mit der Methode des Episodensplittings, die vor allem in sozio-ökonomischen Paneluntersuchungen Anwendung findet.

### 3.3 Der lineare Zählprozeßansatz von Aalen

Ein weiteres Verfahren, das die Berücksichtigung der zeitlichen Entwicklung von Kovariablen erlaubt, ist das nichtparametrische lineare Modell von Aalen, das bereits in Discussion Paper Nr. 25 des SFB 386 ausführlich behandelt wurde (s.a. Aalen, 1989, Andersen et al., 1993, Mau, 1986 u. 1988). Im Unterschied zum time-dependent Cox-Regressionsmodell und dem gepoolten Cox-Modell gestattet dieser Ansatz zudem auch die Möglichkeit die zeitliche Struktur der Kovariableneffekte mitzuschätzen.

Bezeichne im folgenden  $\lambda_i(t)$  die Intensität für das Auftreten eines Ereignisses zum Zeitpunkt  $t$  des  $i$ -ten Individuums, d.h.  $\lambda_i(t)$  ist die Wahrscheinlichkeit dafür, daß ein Ereignis im Intervall  $t+dt$  eintritt, unter der Bedingung, daß zuvor noch kein Ereignis eingetreten ist.  $\lambda(t)$  sei der  $n$ -dimensionale Intensitätsvektor der individuellen Intensitäten (=hazards)  $\lambda_i(t)$ ,  $i=1, \dots, n$ .

Das lineare Modell für die Intensität  $\lambda(t)$  lautet

$$\lambda(t) = Y(t)\alpha(t), \quad (5)$$

Dabei ist  $Y(t)$  die  $n \times (r+1)$ -Datenmatrix der Kovariablen  $Z_j$ ,  $j = 1, \dots, r$ , gemessen zum Zeitpunkt  $t$ , erweitert um den Intercept 1. Der Vektor  $\alpha(t) = (\alpha_0(t), \alpha_1(t), \dots, \alpha_r(t))'$  enthält die Regressionsfunktionen, wobei  $\alpha_j(t)$ ,  $j = 1, \dots, r$ , den Einfluß der  $j$ -ten Kovariable zur Zeit  $t$  wiedergibt und  $\alpha_0(t)$  die Baselineparameterfunktion bezeichnet.

Der lineare Ansatz (5) ermöglicht somit sowohl den Einbezug zeitveränderlicher Kovariablen  $Y(t)$  als auch die zeitliche Modellierung der Regressionsfunktionen  $\alpha(t)$ . Zu jedem Zeitpunkt  $t$  ist in der Datenmatrix  $Y(t)$  die aktuelle Kovariableninformation enthalten. Ist das  $i$ -te Individuum zum Zeitpunkt  $t$  nicht mehr unter Risiko, so wird die entsprechende  $i$ -te Zeile der Matrix  $Y(t)$  mit Nullen besetzt. Liegen lediglich Kovariablen mit ihren zum Baselinezeitpunkt  $t=0$  ermittelten Meßwerten vor, ändert sich die Belegung der Kovariablenmatrix nicht, solange der Patient unter Risiko ist.

Im Zusammenhang mit der zeitlichen Struktur von  $Y(t)$  werden folgende Annahmen zugrunde gelegt :

(A1) Die Stichprobenfunktionen sind linksstetig.

(A2) Der Wert der Matrix  $Y(t)$  hängt nur von der Vergangenheit, und nicht von der Zukunft ab.

Die Annahme (A1) impliziert, daß bei Vorliegen von zeitveränderlichen Kovariablen diejenige Belegung der Kovariablenmatrix  $Y(t)$  zu wählen ist, die unmittelbar vor einem Ereigniszeitpunkt zutrifft.

### *Schätzung im linearen Regressionsmodell*

Der Ansatz der Analyse im linearen Regressionsmodell besteht in der Betrachtung kumulierter Regressionsfunktionen, die über einen Zählprozeßansatz geschätzt werden.

Der Vektor der kumulierten Regressionsfunktionen sei mit  $A(t)$  bezeichnet, d.h.

$$A(t) = (A_0(t), A_1(t), \dots, A_r(t))' = \left( \int_0^t \alpha_0(s) ds, \int_0^t \alpha_1(s) ds, \dots, \int_0^t \alpha_r(s) ds \right)' .$$

Seien  $T_1 < T_2 < \dots$  geordnete Ereigniszeitpunkte. Im weiteren wird angenommen, daß keine gleichzeitigen Ereignisse auftreten. Im Falle von Bindungen in den Daten wird zu den (ganzzahligen) Ereigniszeiten eine auf  $(0,1)$  gleichverteilte Zufallszahl addiert, um eine steigende Ordnung in den Ereigniszeitpunkten zu erhalten. Der Schätzer für die kumulierten Regressionsfunktionen ist somit gegeben durch

$$A^*(t) = \sum_{T_k \leq t} X(T_k) I_k , \quad (6)$$

wobei  $X(t) = \left[ Y(t)' Y(t) \right]^{-1} Y(t)'$  die generalisierte Inverse von  $Y(t)$ , und

$I_k$  der n-dimensionale Einheitsvektor ist (an k-ter Stelle mit 1, sonst mit Nullen besetzt). Eine unverzerrte Schätzung  $A^*(t)$  für die kumulierten Regressionsfunktionen ist nur möglich, solange die Matrix  $Y(t)$  vollen Rang besitzt und somit invertierbar ist, d.h. solange  $rg(Y(t)) = r + 1$  gilt.

Um den Einfluß der j-ten Kovariable zu beschreiben, wird der Plot der geschätzten kumulierten Regressionsfunktion  $A_j^*(t)$  gegen die Zeit betrachtet. Dabei gibt die Steigung dieser Funktion zu einem bestimmten Zeitpunkt den Einfluß der j-ten Kovariable in dem betrachteten Moment wieder.

### *Inferenz im linearen Regressionsmodell*

Das lineare Regressionsmodell von Aalen bietet die Möglichkeit den Einfluß der Kovariablen im gesamten Beobachtungszeitraum auf Signifikanz zu testen. Die Teststatistik für die Hypothese

$$H_j : \alpha_j(t) = 0 \quad \forall t$$

wird formuliert durch die j-te Komponente  $U_j$  des Vektors  $U$

$$U = \sum_{T_k} K(T_k) X(T_k) I_k. \quad (7)$$

Die Teststatistik  $U$  ist somit eine gewichtete Summe der Summanden von  $A^*(t)$ , wobei die Summierung über alle Ereigniszeitpunkte erfolgt. Dabei ist die Gewichtematrix  $K(t)$  eine  $(r+1)$ -dimensionale Diagonalmatrix, in der die Diagonalelemente den Gewichtsfunktionen entsprechen. Die Wahl der Gewichtematrix  $K(t)$  hängt einerseits von den Alternativhypothesen andererseits von den relativen Gewichten der einzelnen Zeitintervalle ab. Eine mögliche Belegung der Gewichtematrix ist, die Gewichtsfunktionen entsprechend der Anzahl der Individuen zu wählen, die zu gegebenen Zeitpunkten im Risikosest enthalten sind. Als weitere Möglichkeit für  $K(t)$ , mit der zudem eine größere Power von  $U$  erreicht wird, schlägt Aalen (1989) folgenden Ausdruck vor

$$K(t) = \left\{ \text{diag} \left[ \left( Y(t)' Y(t) \right)^{-1} \right] \right\}^{-1}.$$



Die Teststatistik  $U$  ist asymptotisch multivariat normalverteilt und die Schätzung für die Kovarianzmatrix  $V$  der Teststatistik  $U$  lautet

$$\text{cov}(U) = V = \sum_{T_k} K(T_k)X(T_k)I_k^D X(T_k)' K(T_k).$$

Wird nur eine Komponente getestet, z.B. der Einfluß der  $j$ -ten Kovariable, so ist die Teststatistik

$$z = U_j V_{jj}^{-1} \quad (8)$$

zu verwenden, die unter der Nullhypothese  $N(0,1)$ -verteilt ist. Entsprechend lautet beim simultanen Testen von  $s$  Hypothesen  $H_j$ ,  $j \in A$ , wobei  $A$  eine  $s$ -dimensionale Subgruppe aus  $\{1, \dots, r\}$  ist, die normalisierte Teststatistik

$$U_A' V_A^{-1} U_A$$

und ist  $\chi^2$ -verteilt mit  $s$  Freiheitsgraden. Dabei ist  $U_A$  der entsprechende Subvektor von  $U$  und  $V_A$  die entsprechende Submatrix der Kovarianzmatrix  $V$ .

Analog zur Schätzung ist auch eine Inferenz im nonparametrischen Ansatz nur solange möglich, solange  $Y(t)$  vollen Rang besitzt und somit invertierbar ist.

### *Diagnostik im linearen Regressionsmodell*

Die Überprüfung der Validität des linearen Modells kann beispielsweise mit Arjas-Plots oder Martingalresiduen vorgenommen werden. Der Arjas-Plot ist dabei ein besonders nützliches graphisches Verfahren, das für Subgruppen von Kovariablenkonstellationen die Anzahl der beobachteten und der erwarteten Ereignisse als eine Funktion über die Zeit vergleicht. Die Beurteilung der Modellanpassung mit Hilfe von Martingalresiduen entwickelte Aalen (1993) in Anlehnung an das Cox-Modell. Im Kontext der Martingalresiduen können zwei komplementäre Ansätze herangezogen werden, zum einen eine graphische Überprüfung der Martingalresiduen, der sogenannte Martingalresiduenplot, zum anderen ein analytisches Verfahren zur Identifizierung nicht-linearer Effekte der Kovariablen.

Eine ausführliche Darstellung dieser Ansätze ist in Aalen (1989, 1993) und in Discussion Paper Nr.25 (S.Aydemir, Ü.Aydemir, P.Dirschedl) zu finden.

## **4 Beispiele**

Die in Kapitel 3 vorgestellten Verfahren zur Berücksichtigung zeitveränderlicher Kovariablenstrukturen werden im folgenden an zwei Beispielen illustriert und diskutiert, nämlich an den in der Literatur mehrfach analysierten Daten der Stanford Heart Transplantation Study und an einer Studie über hochmaligne Hirntumoren.

### **4.1 Stanford Heart Transplantation Study**

Die Stanford Heart Transplantation Data, die bereits in Discussion Paper 25 zur Illustration des Zählprozeßansatzes von Aalen herangezogen wurden, werden diesmal mit allen vorgestellten Modellen zur Berücksichtigung zeitveränderlicher Kovariablenstrukturen analysiert. Die Diskussion der Ergebnisse zeitveränderlicher Analysen und insbesondere der Vergleich mit der Baselinesituation stehen dabei im Mittelpunkt dieser Arbeit.

Die Studie umfaßt insgesamt 103 Patienten, von denen 69 Patienten im Beobachtungszeitraum von 1967 bis 1974 ein Herztransplantat erhielten. 75 Patienten verstarben während der Studiendauer, von denen 45 transplantiert wurden. In den nachfolgenden Analysen werden folgende Kovariablen berücksichtigt :

Z1: Alter, Z2: vorangegangene Operation (0=nein, 1=ja), Z3: Transplantation (0=nein, 1=ja), Z4: Wartezeit bis zur Transplantation, Z5: Aufnahmejahr in die Studie.

Die Transplantation in Verbindung mit der entsprechenden Wartezeit ist die einzige zeitveränderliche Kovariable in diesem Datensatz, alle anderen Kovariablen wurden zum Baselinezeitpunkt  $t=0$  ermittelt. Tab.2 zeigt für die Baselinesituation die Ergebnisse des gewöhnlichen Cox-proportional-

hazards- und des linearen Aalen-Modells. Hier wird die Transplantation zurückprojiziert auf den Baselinezeitpunkt, es wird also lediglich zwischen transplantierten und nicht-transplantierten Patienten unterschieden ohne Beachtung des Transplantationszeitpunktes. Ein Vergleich der beiden Ansätze erfolgt zunächst nur bezüglich der Teststatistiken. Dabei stimmen die Ergebnisse beider Modelle bis auf die Kovariable "vorangegangene Operation" ziemlich gut überein. Während der negative Einfluß einer vorangegangenen Operation auf die Hazardrate im linearen Modell signifikant ist, resultiert aus dem Cox-Regressionsmodell keine Auffälligkeit. Dagegen ergibt sich in beiden Modellen übereinstimmend ein auffällig positiver Effekt des Alters und ein signifikant negativer Effekt der Transplantation. Die im linearen Modell geschätzte kumulierte Regressionsfunktion der fixen Transplantationsvariable, die in Abb.1 zu sehen ist, weist in den ersten 150 Tagen eine starke negative Steigung auf. D.h. in diesem Zeitraum besitzt die Transplantation einen deutlichen negativen Effekt auf die Hazardrate, der sich jedoch im weiteren Verlauf verliert. Bei Betrachtung der Transplantation als "fixe" Kovariable kommt man folglich zu der Aussage, daß transplantierte Patienten länger leben als nicht-transplantierte.

**Tab.2:** Stanford Heart Transplantation Data mit Einbezug der Transplantation als fixe Kovariable \*

	fixes Cox-Modell	fixes lineares Modell
Alter	MLE= 0.06 (0.01) TST = 3.94	TST = 3.31
vorangeg. OP	MLE=-0.65 (0.45) TST = -1.44	TST = -3.18
Aufnahmejahr	MLE=-0.09 (0.07) TST = -1.38	TST = -0.80
Transplantation	MLE=-1.64 (0.28) TST = -5.88	TST = -4.26

\* für das Cox-Modell sind Effektschätzung (Standardfehler) und die Wurzel aus der Wald-Teststatistik (TST) angegeben, für das Aalen-Modell die asymptotisch standardnormalverteilte Teststatistik

Wird im Gegensatz zur Baselineanalyse (Tab.2) die zeitliche Struktur der Transplantationsvariable mitberücksichtigt (Tab.3), verschwindet der auffällig negative Effekt der Transplantation, d.h. es ist nun kein Unterschied mehr bezüglich des Überlebens zwischen transplantierten und nicht-transplantierten Patienten zu erkennen. Tab.3 zeigt die Ergebnisse der drei vorgestellten zeitveränderlichen Ansätze. Um die Vergleichbarkeit zum Aalen-Modell zu gewährleisten, wurden für die Teststatistiken des time-dependent und des gepoolten Cox-Modells bereits die Quadratwurzeln der Wald-Statistik angegeben. Bis auf die Kovariable "Operation" stimmen die Teststatistiken des time-dependent Cox-Modells und des linearen Modells gut überein. Abweichungen zum gepoolten Ansatz bestehen nur für die zeitveränderliche Kovariable "Transplantation", da infolge des Datenpools der Transplantationseffekt etwas überschätzt wird. Vor allem für die fixen Kovariablen sind die Effektschätzungen aus dem time-dependent und dem gepoolten Cox-Modell sehr ähnlich. Die geschätzte kumulierte Regressionsfunktion der Transplantation (Abb.2) zeigt bei Beachtung der Zeitstruktur keine eindeutige Steigung mehr. Der bei der Baselineanalyse in den ersten fünf Monaten vorhandene negative Effekt der Transplantation ist bei einer adäquaten zeitveränderlichen Analyse somit nicht mehr auffällig.

**Tab.3:** Stanford Heart Transplantation Data mit Einbezug der Transplantation als Switchvariable

	time-dependent Cox-Modell	gepooltes Cox-Modell	zeitveränderl. lineares Modell
Alter	MLE= 0.03 (0.01) TST = 2.22	MLE= 0.03 (0.01) TST = 2.22	TST = 2.24
vorangeg. OP	MLE=-0.96 (0.44) TST = -2.21	MLE=-0.98 (0.44) TST = -2.24	TST = -4.26
Aufnah- mejahr	MLE=-0.12 (0.07) TST = -1.82	MLE=-0.13 (0.07) TST = -1.88	TST = -1.78
Trans- plantation	MLE=-0.11 (0.31) TST = -0.37	MLE=-0.28 (0.26) TST = -1.08	TST = -0.30

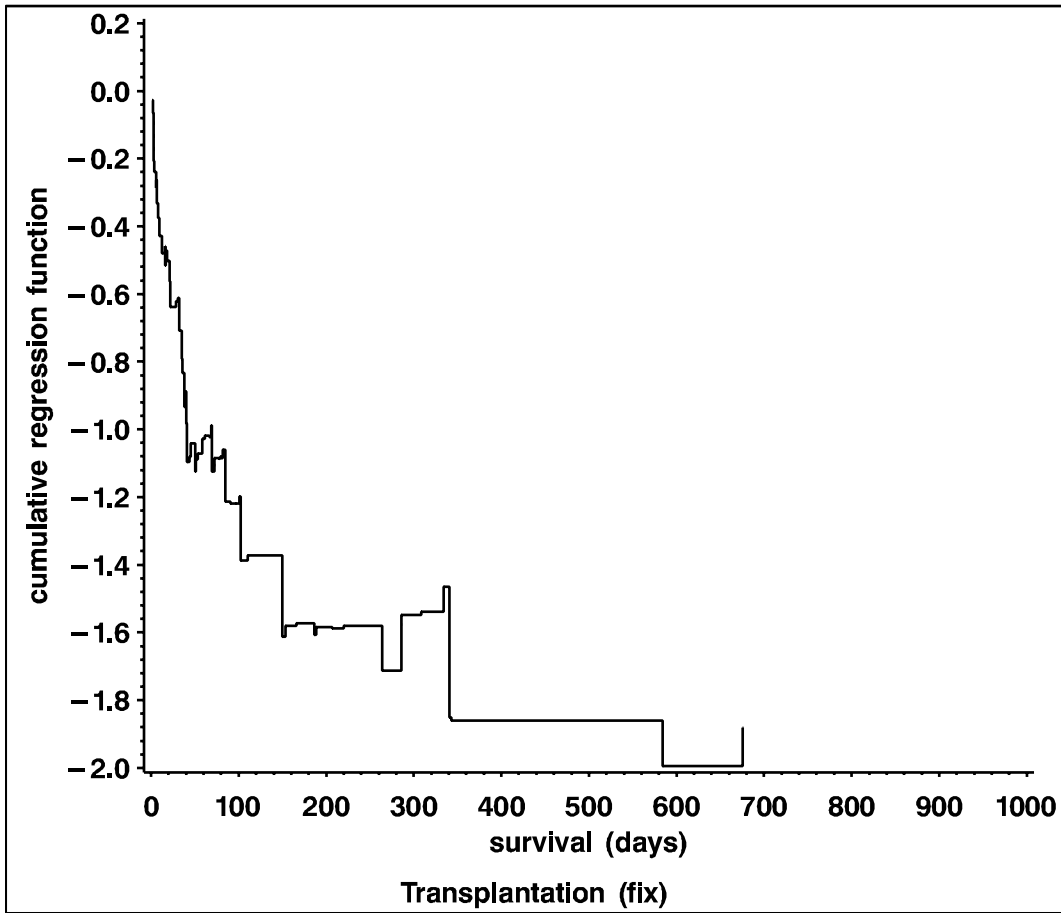


Abb.1 : kumulierte Regressionsfunktion der "*Transplantation* " bei Einbezug als fixe Kovariable (zur Interpretation siehe Seite 15)

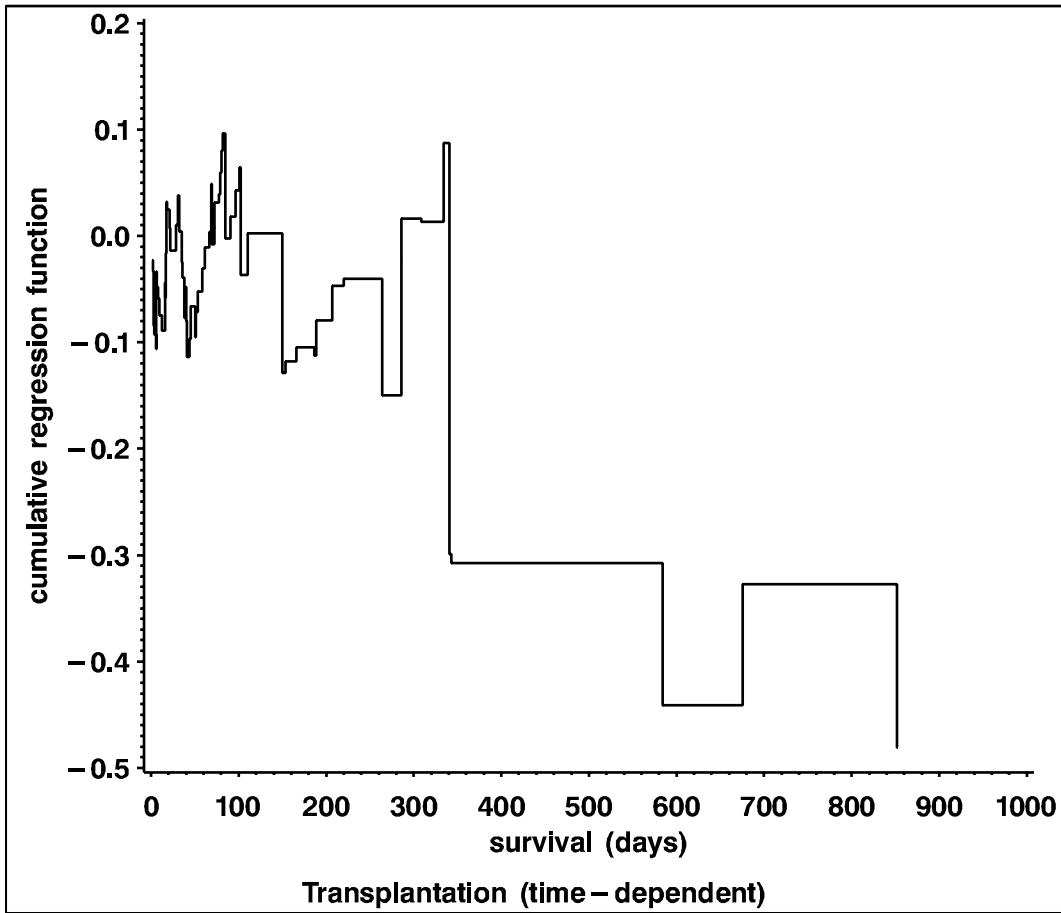


Abb.2 : kumulierte Regressionsfunktion der "Transplantation " bei Einbezug als zeitveränderliche Kovariable

## 4.2 Studie über die Therapie hochmaligner Hirntumoren

Als zweites Beispiel wird die multizentrische BMFT-Therapiestudie über hochmaligne Hirntumoren, auch als Gliomstudie bekannt, vorgestellt. Es handelt sich hierbei um eine postoperative Strahlen- und Chemotherapie mit BCNU und VM26 bei malignen supratentoriellen Gliomen im Erwachsenenalter. Im Zeitraum von 1983 bis 1989 wurden insgesamt 501 Studienpatienten nach operativer Entfernung der malignen Gliome in eine der beiden Chemotherapiegruppen BCNU oder BCNU+VM26 randomisiert. Beide Behandlungsgruppen erhielten zusätzlich zur Chemotherapie, die in 6-wöchigen Zyklen appliziert wurde, eine einheitliche Strahlentherapie. Informationen über potentielle prognostische Faktoren wurden bei der Eingangsuntersuchung sowie bei Verlaufsuntersuchungen im 6-wöchigen Abstand erhoben.

Für die biometrische Betreuung dieser Studie war das Biometrische Zentrum für Therapiestudien in München zuständig. Mittlerweile ist die Gliomstudie von verschiedenen Gruppen analysiert worden, wobei bislang nur die Baselineinformation der Kovariablen berücksichtigt, die Entwicklung im follow-up jedoch außer Acht gelassen wurde (z.B. Ulm et al., 1989). Ausgehend von den Ergebnissen der bisherigen Auswertungen wurden für die Analysen der vorliegenden Arbeit folgende zehn Risikofaktoren ausgewählt, von denen fünf lediglich zum Baselinezeitpunkt und weitere fünf auch im follow-up vorliegen.

Kovariablen, die nur zum Baselinezeitpunkt vorliegen (*fixe Kovariablen*) :

Alter (ordinal: 16-30, 31-40, 41-50, 51-60, 61-70, > 70 Jahre)

Therapiegruppe (0=BCNU, 1=BCNU+VM26)

Malignität (ordinal: Grad 3, Grad 4, Grad 3-4)

Biopsie (0=nein, 1=ja)

Resektion total (0=nein, 1=ja)

Kovariablen, die im Verlauf vorliegen (*veränderliche Kovariablen*) :

Karnofsky ( $\leq 70\%$ , 71-80%,  $\geq 81\%$ )

Cortisongabe (0=nein, 1=ja)

epileptische Anfälle (0=nein, 1=ja)

organisches Psychosyndrom (0=nein, 1=ja)

Bewußtseinsstörungen (0=nein, 1=ja)

Aufgrund fehlender Baselinewerte können in den nachfolgenden Analysen von anfänglich 501 Patienten 476 Patienten berücksichtigt werden. Da die zeitveränderlichen Kovariablen in nur wenigen Fällen öfter als 4 mal wechselten, haben wir aus rechentechnischen Effizienzgründen die Analyse auf maximal 4 Switches beschränkt, so daß ab dem vierten Switch die veränderliche Kovariable auf dem selben Niveau belassen wird. Tab.4 gibt Aufschluß über die Anzahl der dokumentierten Switches der zeitveränderlichen Kovariablen. Bei den Kovariablen Cortisongabe, epileptische Anfälle und organisches Psychosyndrom kam es bei etwa 40% der Fälle zu Änderungen im Verlauf, bei der Kovariable Karnofsky-Index wurde sogar in 71% der Fälle mindestens ein Switch registriert. Dagegen änderten sich die Baselinewerte der Kovariable Bewußtseinsstörung relativ selten, in nur 25% der Fälle wurden hier überhaupt Switches beobachtet, darunter lediglich in 2% zwei Switches und in 1% der Fälle drei Switches.

**Tab.4:** Anzahl der Switches bei den zeitveränderlichen Kovariablen

	kein sw. (in %)	1 sw. (in %)	2 sw. (in %)	3 sw. (in %)	4 sw. (in %)	≥ 5 sw. (in %)
Cortison	59	25	11	4	1	0
Karnofsky	29	20	25	14	7	5
epileptische Anfälle	60	21	10	4	3	2
organisches Psychosynd.	56	25	13	3	2	1
Bewußtseins- störung	75	22	2	1	0	0

Es sei hier vermerkt, daß das Ziel dieser Analyse nicht die Reanalyse der vorliegenden Studie ist, sondern der Vergleich der vorgestellten Ansätze untereinander. Insbesondere steht der Erkenntnisgewinn durch Berücksichtigung der Zeitveränderlichkeit im Mittelpunkt der Diskussion. Das analytische Vorgehen läßt sich zu diesem Zwecke in zwei Schritte gliedern. Zunächst wird die zeitliche Struktur der veränderlichen



Kovariablen nicht beachtet, d.h. alle zehn Risikofaktoren gehen mit ihren Baselinewerten in die Analyse ein. Auf diese Situation wird das gewöhnliche Cox-proportional-hazards-Regressionsmodell und das lineare Modell von Aalen angewendet. Im zweiten Schritt wird jeweils nur von einer der fünf veränderlichen Kovariablen die zeitliche Struktur berücksichtigt, alle anderen Kovariablen gehen wiederum mit ihren Baselinewerten ein. Jede der fünf resultierenden Situationen, die jeweils eine veränderliche und neun fixe Kovariablen enthalten, werden mit dem linearen Regressionsmodell, dem gepoolten und dem time-dependent Cox-Modell geschätzt. In den Ansätzen werden keine Wechselwirkungen berücksichtigt und pro Modell wird auch nur eine Kovariable als veränderlich einbezogen, um eventuelle Beeinflussungen zwischen zeitveränderlichen Kovariablen zu vermeiden.

In Tab.5 sind für die Kovariablen Cortisongabe, Karnofskyindex und Bewußtseinsstörungen unter Berücksichtigung der zeitlichen Struktur der jeweiligen Kovariable die Ergebnisse der drei zeitveränderlichen Ansätze und anschließend schraffiert die Ergebnisse der Baselineanalysen zusammengestellt. Für die Cox-Modelle sind die Maximum-Likelihood-Schätzungen der Effekte und ihre Standardfehler, die Chi-Quadrat verteilten Wald-Teststatistiken und die entsprechenden p-Werte, für die Aalen-Modelle die (asymptotisch standardnormalverteilten) Teststatistiken angegeben.

In der ersten Modellsituation mit der zeitlich veränderlichen Berücksichtigung der Cortisongabe und dem fixen Einbezug der restlichen Kovariablen (Tab.5, Spalte 1), ergibt sich in allen drei zeitveränderlichen Ansätzen auf dem Signifikanzniveau von 5% ein auffälliger Effekt der Cortisonvariable. Die Effektschätzungen im time-dependent und im gepoolten Cox-Modell sind nahezu gleich (time-dependent: 0.91, gepoolt: 0.95).

Wird dagegen die zeitliche Struktur der Cortisongabe nicht berücksichtigt, resultiert aus den Baselineanalysen mit dem gewöhnlichen Cox-proportional-hazards-Modell bzw. dem linearen Aalen-Modell ein weitaus geringerer Effekt (Baseline-Cox-Modell: 0.27).

Der Effektschwund der Baselineanalyse (Abb.3) ist besonders gut im Vergleich zur kumulierten Regressionsfunktion von Abb.4 erkennbar, die

einen wesentlich steileren Anstieg im zeitveränderlichen Fall aufweist. Analog ergeben sich für die zweite Modellsituationen "Karnofsky veränderlich, restliche Kovariablen fix" (Tab.5, Spalte 2) und die dritte Modellsituation "Bewußtseinsstörungen veränderlich, restliche Kovariablen fix" (Tab.5, Spalte 3) annähernd gleiche Ergebnisse in den zeitveränderlichen Analysen, wobei auch hier wieder eine Effektverstärkung gegenüber den Baselineanalysen zu beobachten ist.

**Tab.5:** Modellvergleich - Cortison, Karnofsky und Bewußtseinsstörungen

	1.Modellsituation: Cortison (zeitveränderlich)	2.Modellsituation: Karnofsky (zeitveränderlich)	3.Modellsituation: Bewußtseinsstörung (zeitveränderlich)
<b>time-depend. Cox-Modell</b>	MLE: <b>0.91</b> (0.13) CHI <sup>2</sup> : <b>47.07</b> p: 0.0001	MLE: <b>-0.39</b> (0.08) CHI <sup>2</sup> : <b>25.95</b> p: 0.0001	MLE: <b>1.29</b> (0.12) CHI <sup>2</sup> : <b>118.82</b> p: 0.0001
<b>pooled Cox- Modell</b>	MLE: <b>0.95</b> (0.13) CHI <sup>2</sup> : <b>53.40</b> p: 0.0001	MLE: <b>-0.40</b> (0.08) CHI <sup>2</sup> : <b>29.05</b> p: 0.0001	MLE: <b>1.19</b> (0.12) CHI <sup>2</sup> : <b>99.98</b> p: 0.0001
<b>zeitverän. lin. Modell</b>	Z: <b>6.98</b> p: 0.0001	Z: <b>-5.19</b> p: 0.0001	Z: <b>6.90</b> p: 0.0001
<b>baseline Cox-Modell</b>	MLE: <b>0.27</b> (0.12) CHI <sup>2</sup> : <b>5.46</b> p: 0.019	MLE: <b>-0.25</b> (0.08) CHI <sup>2</sup> : <b>10.49</b> p: 0.001	MLE: <b>0.86</b> (0.35) CHI <sup>2</sup> : <b>6.12</b> p: 0.013
<b>fixes lineares Modell</b>	Z : <b>2.10</b> p: 0.036	Z: <b>-2.54</b> p: 0.011	Z: <b>1.43</b> p: 0.153

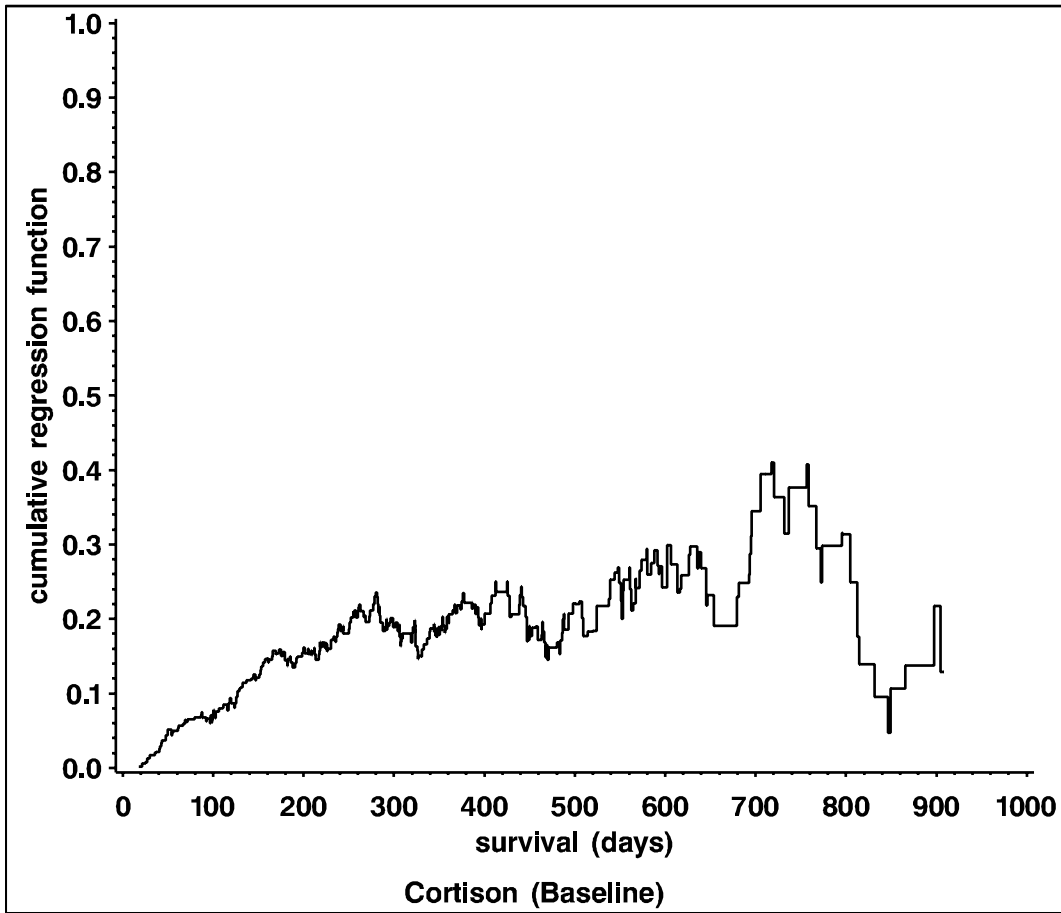


Abb.3: kumulierte Regressionsfunktion für "Cortison" bei Einbezug als fixe Kovariable

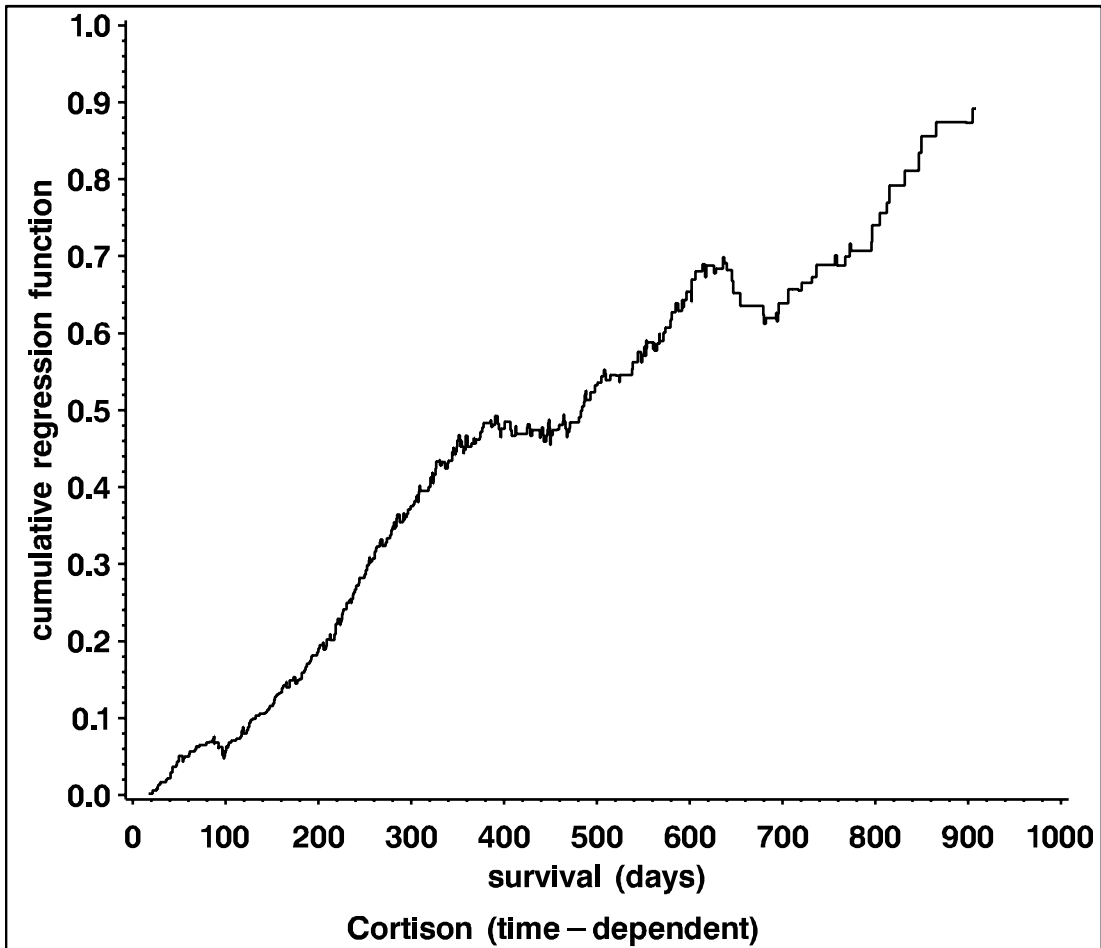


Abb.4: kumulierte Regressionsfunktion für "Cortison" bei Einbezug als zeitveränderliche Kovariable

Weitaus größer ist der Unterschied der Ergebnisse bei Berücksichtigung der Zeitveränderlichkeit der Kovariablen "epileptische Anfälle" und "organisches Psychosyndrom" (Tab.6). Während sich beispielsweise in der Baselineanalyse das Vorhandensein epileptischer Anfälle (4.Modellsituation) negativ auf die Hazardrate auswirkt, resultiert bei Beachtung der zeitlichen Struktur in allen drei zeitveränderlichen Modellen ein auffällig positiver Effekt der epileptischen Anfälle. Die kumulierte Regressionsfunktion besitzt entsprechend im zeitveränderlichen Fall eine durchwegs positive Steigung (Abb.5), während in der Baselinesituation Perioden mit abfallender kumulierter Regressionsfunktion überwiegen (Abb.6).

**Tab.6:** Vergleich der Modellansätze - epileptische Anfälle und organisches Psychosyndrom

	4.Modellsituation epileptische Anfälle (zeitveränderlich)	5.Modellsituation org. Psychosyndrom (zeitveränderlich)
<b>time-dependent Cox-Modell</b>	MLE: <b>0.43</b> (0.12) CHI <sup>2</sup> : <b>13.21</b> p: 0.0003	MLE: <b>0.39</b> (0.12) CHI <sup>2</sup> : <b>10.71</b> p: 0.001
<b>pooled Cox- Modell</b>	MLE: <b>0.38</b> (0.12) CHI <sup>2</sup> : <b>10.17</b> p: 0.001	MLE: <b>0.32</b> (0.12) CHI <sup>2</sup> : <b>6.88</b> p: 0.009
<b>zeitveränderliches lineares Modell</b>	Z: <b>3.04</b> p: 0.002	Z: <b>3.0</b> p: 0.003
<b>baseline Cox-Modell</b>	MLE: <b>-0.17</b> (0.20) CHI <sup>2</sup> : <b>0.76</b> p: 0.382	MLE: <b>-0.14</b> (0.11) CHI <sup>2</sup> : <b>1.73</b> p: 0.188
<b>fixes lineares Modell</b>	Z: <b>-1.31</b> p: 0.190	Z: <b>-1.34</b> p: 0.180

Eine derartige Umkehrung in der Effektrichtung zwischen der Baseline- und der zeitveränderlichen Analyse ist auch bei der Kovariable "organisches Psychosyndrom" zu verzeichnen (Tab.6, 5.Modellsituation). In diesen beiden Modellsituationen kommt es also in Abhängigkeit davon, ob die zeitliche Folge der Kovariablen einbezogen wird oder nicht, zur Umkehrung der Effektvorzeichen und folglich zu gegensätzlichen Schlußfolgerungen. Im Anhang dieser Arbeit findet sich eine ausführliche Zusammenstellung der Ergebnisse der zeitveränderlichen Ansätze und, im Vergleich dazu, der Baselineanalysen für jede der fünf Modellsituationen, in der jeweils eine Kovariable veränderlich und alle restlichen Kovariablen mit ihren Baselinewerten eingehen.

Die Effektschätzungen der Kovariablen, die nur mit zum Baselinezeitpunkt gemessenen Werten vorliegen, sind insgesamt sehr stabil, unabhängig von der Modellwahl und auch unabhängig davon, ob alle Kovariablen fix eingehen oder eine Kovariable veränderlich einbezogen wird. Die Ergebnisse der einzelnen Analysen sind exemplarisch für eine der fixen Kovariablen, nämlich für die Therapie, im Anhang tabelliert.

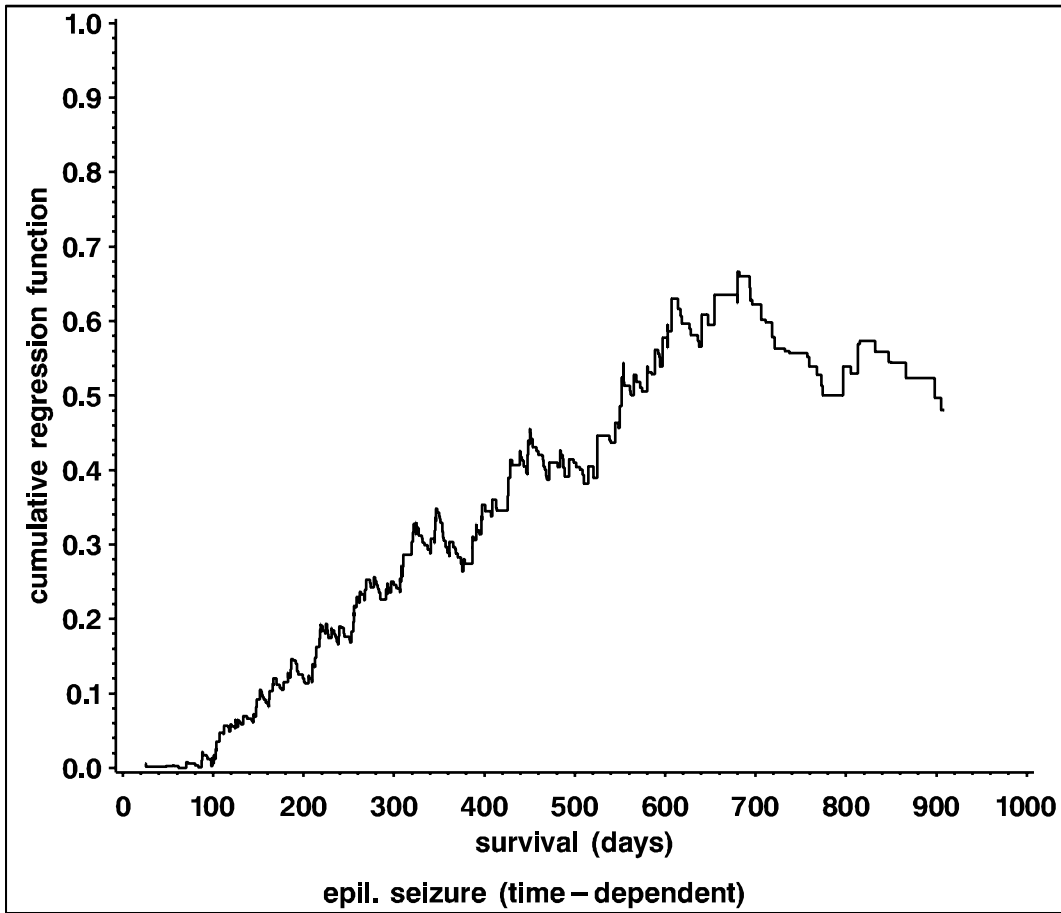


Abb.5 : kumulierte Regressionsfunktion für "epileptische Anfälle " bei Einbezug als zeitveränderliche Kovariable

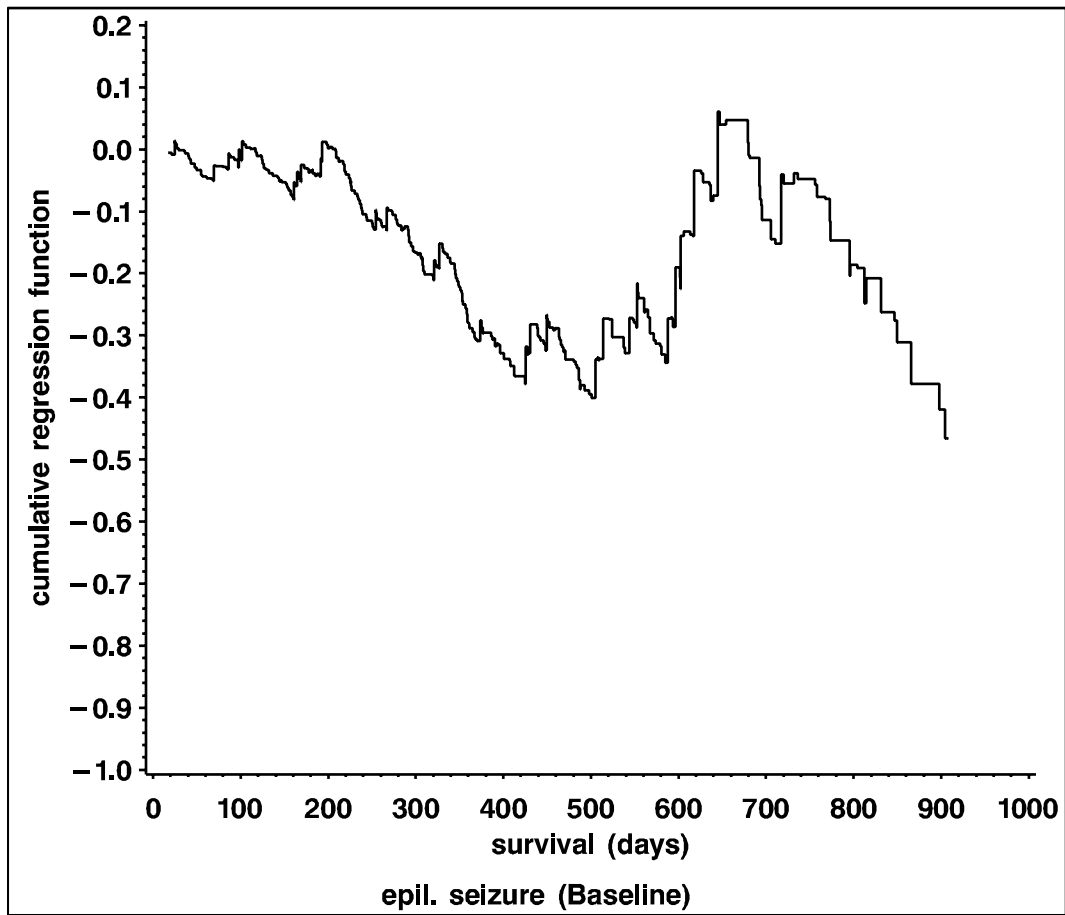


Abb.6 : kumulierte Regressionsfunktion für "epileptische Anfälle" bei Einbezug als fixe Kovariable



## 5 Diskussion

Zur Berücksichtigung der Zeitveränderlichkeit von Kovariablen in multiplen Überlebenszeitanalysen wurden in der vorliegenden Arbeit drei verschiedene Ansätze vorgestellt und an den Daten der Stanford Heart Transplantation Study sowie einer Studie über hochmaligne Hirntumoren illustriert.

Es handelt sich hierbei um das time-dependent Cox-Modell, den linearen Zählprozeßansatz von Aalen und das gepoolte Cox-Regressionsmodell. Obwohl grundsätzlich gepoolte Ansätze aufgrund der Abhängigkeit in der Datenstruktur nicht ganz unbedenklich sind, finden sie in der Literatur bei Vorliegen zeitveränderlicher Strukturen häufig Anwendung (vor allem wegen der einfachen Realisierungsmöglichkeit mit herkömmlichen Softwarepaketen), werden hier aber nur der Vollständigkeit halber aufgeführt.

Ein direkter Vergleich des Cox-Modells und des linearen Aalen-Modells ist infolge struktureller Unterschiede zwischen beiden Ansätzen nicht möglich. Während im Aalen-Modell, ausgehend von einem linearen Einfluß der Risikofaktoren, absolute Zuwächse in den Risiken betrachtet werden, handelt es sich beim Cox-Modell um ein multiplikatives Modell aus dem relative Risiken resultieren. Zudem wird der Vergleich erschwert durch konstante Effektschätzungen im time-dependent Cox-Modell gegenüber zeitveränderlichen Effektschätzungen im linearen Modell.

Trotz dieser modellbedingten Unterschiede lassen sich in allen drei zeitveränderlichen Ansätzen, also bei Einbezug der Kovariablenentwicklung, vergleichbare Effektveränderungen gegenüber den Schätzungen der konventionellen Baselineanalysen erkennen. Besonders deutlich wird die Übereinstimmung der Ergebnisse der zeitveränderlichen Ansätze am Beispiel der Hirntumorendaten. Bei Berücksichtigung der Zeitveränderlichkeit von epileptischen Anfällen beispielsweise, resultiert in allen drei Modellen ein negativer Einfluß auf das Überleben. Im Gegensatz dazu ergeben entsprechende Baselineanalysen mit dem gewöhnlichen Cox-PH-Regressionsmodell oder dem linearen Aalen-Modell, also lediglich bei Einbezug des Baselinewertes der Kovariable, einen positiven und somit protektiven Effekt von epileptischen Anfällen. Dies zeigt, daß die Berücksichtigung der zeitlichen Kovariablenentwicklung in manchen

Fällen durchaus zu neuen Erkenntnissen führen kann, demgegenüber die Beschränkung auf Baselineanalysen und somit der Verzicht auf weitere Kovariableninformationen manchmal völlig andere, unter Umständen verzerrte Resultate liefert.

In der speziellen Situation von Therapiestudien ist eine eventuell mögliche Beeinflussung der Kovariablen durch die Therapie zu beachten. Durch Einbezug der zeitlichen Struktur von "updated covariates" kann es zu einer "Maskierung" bzw. Minderung des Therapieeffektes kommen (Altman et al., 1994). Effekte, die bei der Analyse mit Baselinewerten der Kovariablen, der Therapie zugeschrieben werden, können im time-dependent Ansatz durch die zeitveränderlichen Kovariablen selbst aufgefangen werden. Eine Maskierung des Therapieeffektes wird vor allem in Zusammenhang mit Kovariablen beobachtet, die von der Therapie direkt beeinflußt werden, wie z.B. Laborparameter.

## 6 Anhang : Analyse der Gliomdaten

### Modellsituation 1

veränderliche Kovariable : Cortison

fixe Kovariablen : Karnofsky, epileptische Anfälle, organisches Psychosyndrom, Bewußtseinsstörungen, Alter, Therapie, Malignität, Biopsie, Resektion total.

	time-depend. Cox-Modell	gepooltes Cox-Modell	lineares Modell	Baseline-Cox- Modell (fix)	lineares Modell (fix)
Corti- son	ML: 0.91 (0.13) CHI <sup>2</sup> : 47.07 p: 0.0001	ML: 0.95 (0.13) CHI <sup>2</sup> : 53.40 p: 0.0001	Z: 6.98 p: 0.0001	ML: 0.27 (0.12) CHI <sup>2</sup> : 5.46 p: 0.019	Z: 2.10 p: 0.036
Kar- nofsky	ML:-0.24 (0.08) CHI <sup>2</sup> : 9.92 p: 0.002	ML:-0.28 (0.08) CHI <sup>2</sup> : 13.67 p: 0.0002	Z: -2.67 p: 0.008	ML:-0.25 (0.08) CHI <sup>2</sup> : 10.49 p: 0.001	Z: -2.54 p: 0.011
epi. Anfälle	ML:-0.21 (0.19) CHI <sup>2</sup> : 1.18 p: 0.276	ML:-0.19 (0.19) CHI <sup>2</sup> : 0.96 p: 0.327	Z: -1.35 p: 0.177	ML:-0.17 (0.20) CHI <sup>2</sup> : 0.76 p: 0.382	Z: -1.31 p: 0.190
org. Psych.	ML:-0.17 (0.11) CHI <sup>2</sup> : 2.37 p: 0.124	ML:-0.23 (0.11) CHI <sup>2</sup> : 4.62 p: 0.032	Z: -1.41 p: 0.158	ML:-0.14 (0.11) CHI <sup>2</sup> : 1.73 p: 0.188	Z: -1.34 p: 0.180
Be- wußts.	ML: 0.64 (0.35) CHI <sup>2</sup> : 3.35 p: 0.067	ML: 0.41(0.35) CHI <sup>2</sup> : 1.42 p: 0.233	Z: 1.39 p: 0.164	ML: 0.86 (0.35) CHI <sup>2</sup> : 6.12 p: 0.013	Z: 1.43 p: 0.153
Thera- pie	ML:-0.11 (0.11) CHI <sup>2</sup> : 1.13 p: 0.288	ML: 0.09(0.11) CHI <sup>2</sup> : 0.67 p: 0.413	Z: -1.07 p: 0.285	ML:-0.15 (0.10) CHI <sup>2</sup> : 2.02 p: 0.155	Z:-1.38 p: 0.168

## Modellsituation 2

veränderliche Kovariable : Karnofsky

fixe Kovariablen : Cortison, epileptische Anfälle, organisches Psychosyndrom, Bewußtseinsstörungen, Alter, Therapie, Malignität, Biopsie, Resektion total.

	time-depend. Cox-Modell	gepooltes Cox-Modell	lineares Modell	Baseline-Cox- Modell (fix)	lineares Modell (fix)
Corti- son	ML: 0.18 (0.11) CHI <sup>2</sup> : 2.47 p: 0.116	ML: 0.07 (0.12) CHI <sup>2</sup> : 0.33 p: 0.564	Z: 2.01 p: 0.044	ML: 0.27 (0.12) CHI <sup>2</sup> : 5.46 p: 0.019	Z: 2.10 p: 0.036
Kar- nofsky	ML:-0.39 (0.08) CHI <sup>2</sup> : 25.95 p: 0.0001	ML:-0.40 (0.07) CHI <sup>2</sup> : 29.05 p: 0.0001	Z: -5.19 p: 0.0001	ML:-0.25 (0.08) CHI <sup>2</sup> : 10.49 p: 0.001	Z:-2.54 p: 0.011
epi. Anfälle	ML:-0.16 (0.19) CHI <sup>2</sup> : 0.70 p: 0.401	ML:-0.09 (0.20) CHI <sup>2</sup> : 0.20 p: 0.652	Z: -1.06 p: 0.289	ML:-0.17 (0.20) CHI <sup>2</sup> : 0.76 p: 0.382	Z:-1.31 p: 0.190
org. Psych.	ML:-0.11 (0.11) CHI <sup>2</sup> : 1.02 p: 0.312	ML:-0.16 (0.11) CHI <sup>2</sup> : 2.31 p: 0.128	Z: -0.99 p: 0.322	ML:-0.14 (0.11) CHI <sup>2</sup> : 1.73 p: 0.188	Z:-1.34 p: 0.180
Be- wußts.	ML: 0.92 (0.34) CHI <sup>2</sup> : 7.13 p: 0.007	ML: 0.70 (0.34) CHI <sup>2</sup> : 4.14 p: 0.042	Z: 1.68 p: 0.093	ML: 0.86 (0.35) CHI <sup>2</sup> : 6.12 p: 0.013	Z: 1.43 p: 0.153
Thera- pie	ML:-0.18 (0.10) CHI <sup>2</sup> : 2.86 p: 0.091	ML:-0.18 (0.10) CHI <sup>2</sup> : 3.08 p: 0.079	Z: -1.37 p: 0.171	ML:-0.15 (0.10) CHI <sup>2</sup> : 2.02 p: 0.155	Z:-1.38 p: 0.168

### Modellsituation 3

veränderliche Kovariable : Bewußtseinsstörungen

fixe Kovariablen : Cortison, Karnofsky, epileptische Anfälle, organisches Psychosyndrom, Alter, Therapie, Malignität, Biopsie, Resektion total.

	time-depend. Cox-Modell	gepooltes Cox-Modell	lineares Modell	Baseline-Cox- Modell (fix)	lineares Modell (fix)
Corti- son	ML: 0.22 (0.11) CHI <sup>2</sup> : 3.61 p: 0.057	ML: 0.13 (0.12) CHI <sup>2</sup> : 1.18 p: 0.277	Z: 2.02 p: 0.043	ML: 0.27 (0.12) CHI <sup>2</sup> : 5.46 p: 0.019	Z: 2.10 p: 0.036
Kar- nofsky	ML:-0.28 (0.07) CHI <sup>2</sup> : 13.65 p: 0.0002	ML:-0.27 (0.07) CHI <sup>2</sup> : 13.24 p: 0.0003	Z: -2.82 p: 0.005	ML:-0.25 (0.08) CHI <sup>2</sup> : 10.49 p: 0.001	Z:-2.54 p: 0.011
epi. Anfälle	ML:-0.25 (0.19) CHI <sup>2</sup> : 1.68 p: 0.195	ML:-0.15 (0.20) CHI <sup>2</sup> : 0.57 p: 0.449	Z: -1.45 p: 0.147	ML:-0.17 (0.20) CHI <sup>2</sup> : 0.76 p: 0.382	Z:-1.31 p: 0.190
org. Psych.	ML: 0.23 (0.11) CHI <sup>2</sup> : 4.76 p: 0.029	ML:-0.28 (0.11) CHI <sup>2</sup> : 7.00 p: 0.008	Z: -1.89 p: 0.059	ML:-0.14 (0.11) CHI <sup>2</sup> : 1.73 p: 0.188	Z:-1.34 p: 0.180
Be- wußts.	ML: 1.29 (0.12) CHI <sup>2</sup> : 118.82 p: 0.0001	ML: 1.19 (0.12) CHI <sup>2</sup> : 99.98 p: 0.0001	Z: 6.90 p: 0.0001	ML: 0.86 (0.35) CHI <sup>2</sup> : 6.12 p: 0.013	Z: 1.43 p: 0.153
Thera- pie	ML:-0.12 (0.10) CHI <sup>2</sup> : 1.36 p: 0.244	ML:-0.10 (0.11) CHI <sup>2</sup> : 0.91 p: 0.339	Z: -1.34 p: 0.180	ML:-0.15 (0.10) CHI <sup>2</sup> : 2.02 p: 0.155	Z:-1.38 p: 0.168

## Modellsituation 4

veränderliche Kovariable : epileptische Anfälle

fixe Kovariablen : Cortison, Karnofsky, organisches Psychosyndrom, Bewußtseinsstörungen, Alter, Therapie, Malignität, Biopsie, Resektion total.

	time-depend. Cox-Modell	gepooltes Cox-Modell	lineares Modell	Baseline-Cox- Modell (fix)	lineares Modell (fix)
Corti- son	ML: 0.22 (0.11) CHI <sup>2</sup> : 3.77 p: 0.052	ML: 0.15 (0.12) CHI <sup>2</sup> : 1.56 p: 0.211	Z: 1.59 p: 0.112	ML: 0.27 (0.12) CHI <sup>2</sup> : 5.46 p: 0.019	Z: 2.10 p: 0.036
Kar- nofsky	ML:-0.23 (0.08) CHI <sup>2</sup> : 9.22 p: 0.002	ML:-0.28 (0.08) CHI <sup>2</sup> : 12.66 p: 0.0004	Z: -2.42 p: 0.015	ML:-0.25 (0.08) CHI <sup>2</sup> : 10.49 p: 0.001	Z:-2.54 p: 0.011
epi. Anfälle	ML: 0.43 (0.12) CHI <sup>2</sup> : 13.21 p: 0.0003	ML: 0.38 (0.12) CHI <sup>2</sup> : 10.17 p: 0.001	Z: 3.04 p: 0.002	ML:-0.17 (0.20) CHI <sup>2</sup> : 0.76 p: 0.382	Z:-1.31 p: 0.190
org. Psych.	ML:-0.07 (0.11) CHI <sup>2</sup> : 0.46 p: 0.497	ML:-0.17 (0.11) CHI <sup>2</sup> : 2.38 p: 0.122	Z: -0.86 p: 0.390	ML:-0.14 (0.11) CHI <sup>2</sup> : 1.73 p: 0.188	Z:-1.34 p: 0.180
Be- wußts.	ML: 0.81 (0.35) CHI <sup>2</sup> : 5.41 p: 0.02	ML: 0.59 (0.35) CHI <sup>2</sup> : 2.87 p: 0.090	Z: 1.51 p: 0.131	ML: 0.86 (0.35) CHI <sup>2</sup> : 6.12 p: 0.013	Z: 1.43 p: 0.153
Thera- pie	ML:-0.13 (0.10) CHI <sup>2</sup> : 1.58 p: 0.209	ML:-0.12 (0.11) CHI <sup>2</sup> : 1.29 p: 0.255	Z: -1.19 p: 0.234	ML:-0.15 (0.10) CHI <sup>2</sup> : 2.02 p: 0.155	Z:-1.38 p: 0.168

## Modellsituation 5

veränderliche Kovariable : organisches Psychosyndrom

fixe Kovariablen : Cortison, Karnofsky, epileptische Anfälle, Bewußtseinsstörungen, Alter, Therapie, Malignität, Biopsie, Resektion total.

	time-depend. Cox-Modell	gepooltes Cox-Modell	lineares Modell	Baseline-Cox- Modell (fix)	lineares Modell (fix)
Corti- son	ML: 0.31 (0.11) CHI <sup>2</sup> : 7.28 p: 0.007	ML: 0.25 (0.12) CHI <sup>2</sup> : 4.42 p: 0.035	Z: 2.53 p: 0.011	ML: 0.27 (0.12) CHI <sup>2</sup> : 5.46 p: 0.019	Z: 2.10 p: 0.036
Kar- nofsky	ML:-0.20 (0.08) CHI <sup>2</sup> : 6.88 p: 0.009	ML:-0.24 (0.08) CHI <sup>2</sup> : 9.86 p: 0.002	Z: -2.27 p: 0.023	ML:-0.25 (0.08) CHI <sup>2</sup> : 10.49 p: 0.001	Z:-2.54 p: 0.011
epi. Anfälle	ML:-0.13 (0.19) CHI <sup>2</sup> : 0.48 p: 0.488	ML:-0.11 (0.20) CHI <sup>2</sup> : 0.30 p: 0.584	Z: -1.11 p: 0.267	ML:-0.17 (0.20) CHI <sup>2</sup> : 0.76 p: 0.382	Z:-1.31 p: 0.190
org. Psych.	ML: 0.39 (0.12) CHI <sup>2</sup> : 10.71 p: 0.001	ML: 0.32 (0.12) CHI <sup>2</sup> : 6.88 p: 0.009	Z: 3.00 p: 0.003	ML:-0.14 (0.11) CHI <sup>2</sup> : 1.73 p: 0.188	Z:-1.34 p: 0.180
Be- wußts.	ML: 0.76 (0.35) CHI <sup>2</sup> : 4.84 p: 0.028	ML: 0.54 (0.35) CHI <sup>2</sup> : 2.45 p: 0.118	Z: 1.41 p: 0.158	ML: 0.86 (0.35) CHI <sup>2</sup> : 6.12 p: 0.013	Z: 1.43 p: 0.153
Thera- pie	ML:-0.16 (0.10) CHI <sup>2</sup> : 2.34 p: 0.126	ML:-0.15 (0.10) CHI <sup>2</sup> : 1.89 p: 0.169	Z: -1.55 p: 0.121	ML:-0.15 (0.10) CHI <sup>2</sup> : 2.02 p: 0.155	Z:-1.38 p: 0.168

## 7 Literatur

**Aalen, O.O.** (1989). *A linear regression model for the analysis of life times*. *Statistics in Medicine*, 8, 907-925.

**Aalen, O.O.** (1993). *Further results on the non-parametric linear regression model in survival analysis*. *Statistics in Medicine*, 12, 1569-1588.

**Altman, D.G., De Stavola, B.** (1994). *Practical problems in fitting a proportional hazards model to data with updated measurements of the covariates*. *Statistics in Medicine*, 13, 301-341.

**Andersen, P.K., Borgan, O., Gill, R.D., Keiding, N.** (1993). *Statistical models based on counting processes*. Springer-Verlag, New York.

**Arjas, E.** (1986). *Stanford heart transplantation data revisited: a real time approach*. In Moolgavkar, S.H., Prentice, R.L., editors, *Modern Statistical Methods in Chronic Disease Epidemiology*, 65-81. Wiley, New York.

**Aydemir, S., Aydemir, Ü., Dirschedl, P.** (1996). *Das lineare Regressionsmodell von Aalen zur Analyse von Überlebenszeiten unter Berücksichtigung zeitveränderlicher Kovariablen*. Discussion Paper 25, SFB 386.

**Crowley, J., Hu, M.** (1977). *Covariance analysis of heart transplant survival data*. *Journal of the American Statistical Association*, 72, 27-36.

**D'Agostino, R.B., Lee, M., Belanger, A.J.** (1990). *Relation of pooled logistic regression to time dependent Cox regression analysis: the Framingham Heart Study*. *Statistics in Medicine*, 9, 1501-1515.

**Henderson, R., Milner, A.** (1991). *Aalen plots under proportional hazards*. *Applied Statistics*, 40, 401-409.

**Huffer, F.W., McKeague, I.W.** (1991). *Weighted least squares estimation for Aalen's additive risk model*. *JASA*, 86, 114-129.



**Hughes, M.D., Raskino, C.L., Pocock, S.J. et al.** (1992). *Prediction of short-term survival with an application in primary biliary cirrhosis*. *Statistics in Medicine*, 11, 1731-1745.

**Kalbfleisch, J.D., Prentice, R.L.** (1980). *The statistical analysis of failure time data*. Wiley, New York.

**Mau, J.** (1986). *On a graphical method for the detection of time-dependent effects of covariates in survival data*. *Appl. Statist.*, 35 (3), 245-255.

**Mau, J.** (1988). *A comparison of counting process models for complicated life histories*. *Applied Stochastic Models and Data Analysis*, 4, 283-298.

**Therneau, T.M., Grambsch, P.M., Fleming, T.R.** (1990). *Martingale-based residuals for survival models*. *Biometrika*, 77 (1), 147-160.

**Ulm, K., Schmoor, C., Sauerbrei, W., Kemmler, G., Aydemir, Ü., Müller, B., Schumacher, M.** (1989). *Strategien zur Auswertung einer Therapiestudie mit der Überlebenszeit als Zielkriterium*. *Biometrie und Informatik in Medizin und Biologie*, 20 (4), 171-205.