Aus dem Institut für Medizinische Epidemiologie, Biometrie und Informatik der Medizinischen Fakultät der Martin-Luther-Universität Halle-Wittenberg (Direktor: Prof. Dr. Johannes Haerting)

## Shared Frailty-Modelle zur Analyse von Überlebenszeiten

- ein Vergleich von Software-Lösungen -

Dissertation zur Erlangung des akademischen Grades Doktor rerum medicarum (Dr. rer. medic.) für das Fachgebiet Medizinische Biometrie

> vorgelegt der Medizinischen Fakultät der Martin-Luther-Universität Halle-Wittenberg

von Katharina Hirsch geboren am 02.11.1985 in Görlitz

Betreuer: Prof. Dr. Andreas Wienke

Gutachter: 1. Prof. Dr. Andreas Wienke, Halle (Saale)2. Prof. Dr. Peter Schlattmann, Jena3. Prof. Dr. Jan Beyersmann, Ulm

Eröffnet: 11.02.2014 Verteidigt: 24.10.2014

## Referat

In der Medizin stellt die Analyse von Ereigniszeiten ein wichtiges Themengebiet dar. Fragen nach Lebensdauer, progressionsfreiem Überleben oder medianer Überlebenszeit sind insbesondere in der Onkologie von Interesse. Eine Besonderheit von Ereigniszeitdaten stellt das Auftreten von Zensierungen dar. Die Analyse von solchen unvollständigen Daten erfordert spezielle statistische Methoden, die durch die Ereigniszeitanalyse bereitgestellt werden. Ein Ziel der Ereigniszeitanalyse ist es, den Einfluss von prognostischen Faktoren, beziehungsweise Risikofaktoren auf das Eintreten von Ereignissen bei Patienten zu ermitteln. Ein wichtiger Modellansatz um dies zu erreichen ist das proportionale Hazardsmodell nach Cox [1]. Dieses Regressionsmodell erfordert jedoch Homogenität innerhalb der Studienpopulation, sowie voneinander unabhängige Beobachtungen. Eine Erweiterung des Cox-Modells ist das shared Frailty-Modell. Es ist fähig gruppierte Lebensdauerdaten mit unbeobachteter Heterogenität zu analysieren. Diese shared Frailty-Modelle etablierten sich zunehmend in den letzten Jahren und wurden in bekannte Statistik-Pakete wie SAS und R eingebunden. Es existieren zahlreiche Softwarerealisierungen des shared Frailty-Modells, von denen die Folgenden in der vorliegenden Arbeit betrachtet wurden: SPGAM1, SPLN3, PROC PHREG, %PCFrailty, coxph, coxme und frailtyPenal. Zu diesen Funktionen liegen keine vergleichenden Validierungsstudien vor, was es dem Nutzer erschwert eine Funktion zu präferieren. Ziel der vorliegenden Arbeit war es, die Funktionen bezüglich ihrer Anwendbarkeit, Güte und Limitationen zu betrachten und miteinander zu vergleichen. Aufgrund dessen wurden umfangreiche Simulationen für unterschiedliche Szenarien durchgeführt. Die Simulationsergebnisse zeigen, dass es keine nennenswerten Unterschiede zwischen parametrischen und semiparametrischen Modellansätzen bezüglich ihrer Schätzgüte gibt. Fünf Funktionen erwiesen sich als geeignet und wurden daraufhin auch an einem realen Datensatz angewandt. Benutzt wurden dafür die Daten der Halle Lung Cancer (HALLUCA)-Studie, einer populationsbasierten prospektiven Studie mit 1696 eingeschlossen Lungenkrebspatienten aus den Regionen Halle und Dessau im Zeitraum von 1996 bis 2002. Unabhängig von der Frailtyverteilung führten die verwendeten Funktionen zu sehr ähnlichen Schätzungen der festen Effekte, was die Robustheit dieser Verfahren unterstreicht.

Hirsch, Katharina: Shared Frailty-Modelle zur Analyse von Überlebenszeiten ein Vergleich von Software-Lösungen, Halle (Saale), Univ., Med. Fak., Diss., 80 Seiten, 2014

## Inhaltsverzeichnis

1	Einleitung	1
1.1	Grundlagen der Lebensdaueranalyse	2
1.1.1	Überlebensfunktion	3
1.1.2	Hazardfunktion	4
1.2	Rechtszensierung	4
1.3	Parametrische Modelle	7
1.3.1	${ m Exponential verteilung}$	7
1.3.2	Gompertzverteilung	8
1.3.3	Piecewise-constant Verteilung	8
1.4	Kaplan-Meier Schätzer	10
1.5	Proportionales Hazards Modell nach Cox	11
1.6	Poisson-Regressionsmodell	12
1.7	Frailty-Modelle	14
1.8	Shared Frailty-Modell	16
1.8.1	Shared Gamma Frailty-Modell	17
1.8.2	Shared Log-normales Frailty-Modell	18
1.9	Gemischtes Poission-Regressionsmodell	20
2	Zielstellung	22
3	Material und Methoden	23
3.1	R und SAS Funktionen zur Analyse von gruppierten Lebensdauerdaten	23
3.1.1	R-Funktionen	24
3.1.2	SAS-Makros/Prozeduren	27
3.1.3	Schätzalgorithmen	31
3.2	Validierungsmethoden zur quantitativen Bewertung der untersuchten Funk-	
	tionen	32
3.2.1	Mittlerer quadratischer Fehler	34
3.2.2	Zweiseitige Konfidenzintervalle	35
3.2.3	Überdeckungsrate	37
3.3	Daten	38
3.3.1	Simulationen	38
3.3.2		
	Halluca Studie	39

4.1	Ergebnisse der Funktionen SPGAM1, coxph und frailtyPenal	43		
4.1.1	Ergebnisse der durch ein PH Modell generierten Daten	43		
4.1.2	Ergebnisse der durch ein shared Gamma Frailty-Modell generierten Daten	47		
4.2	Ergebnisse der Funktionen PROC PHREG, %PCFrailty, SPLN3, coxme			
	und coxph	52		
4.2.1	Ergebnisse der durch ein PH Modell generierten Daten	52		
4.2.2	Ergebnisse der durch ein shared Log-normales Frailty-Modell generierten			
	Daten	56		
4.3	Ergebnisse der Halluca-Studie	61		
4.3.1	Ergebnisse des shared Gamma Frailty-Modells	62		
4.3.2	Ergebnisse des shared Log-normalen Frailty-Modells	64		
5	Diskussion	66		
6	Zusammenfassung	72		
7	Literaturverzeichnis	74		
8	Thesen	79		
Leben	slauf			
$\mathbf{Selbst}$	ständigkeitserklärung			

Erklärung über frühere Promotionsverfahren

# Abkürzungsverzeichnis und Symbolverzeichnis

## Abkürzungen

PL Schätzer	Product Limit Schätzer
PH Modell	Proportionales Hazardsmodell
PPL	Penalized Partial Likelihood
EM	Expectation-Maximization
MCMC	Marcov Chain Monte Carlo
SAS	Statistik-Software
R	Statistik-Software die frei erhältlich ist
STATA	Statistik-Software
WinBUGS	Statistik-Software für bayesianische Analysen unter
	der Verwendung von MCMC-Methoden
BLUP	Best Linear Unbiased Prediction (beste lineare
	unverzerrte Vorhersage)
GLMM	Generalized Linear Mixed Model
SF	Standardfehler
KI	Konfidenzintervall
MQF	Mittlerer quadratischer Fehler
$\mathrm{FG}$	Freiheitsgrad
RMSE	Root mean square error
PL-KI	Profile Likelihood Konfidenzintervall
LR	Likelihood-Ratio Teststatistik
HALLUCA-Studie	Halle Lung Cancer Studie
ECOG	Eastern Cooperative Oncology Gruppe
	(Einstufung des Selbstständigkeitsstatuses)
$\operatorname{sclc}$	Small cell lung cancer
nsclc	Non small cell lung cancer
TNM I - IV	Tumorstadien
IQR	Interquartile Range

## Symbole

t	Zeit
$T^*$	Zeitdauer bis Eintreten eines Ereignisses
$f(\cdot)$	Dichtefunktion
$S(\cdot)$	Überlebensfunktion
$\mu(\cdot)$	Hazardfunktion
$M(\cdot)$	Kumulierte Hazardfunktion
$F(\cdot)$	Kumulierte Verteilungsfunktion
C	Zensierungszeit
T	${ m Beobachtungszeitzeit}$
$\delta$	Zensierungsindikator
$Exp(\lambda)$	Exponential verteilung mit den Parameter $\lambda$
$\mathbf{E}$	Erwartungswert
$\mathbf{V}$	Varianz
$W(\lambda, \nu)$	Weibull-Verteilung mit den Parametern $\lambda,\nu$
$G(\lambda, \varphi)$	Gompertz-Verteilung mit den Parametern $\lambda,\varphi$
a	Intervallgrenzen
$\hat{S}(t)$	Kaplan-Meier-Schätzer
x	Kovariablenvektor
$\beta$	${ m Regressionskoeffizientenvektor}$
$\mu_0(\cdot)$	Basishazardfunktion
$L(\cdot)$	Likelihoodfunktion
$\ell(\cdot)$	${ m Log-Likelihoodfunktion}$
Z	Frailtyvariable
$\Gamma(k,\lambda)$	Gamma-Verteilung mit den Parametern $k,\lambda$
$\sigma^2$	Varianz der Frailtyvariable
$N(\mu,\sigma^2)$	Normalverteilung mit den Parametern $\mu,\sigma^2$
W	zufälliger Effekt
$s^2$	Varianz des zufälligen Effektes
Ω	Piece (Intervall)
$\hat{p}$	Parameterschätzer
U/O	untere/obere Grenze eines Konfidenzintervalls
$\alpha$	${ m Irrtums wahrscheinlichkeit}$
h	nominale Überdeckungswahrscheinlichkeit

### 1 Einleitung

Um die Zeiten bis zum Eintreten eines bestimmten Ereignisses auszuwerten, sind spezielle Analyseverfahren notwendig. Diese werden durch die sogenannte Lebensdaueranalyse (survival analysis) realisiert. Von Interesse ist dabei der Zeitpunkt des Zustandswechsels von einem Ausgangszustand zu einem Endzustand. Damit ist der betrachtete Outcome in der Lebensdaueranalyse die Ereigniszeit. Interessierende Outcomes sind beispielsweise die Zeit bis zum Eintritt des Todes, die Zeit bis zur Entlassung aus dem Krankenhaus, die Zeit bis zum Abstillen oder die Zeit bis zum Auftreten einer Erkrankung. Ein wichtiges Ziel der Lebensdaueranalyse ist es, den Einfluss von prognostischen Faktoren/Risikofaktoren auf die Ereigniszeit klären. Durch die Lebensdaueranalyse können unterschiedliche medizinisch und epidemiologisch wichtige Fragen geklärt werden. Typische Fragestellungen sind beispielsweise [2]:

- Ist das Sterberisiko für Empfänger einer Niere bei Lebendspenden geringer als bei Todspenden?
- Wie wirken sich der Body-Mass-Index und das Rauchverhalten auf die Brustkrebsinzidenz aus?
- Haben HIV-Patienten eine bessere Prognose durch eine neu entwickelte Therapie?

Neben den Bereichen der Medizin und Epidemiologie werden die Methoden der Lebensdaueranalyse auch in der Demografie und Soziologie als auch im Bereich der Technik verwendet. In der Technik spricht man von der sogenannten Zuverlässigkeitstheorie (Reliability Theory). In all diesen Gebieten werden Verfahren angewendet, welche auf denselben grundlegenden mathematischen Gesetzmäßigkeiten beruhen.

Die Lebensdaueranalyse, oder auch Ereigniszeitanalyse, unterscheidet sich von anderen Gebieten der Statistik durch das sogenannte "conditioning" und das Auftreten von Zensierungen [3]. Hierbei gelten folgende Definitionen:

- conditioning: Das zur Berechnung verwendete Wissen ist nur zum Zeitpunkt der Analyse vorhanden. Es verändert sich jedoch im Zeitverlauf und ist damit zeitabhängig. Es wird für die Analyse des individuellen Sterberisikos nur der Teil der Population verwendet, welcher zum Zeitpunkt der Beobachtung unter Risiko für das Ereignis steht.
- Zensierung: Dabei handelt es sich um einen Mechanismus, welcher zu unvollständigen Beobachtungen führt. Dies bedeutet, von einem Individuum ist zum

Beispiel nur bekannt, dass es bis zum Zensierungszeitpunkt ereignislos blieb. Die eigentliche Ereigniszeit ist jedoch unbekannt.

Ferner konzentriert sich diese Arbeit auf einmalige Ereignisse, wobei jedoch die Untersuchung von rekurrenten Ereignissen, wie wiederholte Asthmaanfälle oder Infektionen, ebenfalls möglich wäre. Wichtig ist allerdings, dass das zu analysierende Ereignis sowie der Anfangszeitpunkt der Zeitmessung vor der Datenerhebung genau definiert sind. In dieser Arbeit werden die Begriffe, Überlebenszeit, Lebensdauer, sowie die Zeit bis zum Eintreten eines Ereignisses als Synonyme für die Ereigniszeit verwendet.

Die vorliegende methodische Arbeit hat Anwendungsmöglichkeiten in den Gebieten Medizin und Epidemiologie. Sie hat die Untersuchung des Einflusses von prognostischen Faktoren bzw. Risikofaktoren sowie von Gruppeneffekten mittels der Lebensdaueranalyse zum Gegenstand. Eben dieser Effekt der Gruppenzugehörigkeit erfordert spezialisiertere Methoden der Lebensdaueranalyse. Eines dieser Modelle ist das shared Frailty-Modell [4], [5]. Solch spezialisierte Modelle lassen sich oft nur mit erheblichem technischen Aufwand in Softwarepakete implementieren. Die ersten Umsetzungen bilden daher meist frei erhältliche Makros und Funktionen. Der daraus resultierende Nachteil liegt häufig in der wenig geprüften Validität solcher Makros und Funktionen. Da es im Bereich der Softwarevalidierung von shared Frailty-Modellen noch keine vergleichenden Studien gibt, versucht die vorliegende Arbeit diese Lücke zu schließen.

Diese Arbeit stellt zunächst die grundlegenden Eigenschaften, Funktionen und Methoden der Lebensdaueranalyse dar. Dabei wird besonders auf den Unterschied zwischen semiparametrischen und parametrischen Modellansätzen eingegangen (Abschnitte 1.5 und 1.6). Im Anschluss daran wird die Motivation der Arbeit im Kapitel Fragestellung genauer erläutert. Im darauffolgenden Kapitel 3 wird das shared Frailty-Modell sowie der Alternativansatz des gemischten Poisson Regressionsmodells eingeführt und erläutert. Zum Materialteil gehören neben den beschriebenen simulierten und realen Datensätzen auch die verwendeten Funktionen. Diese werden in Abschnitt 3.1 bezüglich ihrer Funktionsweise und Verwendung näher erläutert. Mithilfe der simulierten Daten soll eine Validierung durchgeführt werden. Die Ergebnisse der untersuchten Funktionen sind in Abschnitt 4 dargestellt. Diese Ergebnisse werden anschließend im Kapitel 5 der Arbeit besprochen. Zum Abschluss folgt eine Zusammenfassung der Arbeit.

#### 1.1 Grundlagen der Lebensdaueranalyse

In der Lebensdaueranalyse wird die Verteilung der nicht-negativen Variable Ereigniszeit  $T^*$  oft durch eine der folgenden Funktionen beschrieben

- Dichtefunktion f(t),
- Überlebensfunktion S(t),
- Hazardrate bzw. Hazardfunktion  $\mu(t)$  und
- kumulative Hazard function M(t).

Wenn eine dieser vier Funktionen bekannt ist, können die anderen daraus eindeutig abgeleitet werden. Soweit nicht anders angegeben, sind alle Funktionen über dem Intervall  $(0, \infty)$  definiert. In der vorliegenden Arbeit werden ausschließlich stetige Ereigniszeiten betrachtet.

#### 1.1.1 Überlebensfunktion

Um "time-to-event"-Daten zu beschreiben ist die Überlebensfunktion S(t) geeignet. Sie wird als Wahrscheinlichkeit des Überlebens eines Individuums über den Zeitpunkt t hinaus definiert. Sei F(t) die Verteilungsfunktion der Ereigniszeit  $T^*$ :

$$F(t) = P(T^* \le t) = \int_0^t f(s) ds.$$

Damit ist S(t) = 1 - F(t):

$$S(t) = P(T^* > t).$$

Für jede stetige Zufallsvariable  $T^*$  ist S(t) eine stetige Funktion. Allgemein ist S(t) eine monoton fallende Funktion mit S(0) = 1. Wenn die Überlebensfunktion im zeitlichen Verlauf  $(t \to \infty)$  gegen 0 strebt  $(S(t) \to 0)$  spricht man von einer nicht entarteten Verteilung.

Darüber hinaus ist die Überlebensfunktion das Integral über die Dichtefunktion f(t):

$$S(t) = \int_{t}^{\infty} f(s) ds,$$

demnach ist

$$f(t) = -\frac{dS(t)}{dt}$$

Die Dichte f(t) ist eine nicht-negative Funktion, welche die Verteilung der Zufallsvariable  $T^*$  vollständig und eindeutig bestimmt.

#### 1.1.2 Hazardfunktion

Eine wichtige Funktion in der Lebensdaueranalyse ist die Hazardfunktion  $\mu(t)$ . Je nach Anwendungsbereich wird sie auch Risikofunktion, Sterberate, Inzidenzrate, Ausfallrate oder Hazardrate genannt. Sie wird definiert durch:

$$\mu(t) = \lim_{\varepsilon \to 0} \frac{P(t < T^* \le t + \varepsilon | T^* > t)}{\varepsilon}.$$

Die Hazardfunktion kann als Rate oder Intensität interpretiert werden [6].

Ist  $T^*$  eine stetige, nicht-negative Zufallsvariable, dann lässt sich  $\mu(t)$  mittels Dichte und Überlebensfunktion darstellen:

$$\mu(t) = \frac{f(t)}{S(t)}.$$

Damit verbunden ist die kumulative Hazardfunktion M(t), welche die Form

$$M(t) = \int_0^t \mu(s)ds = -\ln[S(t)]$$

hat. Folglich kann die Überlebensfunktion durch Hazard- und kumulative Hazardfunktion dargestellt werden

$$S(t) = e^{-M(t)} = e^{-\int_0^t \mu(s)ds}$$

Diese Beziehung spielt eine zentrale Rolle in der gesamten Lebensdaueranalyse. Sie zeigt die Charakterisierung der Verteilungs- und Überlebensfunktion durch die Hazardfunktion. Die Hazardfunktion ist am besten dazu geeignet Aussagen über die Verteilung der Ereigniszeiten zu treffen. Sie wird genutzt, um qualitative Informationen über den Ausfallmechanismus zu geben und Veränderungen über die Zeit zu beschreiben. Grund hierfür ist ihre aussagekräftige Interpretation und ihre Bedeutung für die einfache Darstellung der Likelihoodfunktion. Hazardfunktionen können unterschiedliche Formen annehmen. Abbildung 1 illustriert bespielhaft unterschiedliche Zeitverläufe, die für Hazardfunktionen auftreten können.

#### 1.2 Rechtszensierung

Oft ist die Ereigniszeit  $T^*$  nicht vollständig beobachtbar. Dies wird durch Zensierung hervorgerufen. Ist das interessierende Ereignis bis zum Ende des Studienzeitraumes nicht beobachtet worden, gelten die Daten als rechts zensiert. Es sind zwei unterschiedliche Zeiten zu betrachten. Die Ereigniszeit  $T^*$  stellt die Zeit von einem definierten



Abbildung 1: Mögliche Formen der Hazardfunktion.

Startpunkt bis zum Eintreten des interessierenden Ereignisses dar. Die Ereigniszeiten  $T_i^*$  (i = 1, ..., n) aus einer Stichprobe vom Umfang n seien voneinander unabhängig und identisch verteilt mit der Dichtefunktion f(t) und der Überlebensfunktion S(t). Auch die Zensierungszeiten  $C_i$  (i = 1, ..., n) werden als unabhängig und identisch verteilt angenommen. Weiterhin seien alle  $T_i^*$  und  $C_i$  voneinander unabhängig. Die Zensierungszeiten C verlaufen für jedes Individuum parallel zu den Ereigniszeiten und beginnen ebenfalls am vordefinierten Startzeitpunkt. Die exakte Überlebenszeit  $T_i^*$  eines Individuums ist dann bekannt, wenn  $T_i^* \leq C_i$  gilt. Wenn die Überlebenszeit jedoch größer als die Zensierungszeit ist, wird die Beobachtungszeit als rechts zensiert bezeichnet. Die Beobachtungszeit eines Individuums kann mithilfe des Wertepaares  $(T_i, \delta_i)$  dargestellt werden. Dabei wird  $\delta_i$  als Zensierungsindikator bezeichnet. Er zeigt an, ob ein Ereignis eingetreten ist ( $\delta_i = 1$ ) oder Zensierung vorliegt ( $\delta_i = 0$ ). Die Beobachtungszeit  $T_i$ nimmt dabei den Wert des Minimums von Überlebenszeit und Zensierungszeit eines Individuums an  $(T_i = \min(T_i^*, C_i))$ . Rechtszensierungen innerhalb einer klinischen Studie können durch folgende, unterschiedliche Gründe hervorgerufen werden. Die Patienten können:

• ein konkurrierendes Ereignis erfahren (competing risks). Unter einem konkurrierenden Ereignis versteht man, dass der Patient beispielsweise an einer Lungenentzündung verstirbt, das interessierende Ereignis jedoch der Tod durch Lungenkrebs ist.

- sich für das Ausscheiden aus der Studie entschließen (drop out). Gründe dafür können beispielsweise familiäre Probleme sein.
- nicht mehr erreichbar sein (lost to follow-up), also wenn der Patient im Untersuchungszeitraum unbekannt verzieht.

Weiterhin wird eine Studie zu einem vorher festgelegten Zeitpunkt enden. Hierbei wird von einer administrativen Zensierung gesprochen.



Abbildung 2: Ereignis- und Zensierungszeiten von Patienten in einer klinischen Studie.

Die Abbildung 2 verdeutlicht die oben angesprochenen Sachverhalte. Es sind 5 Patienten mit unterschiedlichen Beobachtungszeiten in einer klinischen Studie dargestellt. Es ist zu erkennen, dass die Patienten zu unterschiedlichen Zeitpunkten in die Studie aufgenommen wurden. Die Patienten 2 und 4 weisen eine vollständige Beobachtung der Ereigniszeit auf, während Patient 5 im Untersuchungszeitraum ereignislos blieb und seine Ereigniszeit am Ende der Studie zensiert wurde. Die Zensierungen von Patient 1 und 3 haben jedoch andere Ursachen, wie lost to follow up, drop out oder ein konkurrierendes Ereignis. Alle Patienten, ob unvollständig oder vollständig beobachtet, liefern wichtige Informationen und die Lebensdaueranalyse ist in der Lage diese vollständig für die Auswertung zu nutzen.

Es ist möglich die Rechtszensierung in unterschiedliche Typen einzuteilen. Nach Klein und Moeschberger [7] lautet diese Unterteilung wie folgt:

- Die "Typ I"-Rechtszensierung steht dafür, dass ein Ereignis nur bis zu einem vorher festgelegten Zeitpunkt beobachtet werden kann (administrative Zensierung). Demzufolge besitzt jedes Individuum, sobald es in die Studie eingeschlossen wird, eine feste Zensierungszeit.
- Die "Typ II"-Rechtszensierung tritt dann auf, wenn vorher festgelegt wird, wie viele Ereignisse in der Stichprobe beobachtet werden sollen. Beispielsweise werden bei einer einarmigen klinischen Studie<sup>1</sup> 45 Patienten rekrutiert und die Studie solange durchgeführt, bis 10 Patienten eine bestimmte Nebenwirkung erfahren haben [7].
- Die "Typ III"-Rechtszensierung wird auch Zensierung durch konkurrierende Risiken genannt. Das Ereignis von Interesse kann für die Patienten nicht beobachtet werden, welche ein anderes Ereignis erfahren haben. Diese Patienten sind zu diesem Zeitpunkt rechtszensiert.

#### 1.3 Parametrische Modelle

In der Lebensdaueranalyse gibt es einige typische, oft verwendete Verteilungen. Prinzipiell sind alle nicht-negativen Verteilungen möglich, um Ereigniszeiten zu beschreiben<sup>2</sup>. Die im Folgenden dargestellten Lebensdauerverteilungen sind alle stetig. Betrachtet werden hier die Exponential-, die Gompertz- und die piecewise-constant-Verteilung.

#### 1.3.1 Exponentialverteilung

Die Exponentialverteilung spielt eine zentrale Rolle in der Lebensdaueranalyse. Sie ist ein Spezialfall vieler anderer Lebensdauerverteilungen. Diese Verteilung besitzt nur den Parameter  $\lambda$  ( $T \sim Exp(\lambda)$ ). Der Parameter  $\lambda$  kann alle positiven Werte von 0 bis  $\infty$ annehmen ( $\lambda > 0$ ), wobei die Verteilung mit  $\lambda = 1$  als Standardexponentialverteilung bezeichnet wird. Unter der Exponentialverteilung gilt:

Dichtefunktion	$f(t) = \lambda e^{-\lambda t}$
$\ddot{\mathrm{U}}\mathrm{berlebens funktion}$	$S(t) = e^{-\lambda t}$
Hazardfunktion	$\mu(t) = \lambda$
kumulative Hazardfunktion	$M(t) = \lambda t$

<sup>&</sup>lt;sup>1</sup>Einarmige klinische Studie: klinische Studie, welche keine Vergleichsgruppe beinhaltet.

<sup>&</sup>lt;sup>2</sup>Nicht-negative Verteilungen werden hier verwendet, da die Zeit t eine nicht-negative Zufallsvariable ist (siehe Abschnitt 1.1).

Der Erwartungswert der exponentialverteilten Ereigniszeiten ist  $\mathbf{E}T = 1/\lambda$  und die Varianz  $\mathbf{V}(T) = 1/\lambda^2$ .

Dieses Modell ist wenig flexibel, da es nur einen Parameter besitzt. Obwohl es eher selten ist, dass reale Ereigniszeiten dieser einparametrischen Verteilung folgen, verdeutlicht sie dennoch die wichtigsten Aspekte vieler typischer Lebensdauerverteilungen.

#### 1.3.2 Gompertzverteilung

Eine Zufallsvariable T folgt einer Gompertzverteilung  $(T \sim G(\lambda, \varphi))$  mit den Parametern  $\lambda$  ( $\lambda > 0$ ) und  $\varphi$  ( $\varphi \in (\infty, -\infty)$ ), wenn folgendes gilt:

$f(t) = \lambda e^{\varphi t} e^{-\frac{\lambda}{\varphi}(e^{\varphi t} - 1)}$
$S(t) = e^{-\frac{\lambda}{\varphi}(e^{\varphi t} - 1)}$
$\mu(t) = \lambda e^{\varphi t}$
$M(t) = \frac{\lambda}{\varphi} (e^{\varphi t} - 1)$

Die Hazardfunktion steigt im Zeitverlauf an, beginnend zum Zeitpunkt 0 bei dem Wert  $\lambda$ . Für  $\varphi < 0$  fällt die Hazardfunktion ab.



Abbildung 3: Gompertzverteilte Hazardfunktionen mit unterschiedlichen Parametern.

#### 1.3.3 Piecewise-constant Verteilung

Diese Verteilung stellt einen einfachen Weg dar um  $\mu(t)$  flexibel zu modellieren. Dabei werden Zeitintervalle definiert, in denen die Hazardfunktion als konstant angenommen wird. Die Intervallgrenzen  $a_k$  (k = 0, ..., m) liegen im Bereich  $0 = a_0 < a_1 < ... < a_m$ , wobei  $a_m$  größer als die maximal gemessene Ereigniszeit ist [4]. Dies bedeutet, sie kann theoretisch den Wert  $\infty$  annehmen. Die Intervallgrenzen  $a_1, ..., a_{m-1}$  werden im Folgenden als Sprungpunkte bezeichnet, da es in diesen Zeitpunkten zu Änderungen der Hazardfunktion kommen kann. Es sei  $a_{k-1} \leq t < a_k$  mit j = 1, ..., m, damit die nachstehenden Funktionen wie folgt definiert sind:

Dichtefunktion	$f(t) = \lambda_k e^{-\sum_{q=1}^{k-1} \lambda_k (a_q - a_{q-1}) - \lambda_k (t - a_{k-1})}$
Überlebensfunktion	$S(t) = e^{-\sum_{q=1}^{k-1} \lambda_k (a_q - a_{q-1}) - \lambda_j (t - a_{k-1})}$
Hazardfunktion	$\mu(t) = \lambda_k$
kumulative Hazardfunktion	$M(t) = \sum_{q=1}^{k-1} \lambda_k (a_q - a_{q-1}) + \lambda_k (t - a_{k-1})$

Der Nachteil dieser Verteilung liegt darin, dass die Hazardfunktion an den Sprungpunkten  $a_1, \ldots, a_{m-1}$  nicht stetig ist. Werden jedoch geeignete Sprungpunkte gewählt, können beliebige Formen von  $\mu(t)$  approximiert werden [8]. Damit stellt die piecewiseconstant Hazardfunktion ein geeignetes Bindeglied zwischen parametrischen und nicht parametrischen Modellen dar.

Eine interessante Erweiterung ist die lineare piecewise Funktion. Indem ein Parameter hinzugefügt wird, entfallen die Sprünge am Ende der Intervalle. Diese jetzt stetige Funktion ist jedoch schwieriger anzupassen [4]. Weitere Erweiterungen in diese Richtung umfassen Spline-Funktionen.



Abbildung 4: Piecewise-constant verteilte Hazardfunktionen mit unterschiedlichen Parametern.

Liegen Informationen zur Gestalt der Hazardfunktion vor, so können diese für die Auswahl eines geeigneten parametrischen Lebensdauermodells genutzt werden. Können aufgrund der Datenlage keine Annahmen über die Verteilung getroffen werden, müssen nicht-parametrische Verfahren angewendet werden. Eine der gebräuchlichsten Methoden zur Schätzung der Survivalfunktion bei zensierten Daten ist der sogenannte Kaplan-Meier Schätzer [9].

#### 1.4 Kaplan-Meier Schätzer

Eine Möglichkeit, die Überlebensfunktion S(t) nichtparametrisch zu schätzen, ist der Kaplan-Meier Schätzer [9]. Die Zeitachse wird durch die unterschiedlichen Ereigniszeiten in Intervalle unterteilt. Für jedes Intervall wird die empirische Überlebensfunktion  $\hat{S}(t)$  (Kaplan-Meier Schätzer) als ein Produkt von bedingten Wahrscheinlichkeiten gebildet. Daraus resultiert eine weitere Bezeichnung des Schätzers als "product-limitestimator" (PL-Schätzer).

Bei zensierten Beobachtungen gilt ein Patient bis zum Eintreten des nächsten Ereignisses nach der Zensur als in der Studie enthalten. Beginnt dann das nächste Zeitintervall, so wird er aus der Beobachtung ausgeschlossen. Während in der Orginalarbeit von Kaplan und Meier [9] die Intervalle als variabel wählbar betrachtet wurden, hat sich in den folgenden Jahren eine möglichst kleine Intervalleinteilung durchgesetzt. Mittlerweile werden die Intervalle für den PL-Schätzer so definiert, dass jedes Ereignis eine Intervallgrenze beschreibt. Ausgehend von dieser Intervalldefinition hat der Kaplan-Meier Schätzer die Gestalt

$$\hat{S}(t) = \prod_{i:T_{(i)} < t} \left( 1 - \frac{\delta_{(i)}}{n - i + 1} \right).$$

Der Definition von Abschnitt 1.2 (Rechtszensierung) folgend, wird jede Beobachtung durch ein Wertepaar  $(T_i, \delta_i)$  charakterisiert, wobei  $T_i = min\{T_i^*, C_i\}$  und  $\Delta = 1(T_i^* \leq C_i)$  sind. Dabei sind  $T_{(1)} < T_{(2)} < ... < T_{(n)}$  die geordneten Beobachtungszeiten. Mit Hilfe des Kaplan-Meier Schätzers ist es möglich, sowohl die mediane Überlebenszeit, als auch die Überlebenswahrscheinlichkeit zu einem beliebigen Zeitpunkt darzustellen.

#### 1.5 Proportionales Hazards Modell nach Cox

Der Kaplan-Meier Schätzer erlaubt lediglich die Analyse von kategorialen Kovariablen mit wenigen Kategorien. Daher ist oft ein Regressionsmodell von Interesse, das sowohl mit kategorialen wie auch mit stetigen Variablen umgehen kann. Ein proportionales Hazards Modell beschreibt die Hazardfunktion als Produkt einer Basishazardfunktion und einer nicht-negativen Funktion. Diese nicht-negative Funktion modelliert die Abhängigkeit der Hazardfunktion von den Kovariablen [10]. In diesem Modell ist die Hazardfunktion zum Zeitpunkt t des i-ten Individuums:

$$\mu(t, x_i) = \mu_0(t) \exp\left(\beta' x_i\right). \tag{1}$$

Dabei ist  $x'_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$  der Kovariablenvektor, wobei die Kovariablen hier zeitunabhängig sein sollen. Die dazugehörigen Regressionskoeffizienten werden durch den Vektor  $\beta = (\beta_1, \beta_2, \ldots, \beta_p)$  dargestellt. Die Basishazardfunktion  $\mu_0$  ist zeitabhängig und für alle Individuen gleich.

Es gibt zwei Möglichkeiten ein solches Überlebenszeitmodell anzupassen. Zum einen kann ein parametrischer Ansatz gewählt werden, bei dem die Basishazardfunktion z.B. einer Weibull- oder piecewise-constant Verteilung folgt (Abschnitt 1.6). Zum Anderen ist es möglich einen semiparametrischen Modellansatz zu wählen. Ein solches Modell wurde 1972 von Cox publiziert [1]. Umgangssprachlich wird es oft als Cox-Modell be-zeichnet.

Der einfachste Fall beinhaltet eine einzelne binäre Kovariable X mit den Ausprägungen  $x_1 = 1$  bzw.  $x_2 = 0$ , z. B. in Abhängigkeit von der Gruppenzugehörigkeit.

$$\frac{\mu(t|x_1)}{\mu(t|x_2)} = \frac{\mu_0(t)\exp(\beta x_1)}{\mu_0(t)\exp(\beta x_2)} = \exp[\beta(x_1 - x_2)] = \exp[\beta]$$
(2)

Dieses Modell gilt nur unter der Annahme, dass die Ausfallraten von zwei Individuen proportional zueinander und damit die Hazardquotienten von Gleichung (2) nicht zeitabhängig sind. Diese Annahme bietet einen bequemen Weg um den Effekt einer Kovariablen auf das Überleben in einer Zahl zusammenzufassen. Bei der Einführung einer logarithmischen Skalierung erhalten wir aus Gleichung (2):

$$\ln \mu(t|x_1) - \ln \mu(t|x_2) = \beta(x_1 - x_2).$$

Das Cox-Modell beinhaltet den Hazardquotienten (2), welcher nur vom Regressionskoeffizienten  $\beta$  abhängt. Zur Schätzung von  $\beta$  wird der partielle Likelihood [1] maximiert, wobei  $\mu_0(t)$  unspezifiziert bleibt. Die Flexibilität des Cox-Modells ist durch die nicht-parametrische Basishazardfunktion bestimmt.

#### **1.6** Poisson-Regressionsmodell

Das Cox Modell ist in den letzten Jahrzehnten zu einem der wichtigsten Modelle in der Überlebenszeitanalyse geworden. Das in diesem Abschnitt besprochene Poisson-Regressionsmodell ist als Alternative zum Proportionalen Hazardsmodell (PH Modell) von Cox zu sehen. Die Gleichungen (1) (Cox-Modell) und (3) (Poisson-Regressionsmodell) unterscheiden sich bei oberflächlicher Betrachtung nicht voneinander. Der Kontrast beider Modelle besteht darin, dass die Basishazardfunktion bei Cox nicht spezifiziert wird. Dahingegen ist das Poisson-Regressionsmodell ein Proportionales Hazardmodell mit einer piecewise-constant Basishazardfunktion.

$$\mu(t|x_i) = \mu_0(t) \exp\left(\beta' x_i\right) \tag{3}$$

Bei der Poisson-Regression wird der Logarithmus der Ereignisrate als lineare Funktion der Einflussgrößen dargestellt. Jedoch folgt die Basishazardfunktion  $\mu_0(t) = \mu_{0k}$ , mit  $t_{k-1} \leq t < t_k$  einer piecewise-constant Verteilung. Der theoretische Beobachtungszeitraum verläuft von 0 bis  $\infty$ . Wie im Cox-Modell ist der Kovariablenvektor  $x_i$  und  $\beta'$ beschreibt den Vektor der Regressionskoeffizienten.

Doch warum spricht man bei einem piecewise-constant PH-Modell von einem Poission Regressionsmodell, obwohl keine Poissonverteilung im Modell enthalten ist? In diesem Modellansatz nutzt man die Tatsache, dass die Likelihoodfunktion eines piecewiceconstant Proportionalen Hazardsmodell proportional zur Likelihoodfunktion von poissonverteilten Zähldaten ist [6]. Um dies zu zeigen, wird der folgende einfache Fall betrachtet.



Abbildung 5: Baslinehazard approximiert durch eine piecewise-constant Verteilung bestehend aus 6 Pieces.

Es sei eine Stichprobe von n rechtszensierten Beobachtungen mit exponentialverteilten Überlebenszeiten gegeben, was einen Spezialfall der piecewise-constant Verteilung darstellt (ein Piece über gesamten Beobachtungszeitraum). Zur Vereinfachung werden die Kovariablen nicht mit in die Betrachtung einbezogen. Ausgehend davon ergibt sich für die Likelihoodfunktion, welche aus Hazard- und Survialfunktion unter berücksichtigung des Zensierungsindikators besteht, folgende Form:

$$L(\lambda) = \prod_{i=1}^{n} \lambda^{\delta_i} e^{-\lambda t_i}$$

Daraus ergeben sich die Log-Likelihoodfunktion und deren Ableitung nach  $\lambda$ :

$$\ell(\lambda) = \sum_{i=1}^{n} \delta_i \log \lambda - \lambda \sum_{i=1}^{n} t_i$$
$$\frac{\partial \ell}{\partial \lambda} = \frac{\sum_{i=1}^{n} \delta_i}{\lambda} - \sum_{i=1}^{n} t_i.$$

Ausgehend von dieser Log-Likelihoodfunktion erhält man ein Maximum für das

piecewise-constant PH Model bei:

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} \delta_i}{\sum_{i=1}^{n} t_i} = \frac{d}{T^{\star}} \tag{4}$$

Wenn  $\delta_i$  als poissonverteilt mit einem Erwartungswert  $\psi t_i$  angenommen wird, dann ergibt sich als Likelihoodfunktion der Poissonverteilung:

$$L_p(\psi) = \prod_{i=1}^n (\psi t_i)^{\delta_i} e^{-\psi t_i}$$

Die dazugehörige Log-Likelihoodfunktion und deren Ableitung sind:

$$\ell(\psi) = \sum_{i=1}^{n} \delta_i \left( \log \psi + \log \sum_{i=1}^{n} t_i \right) - \psi \sum_{i=1}^{n} t_i$$
$$\frac{\partial \ell}{\partial \psi} = \frac{\sum_{i=1}^{n} \delta_i}{\psi} - \sum_{i=1}^{n} t_i.$$

Danach ergibt sich ein Maximum für das Poissonmodell, welches identisch ist zu Gleichung 4:

$$\hat{\psi} = \frac{\sum_{i=1}^{n} \delta_i}{\sum_{i=1}^{n} t_i} = \frac{d}{T^{\star}}.$$

Da die Likelihoodfunktionen dieser zwei verschiedenen Modelle proportional zueinander sind, ist es möglich die Poisson-Regression genauso zu verwenden wie das piecewise-constant PH Modell [6]. Obwohl es sich bei der Poisson-Regression um einen parametrischen Modellansatz handelt, bietet sie eine große Flexibilität. Man kann bei einer adäquaten Anzahl von Intervallen von einem "quasi-semiparametrischen" Modell sprechen. Modelle, die zwischen drei und zehn Intervalle mit äquidistanten Sprungstellen enthalten, sind für die meisten praktischen Anwendungen flexibel genug [11].

#### 1.7 Frailty-Modelle

In Abschnitt 1.5 wurde das proportionale Hazards Modell nach Cox (1) eingeführt. Mit Hilfe dieses Modells können die Einflüsse von verschiedenen Kovariablen auf die Ereigniszeit untersucht werden. Das Cox-Modell erfordert zwei wichtige Annahmen. Erstens müssen die Ereigniszeiten unabhängig voneinander sein, zweitens müssen die Hazardraten proportional zueinander sein. Diese Annahmen können jedoch in verschiedenen Situationen verletzt sein (z. B. bei gruppierten Daten).

In der Medizin treten gruppierte Daten beispielsweise in multizentrischen Studien auf. Dabei ist jedes Studienzentrum ein Cluster. Bei der Betrachtung von Zwillingsdaten ist ein Zwillingspaar ebenfalls ein Cluster. Aber auch geografische Aspekte können zur Clusterbildung führen, zum Beispiel Stadtviertel oder Länder. Es gibt einige Modellansätze, welche Gruppierungen von Beobachtungen berücksichtigen. Das können beispielsweise: Copula-Modelle, marginale Modelle, Verallgemeinerte lineare gemischte Modelle und Frailty-Modelle sein. Copula-Modelle können verwendet werden um kleine Gruppen mit identischer Größe zu analysieren. Die Copula ist eine Funktion, welche die Verbindung zwischen der marginalen Überlebensfunktion der einzelnen Individuen im Cluster (individuell) und deren gemeinsamer Überlebensfunktion (gruppenspezifisch) erklärt [3], [12], [13]. Bei den marginalen Modellen wird zuerst die Korrelation innerhalb der Cluster ignoriert, um eine Schätzung der festen Effekte zu erhalten. Anschließend werden durch einen Sandwich-Schätzer die Standardfehler der festen Effekte adjustiert [12], [14]. Für den Ansatz eines verallgemeinerten linearen gemischten Modells (Generalized Linear Mixed Model - GLMM) kann man eine piecewise-constant verteilte Basishazardfunktion (Parametrische Modellierung) mit deren log link Funktion und spezifischem Offset benutzen. Bei diesem Modell wird der Clustereffekt mittels eines zufälligen Effektes modelliert [10], [15].

Eine weitere Möglichkeit stellt das Frailty-Modell dar, welches als multiplikatives Hazards-Modell verstanden werden kann. Betrachtet man den Sonderfall, dass jeder Patient einen Cluster darstellt, handelt es sich um ein univariates Frailty-Modell. Dabei liegt der Fokus darauf, die unbeobachtete Heterogenität in der Studie zu modellieren. Diese Heterogenität entsteht dadurch, dass beeinflussende Kovariaten entweder nicht erfasst wurden oder nicht erfasst werden konnten. Ein Frailty-Modell kann aber auch Abhängigkeiten zwischen Ereigniszeiten innerhalb eines Clusters modellieren [3]. Ein solches Modell bezeichnet man als shared Frailty-Modell.

Was genau verbirgt sich hinter dem Begriff "Frailty"? Frailty bedeutet Anfälligkeit und wird auf unterschiedlichste Weise definiert. Beispielsweise versteht man in der Gerontologie die Frailty als körperliche Anfälligkeit, also ein Maß der Einschränkung körperlicher Funktionen [16]. Zu beachten ist hier, dass keine spezifische Krankheit untersucht wird, sondern die Frailty als eine Art Syndrom verstanden wird [17]. Das Augenmerk liegt darauf, Kenngrößen für die Frailty in Form von Scores zu erarbeiten, da davon ausgegangen wird, dass das Morbiditäts- und Mortalitätsrisiko einer Person mit wachsender Anfälligkeit steigt [17]. Weiterhin geht man in der Gerontologie davon aus, dass die Frailty zeitabhängig ist. Dahingegen stellt in der Lebensdaueranalyse die Frailty einen zufälligen, unbeobachteten Effekt dar, der über die Zeit konstant ist. Wie bereits erwähnt, kann dabei unterschieden werden, ob es sich um eine patientenspezifische oder eine clusterspezifische Frailty handelt.

#### **1.8 Shared Frailty-Modell**

Das shared Frailty-Modell kann auch als ein Mischungsmodell bezeichnet werden [4], da das gemeinsame Risiko als zufällig angenommen wird. Hierbei ist der "Mischungsterm" die Frailty. Eine der zentralen Annahmen des Modells ist, dass die Ereigniszeiten bedingt unabhängig sind. Dies gilt unter der Bedingung, dass die Frailty gegeben ist. Der Frailtyterm wird als konstanter, für alle Individuen in einem Cluster identischer, zufälliger Effekt verstanden. Daher leitet sich auch der Name des Modells ab ("shared" zu Deutsch "geteilt", "gemeinsam haben"). Als erstes wurde dieser Modellansatz von Clayton [5] zur Analyse von Ereigniszeiten von Vätern und deren Söhnen verwendet.

Das shared Frailty-Modell unterscheidet sich von dem in Abschnitt 1.6 besprochenen Cox-Modell durch den zusätzlichen Frailtyterm und stellt damit eine Erweiterung dar. Ausgegangen wird von einem Datensatz mit n Clustern (z.B. Studienzentren). In jedem Cluster befinden sich  $n_j$  Individuen, die eine clusterspezifische Frailty  $Z_j$  (j = 1, ..., l)aufweisen. Die Frailty ist für jeden Patienten i  $(i = 1, ..., l_j)$  im j-ten Cluster identisch. Damit ergibt sich eine Hazardfunktion des i-ten Patienten im j-ten Cluster zum Zeitpunkt t von:

$$\mu(t \mid Z_j, X_{ij}) = Z_j \mu_0(t) e^{\beta' X_{ij}}.$$
(5)

Die Hazardfunktion (5) setzt sich zusammen aus dem Kovariablenvektor  $X_{ij} = (X_{ij1}, \ldots, X_{ijp})$ , dem dazugehörigen Vektor der Regressionsparameter  $\beta = (\beta_1, \ldots, \beta_p)$  und der Frailtyvariable  $Z_j$ . Die Basishazardfunktion  $\mu_0(t)$ , welche für jedes Individuum identisch sei, kann entweder als unbekannt (semiparametrisches Modell) angenommen werden oder einer speziellen Verteilung folgen (parametrisches Modell). Diese Verteilung ist abhängig von einem endlichdimensionalen unbekannten Parameter (z. B. Weibull, Gompertz, Piecewice Constant). In dieser Arbeit werden vor allem semiparametrische

shared Frailty-Modelle betrachtet.

Für das semiparametrische shared Frailty-Modell sind herkömmliche Maximum-Likelihood-Methoden zur Parameterschätzung nicht anwendbar [3]. Stattdessen können Verfahren wie der Penalised-Partial-Likelihood (PPL) Ansatz, der Expectation-Maximization (EM) Algorithmus oder Markov Chain Monte Carlo (MCMC) Methoden verwendet werden.

Um ein shared Frailty-Modell zu analysieren, muss die Verteilung der Frailtyvariable spezifiziert werden. Besonders häufig werden die Gamma- und die Log-normal Verteilung verwendet. Die Verteilung der Frailty impliziert, in welchem Maß sich die Abhängigkeit zwischen den Ereigniszeiten in den Clustern im Lauf der Zeit verändert.

In einem shared Frailty-Modell bezieht sich der Hazardquotient auf den Vergleich zwischen Individuen in einem Cluster [3]. Die Varianz der Frailty wird dabei als ein Maß für die Assoziation zwischen den Ereigniszeiten interpretiert. Weiterhin muss bedacht werden, dass bei einem shared Frailty Ansatz starke Annahmen an das Modell gestellt werden (symmetrische Korrelationsstruktur in den Clustern).

#### 1.8.1 Shared Gamma Frailty-Modell

Die Gamma-Verteilung  $\Gamma(k, \lambda)$  mit  $(\lambda > 0)$  wird häufig als Frailtyverteilung verwendet [3] [4] [19]. Grund dafür sind ihre mathematischen Eigenschaften. Besonders hervorzuheben ist dabei die einfache Form der Laplace-Transformierten. Die Gamma-Verteilung ist eine flexible Verteilung, deren Dichte unterschiedliche Formen für unterschiedliche Werte von k annimmt. Im Fall k=1 ist sie mit der Exponentialverteilung identisch. Für k > 1 ändert sich die Form zu einer Glockenkurve, ähnlich wie bei der Normalverteilung (Abbildung 6). Weiterhin modelliert sie eine stärkere späte Abhängigkeit zwischen den Ereigniszeiten in den Clustern [12].

Trotz dieser mathematischen Vorteile gibt es keine biologischen Gründe, welche die Gamma-Verteilung zu einer bevorzugten Frailtyverteilung machen. Aus Gründen der Identifizierbarkeit wird  $k = \lambda$  angenommen. Damit wird aus der zweiparametrigen Gamma-Verteilung eine einparametrige Verteilung  $\Gamma(\frac{1}{\sigma^2}, \frac{1}{\sigma^2})$  mit  $k = \lambda = \frac{1}{\sigma^2}$ . Daraus ergeben sich der Erwartungswert  $\mathbf{E}Z = 1$  und die Varianz  $\mathbf{V}(Z) = \sigma^2$  für die Frailtyvariable.



Abbildung 6: Dichtefunktion der Gamma-Verteilung mit  $\lambda = 1$  und k = 0,5; 1 und 2.

Die Likelihoodfunktion des shared Gamma Frailty-Modells wird wie folgt definiert:

$$L(\beta, \sigma^2 | z) = \prod_{j=1}^l \int_0^\infty \prod_{i=1}^{l_i} \left( z_j \mu_0(t_{ij}) e^{\beta' x_{ij}} \right)^{\delta_{ij}} e^{-z_j M_0(t_{ij}) e^{\beta' x_{ij}}} f\left( z_j, \sigma^2 \right) dz_j.$$
(6)

Die Likelihoodfunktion besteht aus zwei Teilen. Teil 1 stellt die Beobachtungszeiten in Abhängigkeit der Frailty dar, während Teil 2 die Dichtefunktion der Frailtyvariable abbildet.

#### 1.8.2 Shared Log-normales Frailty-Modell

Eine weitere wichtige Frailtyverteilung ist die log-Normal-Verteilung. Die Popularität dieser Verteilung rührt hauptsächlich von ihrer Verbindung mit den verallgemeinerten linearen gemischten Modellen her. Bei diesen ist die Standardannahme, dass der zufällige Effekt im linearen Prädiktor normalverteilt ist  $W \sim N(0, s^2)$ . Weiterhin soll die Frailty durch  $Z = e^W$  definiert sein. Für diesen Fall sind Erwartungswert und Varianz der Frailty Z Funktionen des Parameters  $s^2$  [20].

$$\mathbf{E}Z = e^{\frac{s^2}{2}}$$
  
 $\mathbf{V}(Z) = e^{s^2}(e^{s^2} - 1).$ 



Abbildung 7: Dichtefunktionen der log-Normalen-Verteilung mit  $s^2 = 0,1; 1$  und 4.

Im shared Gamma Frailty-Modell bezeichnet der Parameter  $\sigma^2$  die Varianz der Frailty Z, wohingegen im log-Normal Modell  $s^2$  die Varianz des zufälligen Effektes  $W = \ln Z$ angibt. Diese beiden Ausdrücke können nicht direkt miteinander verglichen werden. Weiterhin ist im log-Normal Modell der Erwartungswert der Frailty  $Z \neq 1$ , obwohl der Erwartungswert des zufälligen Effektes W = 0 ist. Das shared Log-normal Frailty-Modell hat die Form:

$$\mu(t|\mathbf{X}_{ij}, W_j) = \mu_0(t)e^{\boldsymbol{\beta}'\mathbf{X}_{ij}+W_j} \quad \text{mit} \quad W_j \sim N(0, s^2).$$
(7)

Die eben beschriebenen Modelle können sowohl parametrisch als auch semiparametrisch umgesetzt werden. Semiparametrische Modelle benötigen oft einen großen Stichprobenumfang, um gute Schätzungen zu liefern [21], da  $\mu_0(t)$  nicht spezifiziert wird. Zur Zeit sind nur wenige semiparametrische shared Frailty-Modelle in gängiger Software implementiert (SAS und R). Im Gegensatz dazu ist die Implementierung eines parametrischen Modells einfacher. Bereits Datensätze mit geringem Stichprobenumfang können durch diese ausgewertet werden. Der Nachteil dieser Modelle besteht darin, sich auf eine geeignete Basishazard  $\mu_0(t)$  festlegen zu müssen. Eine falsche Wahl kann zur Verzerrung der Schätzung führen. Um qualitativ gute Schätzungen bei zensierten, geclusterten Lebensdauerdaten zu erhalten, wäre ein quasi-semiparametrischer Ansatz (Mischung zwischen parametrisch und semiparametrisch) ideal.

#### 1.9 Gemischtes Poission-Regressionsmodell

Als eine solche Kombination könnte man das gemischte Poisson-Regressionsmodell verstehen. Dieses Modell ist als Bindeglied zwischen parametrischen und semiparametrischen Modellen anzusehen, obwohl es theoretisch zu den parametrischen Modellen zählt. Wenn die Anzahl an Intervallen groß genug ist, kann die Basishazardfunktion sehr flexibel gestaltet werden, wodurch es ein nahezu semiparametrisches Modell ist. Wie bereits in Abschnitt 1.6 beschrieben, kann man die Poisson-Regression genauso verwenden wie das piecewise-constant PH Modell [6], [22], [23]. Diesen Ansatz führten Feng et. al [15] fort. Sie zeigten, dass ein Frailty-Modell mit einem log-Normal Frailtyterm als ein verallgemeinertes lineares gemischtes Modell mit einem binären Endpunkt und einer Poissonlikelihood mit spezifischem Offset interpretiert und geschätzt werden kann.

Im Folgenden wird diese Äquivalenz für den Fall eines univariaten Modells gezeigt. Es sei eine Stichprobe von n rechtszensierten Beobachtungen gegeben. Der Beobachtungszeitraum werde in m Intervalle unterteilt, hier Pieces genannt. Die vollständige Beobachtungszeit eines Individuums i im Piece  $\Omega_k$  ist definiert als  $t_{ik}$ . Weiterhin wird angenommen, dass in jedem Piece die Basishazard konstant ist. Außerdem sei  $\delta_{ik}$  der Zensierungsindikator des Individuums i in  $\Omega_k$ , mit  $\delta_{ik} = 1$ , wenn das Individuum i im k-ten Piece das Ereignis erfährt und  $\delta_{ik} = 0$ , wenn dies nicht der Fall sein sollte.

$$\mu_0(t) = \mu_k, \qquad t \in \Omega_k = [t'_{k-1}, t'_k), \qquad (k = 1, \dots, m)$$

Weiterhin gilt  $t_{ik} = 0$  und  $\delta_{ik} = 0$  für alle Pieces, die auf ein Piece folgen, in welchem die Beobachtungszeit durch ein Ereignis oder eine Zensierung beendet wurde. Wenn jedes  $t_k$  eine Ausfallzeit repräsentiert und m die Gesamtanzahl der unterschiedlichen Ausfallzeiten ist, dann stellt es ähnlich dem Cox-Modell einen nicht parametrischen Ansatz zur Modellierung der Basishazardfunktion dar. Die bedingte Likelihoodfunktion (bedingt bezüglich  $\mathbf{W} = (W_1, \ldots, W_n)$ ) hat die Form [24]:

$$L(\mu_1, \dots, \mu_m, \boldsymbol{\beta} | \mathbf{W}) = \prod_{i=1}^n \prod_{k=1}^m \left( \mu_k \exp(\boldsymbol{\beta}' \mathbf{X}_i + W_i) \right)^{\delta_{ik}} \exp\left( -\mu_k t_{ik} \exp(\boldsymbol{\beta}' \mathbf{X}_{ik} + W_i) \right)$$
$$\propto \prod_{i=1}^n \prod_{k=1}^m \left( \mu_k t_{ik} \exp(\boldsymbol{\beta}' \mathbf{X}_i + W_i) \right)^{\delta_{ik}} \exp\left( -\mu_k t_{ik} \exp(\boldsymbol{\beta}' \mathbf{X}_{ik} + W_i) \right)$$

Diese Funktion ist proportional zu der bedingten Likelihoodfunktion eines Poisson-Regressionsmodells mit Mittelwert  $\mu_k t_{ik} \exp(\boldsymbol{\beta}' \mathbf{X}_i + W_i)$  und einem spezifischen Offset  $\ln(t_{ik})$ . Anschließend kann der zufällige Effekt ausintegriert werden [15]:

$$L(\mu_1,\ldots,\mu_m,\boldsymbol{\beta}) = \prod_{i=1}^n \prod_{k=1}^m \int_{-\infty}^\infty \left( \mu_k t_{ik} \exp(\boldsymbol{\beta}' \mathbf{X}_i + w) \right)^{\delta_{ik}} \exp\left(-\mu_k t_{ik} \exp(\boldsymbol{\beta}' \mathbf{X}_{ik} + w)\right) f(w) dw$$

Diese Likelihoodfunktion ist proportional zur Likelihoodfunktion eines log-normalen shared Frailty-Modells (Abschnitt 1.8.2). Dies führt dazu, dass geclusterte Lebensdauerdaten mittels eines verallgemeinerten gemischten Modells analysiert werden können. Damit stehen zusätzliche Prozeduren bei der Auswertung zur Verfügung.

### 2 Zielstellung

Um gruppierte Lebensdauerdaten mit vermutetem Clustereffekt zu analysieren gibt es eine Vielzahl von Modellansätzen. Fällt die Wahl auf ein shared Frailty-Modell steht dem Nutzer, nach Auswahl einer geeigneten Frailtyverteilung, eine Vielzahl von Funktionen zur Verfügung. Die meisten Funktionen sind frei erhältlich und deren Güte schwer einzuschätzen. Von Interesse sind in der vorliegenden Arbeit quasi-semiparametrische beziehungsweise semiparametrische Funktionen, welche für die Statistik-Pakete SAS und R entwickelt wurden. Um Vorteile sowie Limitationen von Funktionen abschätzen zu können, sind Simulationsstudien erforderlich. Simulationsstudien bieten die Möglichkeit etablierte Gütekriterien zur Bewertung von Funktionen zu nutzen.

Die Hauptfrage der Arbeit lautet: Welche Funktionen liefern die besten Schätzungen bezüglich fester und zufälliger Effekte?

Von Interesse sind dabei die Abweichung der Punktschätzer vom wahren Wert, der mittlere quadratische Fehler der Schätzungen, die Unterschiede zwischen geschätzten und empirischen Standardfehlern sowie die Überdeckungsraten der 95% Konfidenzintervalle. Neben der Güte der Schätzalgorithmen soll auch deren Robustheit betrachtet werden. Daraus soll eine Empfehlung für den Nutzer bezüglich der Auswertungsmöglichkeiten bei shared Frailty-Modellen abgeleitet werden.

Ist diese Frage beantwortet, soll mit einem Beispieldatensatz (Halluca-Studie) die Anwendbarkeit der Funktionen illustriert werden.

### 3 Material und Methoden

### 3.1 R und SAS Funktionen zur Analyse von gruppierten Lebensdauerdaten

Zunächst werden die R<sup>3</sup>-Funktionen erläutert. Zwei dieser Funktionen wurden von Therenau entwickelt [21]. Implementiert ist der shared Frailty-Modellansatz zum einen in einer Basis-Library<sup>4</sup> und zum anderen in einer speziell für geclusterte Lebensdauerdaten entwickelten Library<sup>5</sup>. Weiterhin ist eine Library von Rondeau erhältlich, welche eine Funktion beinhaltet die mittels quasi-semiparametrischen Modellansatzes ein shared Frailty-Modell analysiert. Im Anschluss daran wird die Software SAS<sup>6</sup> betrachtet. SAS bietet nur einige wenige Makros<sup>7</sup>/Prozeduren<sup>8</sup>, welche shared Frailty-Modelle auswerten können. Zwei der hier betrachteten Makros wurden von Vu [25], [26], [27] entwickelt. Die Prozedur PHREG wurde unter SAS 9.3 für Frailty-Modelle erweitert und der in Abschnitt 1.9 besprochene Ansatz des gemischten Poisson-Regressionsmodells wurde im %PCFrailty Makro realisiert. Da die meisten Prozeduren/Makros/Funktionen mit der Breslow-Approximation (Methode zur Analyse von Daten mit Bindungen) arbeiten, sollte bei den Einstellungen der R-Funktionen auf diese geachtet werden. Alle hier beschriebenen Funktionen/Makros schätzen im Log-normalen Modell immer die Varianz des zufälligen Effektes  $s^2$  (7), während im Gamma Modell die Varianz der Frailtyvariable  $\sigma^2$  (5) ausgegeben wird. Einen Überblick über die R-Funktionen und SAS-Makros liefern die Tabellen 3 und 4.

Die Statistik-Softwarepakete STATA und WinBUGS enthalten ebenfalls Analyseverfahren für Frailty-Modelle. STATA bietet seit Version 7 die Möglichkeit an, sowohl Gamma als auch Log-normale Frailty-Modelle parametrisch zu schätzen. Ein semiparametrischer Ansatz ist nicht implementiert. In WinBUGS werden ausschließlich bayesianische Ansätze verwendet. In dieser Arbeit wird auf diese jedoch nicht weiter eingegangen.

 $<sup>{}^{3}</sup>$ R ist eine objektorientierte und interpretierte Programmiersprache für Datenanalyse und Grafik.

 $<sup>^4\</sup>mathrm{Paket},$  welches in der Standard installation von R enthalten ist.

<sup>&</sup>lt;sup>5</sup>Paket, welches zusätzlich in R eingebunden werden kann. Erhältlich sind diese Pakete unteranderem auf http://cran.R-project.org.

<sup>&</sup>lt;sup>6</sup>SAS ist eine herstellergebundene Softwareumgebung (lizensierte Software).

<sup>&</sup>lt;sup>7</sup>SAS Makros sind Programmcodefragmente oder Variablen, welche einmalig erstellt wurden und genutzt werden um wiederholt auftretende Fragestellungen zu bearbeiten.

<sup>&</sup>lt;sup>8</sup>SAS Prozeduren sind fest in der SAS Software implementiert. Sie werden auf Datensätze angewandt und dienen der Analyse, der statistischen Auswertung und der grafischen Darstellung.

#### 3.1.1 R-Funktionen

Die vier verfügbaren Funktionen für semiparametrische (bezeihungsweise quasi-semiparametrische) shared Frailty-Modelle sind coxph, coxme, frailtyPenal und phmm. Ausgenommen aus der vorliegenden Arbeit ist die Funktion phmm (Version: 0.7 - 4). Sie ist in der gleichnamigen Library enthalten und stammt von Donohue und Xu. Die Funktion phmm analysiert ein proportionales Hazards Modell mit normalverteilten zufälligen Effekten [28]. Jedoch beinhaltet sie kein Konvergenzkriterium. Dies macht die Einschätzung der Konvergenz des Schätzalgorithmus innerhalb einer Simulationsstudie nahezu unmöglich, weshalb die Funktion aus dieser Arbeit ausgeschlossen wurde.

Tabelle 3: Hauptmerkmale der R Funktionen.				
Funktion	coxph	coxme	frailtyPenal	
Library	$\operatorname{survival}$	coxme	frailtypack	
Version	2.36 - 14	2.2-3	2.3	
$Modell^1$	sp Shared	sp Shared	$sp^2$ Shared	
	Frailty-Modell	Frailty-Modell	Frailty-Modell	
			joint Frailty,	
			nested Frailty	
Verteilung	Log-normal,t, Gamma	Log-normal	Gamma	
$Algorithmus^3$	PPL	PPL	$PL^4$	
Zensierung	rechts, links,	rechts, links,	$\mathrm{rechts}$	
	interval	interval		
Trunkierung	ja	ja	ja	
Regulierungs-	method, eps,	eps, iter.max	n.knots, maxit,	
parameter	iter.max,	inner.iter	kappa	
	outer.max			
$SF^5$ (Varianz)	nein	nein	ja	

 $^1$  Frailty-Modelle, welche in Library enthalten sind / sp - semiparametrisch

 $^2$ quasi-semiparametrisches Modell

<sup>3</sup> PPL - Penalised Partial Likelihood

 $^4$  PL - Penalised Likelihood der Hazardfunktion mit Nutzung von Marquardt-Algorithmus und Approximation mittels Splines

 $^5$  SF (Varianz) - Standardfehler für die Varianz der Frailty/des zufälligen Effektes

#### Coxme

Die Coxme-Library beinhaltet die gleichnamige Funktion coxme. Diese analysiert proportionale Hazards Modelle mit normalverteilten zufälligen Effekten (Log-normales Frailty-Modell). Erstellt wurde sie von Therneau, der als Schätzalgorithmus die Penalized-Partial-Likelihood (PPL)-Methode verwendet [29], [30]. Der Nutzer muss für die Analyse verschiedene Argumente spezifizieren. Diese Argumente sind:

- formula: die Modellbeschreibung mit den festen Effekten und dem zufälligen Effekt.
- data: der zu analysierenden Datensatz.
- ties: die Festlegung der Methode zum Umgang mit Bindungen.

Es hat sich als nützlich erwiesen bestimmte, vom Programm voreingestellte Kontrollparameter zu verändern. Mittels coxme.control kann das Konvergenzkriterium eps von  $10^{-9}$  auf  $10^{-11}$  und die maximale Iterationsanzahl iter.max (für die abschließende Anpassung) von 20 auf 40 Schritte gesetzt werden. Ebenso kann die Iterationszahl für die innere Schleife inner.iter von 10 auf 40 erhöht werden. Diese Änderungen führen zu einer Erhöhung der Schätzpräzision, jedoch verlängert sich dadurch die Berechnungszeit.

```
control <- coxme.control(eps = 1e-09, iter.max = 20, inner.iter = 10)
coxme(formula, data, control,
    ties = c('efron', 'breslow'), random, ...)</pre>
```

#### Survival

Diese Library wurde ebenfalls von Therneau entwickelt. Eine ihrer Funktionen ist coxph. Sie analysiert proportionale Hazards Modelle mit Hilfe der PPL-Methode. Die Funktion kann mit zeitabhängigen Kovariablen, zeitabhängigen Strata [21] und anderen Modellerweiterungen umgehen. Eine dieser Erweiterungen ist das shared Frailty-Modell. Die Syntax ist vergleichbar mit der von coxme. Hier kann jedoch alternativ zu ties auch die Option methods verwendet werden. Zusätzlich muss der Nutzer die Frailtyverteilung frailty (gamma, log-normaloder log-t) und die Clustervariable x angeben. Genau wie bei der Funktion coxme sollten bestimmte Kontrollparameter (Konvergenzkriterium und maximale Iterationsanzahl) angepasst werden. Der Aufruf der Funktion lautet wie folgt:

```
control<-coxph.control(eps = 1e-09, iter.max = 20, outer.max = 10)
coxph(formula, data, control,
    ties=c('efron','breslow','exact'), method,
    frailty(x, distribution='gamma', ...) ...)</pre>
```

Sowohl coxph als auch coxme liefern keine Schätzung der Standardfehler für den Schätzer der Frailtyvarianz. Es ist jedoch möglich, durch die von Therneau und Grambsch in ihrem Buch beschriebene Methode, Profile-Likelihood Konfidenzintervalle zu erhalten [21]. Dabei werden die Werte der Profile-Likelihoodfunktion aus der Ergebnisliste genutzt. Die Angaben im Buch entsprechen dem Grundaufbau der Ergebnisliste von coxph. Für die Funktion coxme gilt es zu beachten, dass die Ergebnisliste eine andere Struktur aufweist. Die Verwendung der Profile-Likelihood Konfidenzintervalle ist für einzelne Datensätze geeignet. In der vorliegenden Arbeit wurden sie nicht auf die Simulationsdaten angewandt.

#### Frailtypack

Die Library frailtypack wurde von Rondeau und Gonzalez [31] entwickelt und 2012 erweitert [32]. Sie enthält Funktionen, welche sowohl ein proportionales Hazards Modell nach Cox, als auch shared, joint, nested und additive Frailty-Modelle analysieren. Hierzu wird ein kubischer M-Spline zur Approximierung der Basishazardfunktion genutzt [33]. Diese speziellen Splines führen dazu, dass die Library auf einem parametrischen Ansatz basiert. Genauso wie piecewise-constant Basishazardsmodelle ist dieser Ansatz sehr viel flexibler im Vergleich zu klassischen parametrischen Modellen, wie Gompertzoder Weibullmodellen, da sich die Splines jeder beliebigen Form der Basishazardfunktion anpassen können. Darum werden die in frailtypack enthaltenen Modelle hier als quasi-semiparametrische Modelle bezeichnet. Die in der Library verwendete Schätzmethode ist die Maximierung der Penalised-Loglikelihoodfunktion. Genutzt wird dafür der robuste Marquardt-Algorithmus, welcher eine Kombination aus Newton-Raphson-Algorithmus und dem Verfahren des steilsten Abstiegs darstellt. Die Library bietet die Möglichkeit, sowohl links-trunkierte als auch rechtszensierte Daten zu analysieren. Weiterhin kann sie mit rekurrenten Ereignissen und zeitabhängigen Variablen umgehen. Sie bietet eine glatte Schätzung und eine grafische Ausgabe der Hazardfunktion. Ebenso ist eine Analyse mit maximal zwei Strata möglich. Zur Analyse eines shared Gamma Frailty-Modells kann die Funktion frailtyPenal verwendet werden. Jedoch bietet frailtyPenal ebenso die Möglichkeit ein Cox-Modell, sowie nested und joined Frailty-Modelle anzupassen. Der Funktionsaufruf zur Anpassung eines shared Gamma Frailty-Modells lautet wie folgt:

Das Modell wird mittels des formular-Statement angegeben. Weiterhin werden die interessierenden Kovariaten und die Clustervariable an den formular Befehl angeschlossen. Zusätzlich wird im data-Statement der verwendete Datensatz definiert. Mittels n.knots wird die Knotenanzahl des Splines für die Approximation der Basishazardfunktion festgelegt. Diese sind in einem Bereich zwischen 4 bis 20 Knoten definiert. Um Einfluss auf die Penalised-Likelihoodfunktion zu nehmen, kann der Glättungsparameter kappa1 selbst gewählt werden. Dieser sollte immer im positiven Bereich liegen. Wenn die Funktion selbst einen Glättungsparameter zur Verfügung stellen soll, ist die Nutzung des Arguments cross.validation von Vorteil. Wird dieses Statement auf TRUE gesetzt läuft ein Kreuzvalidierungsverfahren ab, bei dem der Wert von kappa1 als Startwert verwendet wird. Damit ein shared Frailty-Modell analysiert werden kann, muss das Argument Frailty auf TRUE gesetzt werden. Weiterhin ist das RandDist-Statement per Default auf Gamma gestellt. Sollte die Funktion innerhalb der maximalen Anzahl von Iterationsschritten nicht konvergieren, ist es möglich die Schrittanzahl mittels maxit nach oben zu korrigieren.

#### 3.1.2 SAS-Makros/Prozeduren

Um semiparametrische und quasi-semiparametrische shared Frailty-Modelle mittels SAS zu analysieren, stehen dem Nutzer die Prozedur PROC PHREG sowie verschiedene Makros zur Verfügung. SAS biete seit der Version 9.3 eine Erweiterung der Prozedur PROC PHREG an. Mit dieser ist es möglich, einen zufälligen Effekt im RANDOM-Statement zu definieren. Die Makros SPGAM1 und SPLN3 wurden von Vu [25] erstellt. Weiterhin wurde das Makro %PCFrailty entwickelt, welches SAS Prozeduren sowie das %Lexis Makro von B. Carstensen [34] zur Analyse von Überlebensdauerdaten mittels gemischter Poisson-Regressionsmodelle verwendet [18]. Ein weiteres Makro zur Analyse von semiparametrischen shared Gamma Frailty-Modellen ist Gamfrail von Klein<sup>9</sup>. Dieses Makro weist jedoch eine lange Rechenintensität auf, die mit hohen Rechenzeiten verbunden ist, wodurch es für die vorliegende Arbeit nicht in Betracht gezogen werden konnte.

#### PROC PHREG

Eine neue Erweiterung in SAS 9.3 ist das RANDOM-Statement in der Prozedur PROC PHREG zur Analyse von shared Log-normalen Frailty-Modellen. Der Aufruf ist identisch zum normalen PROC PHREG Aufruf. Zu beachten ist, dass die Clustervariable im CLASS-Statement gelistet werden muss. Ebenso sollte als Option innerhalb dieses Statements PARAM zur Angabe der Referenzgruppe genutzt werden. Im MODEL-Statement werden die Varia-

 $<sup>^{9} \</sup>rm Das$  Makro ist unter http://www.mcw.edu/biostatistics/statistical resources/CollaborativeSoftware.htm zu erhalten.

Tabelle 1. Hauptmerkinale der Srib i Tozedaren/ Manios.				
Prozedur/Makro	SPGAM1	SPLN3	PROC PHREG	%PCFrailty
Version	ab 8.1	ab 8.1	9.3	ab 9.2
$Modell^1$	sp Shared	sp Shared	gemischtes Poission	sp Shared
	Frailty-Modell	Frailty-Modell	${ m Regressionsmodell^2}$	Frailty-Modell
Verteilung	Gamma	Log-normal	Log-normal	$\operatorname{normal}$
$ m Algorithmus^{3}$	EM	EM	PPL	RSPL
Zensierung	$\operatorname{rechts}$	$\operatorname{rechts}$	rechts, links	$\operatorname{rechts}$
Trunkierung	nein	nein	ja	
Regulierungs-	maxiter,	n, epsilon,	PCONV, METHOD,	PCONV,
parameter	epsilon	maxiter	INITIAL	NLOPT,
				pieces
$SF^4$ (Varianz)	ja	ja	ja	ja

Tabelle 4: Hauptmerkmale der SAS Prozeduren/Makros.

 $^1$  sp - semiparametrisch

 $^2$  quasi-semiparametrisches Modell

<sup>3</sup> PPL - Penalised Partial Likelihood, EM - Expectation-Maximization, RSPL - Restricted Pseudo Likelihood

 $^4$  SF (Varianz) - Standardfehler für die Varianz der Frailty/des zufälligen Effektes

blen festgelegt, welche als Ausfallzeit, als Zensierungsindikator und als interessierende Variablen in das Modell aufgenommen werden sollen. Eine Option des MODEL-Statements ist MAXITER. Sie dient der Angabe der maximalen Iterationsanzahl (Default 25). Um eine bessere Schätzung zu erreichen sollte sie jedoch erhöht werden. Zur Anpassung an das shared Frailty-Modell wird die PPL-Methode verwendet. Es gibt innerhalb des RANDOM-Statements verschiedene optionale Argumente, um auf die Schätzung Einfluss zu nehmen. Mittels PCONV spezifiziert man das Konvergenzkriterium der Varianzschätzung. Die Option METHOD bietet die Möglichkeit die Schätzmethode der Varianz zu ändern. Zwei Schätzmethoden stehen hierbei zur Auswahl: die Standardeinstellung REML (residual Maximum Likelihood) oder ML (Maximum Likelihood). Zusätzlich ermöglicht die Verwendung von INITIAL die Änderung des Startwertes der Varianzschätzung. Die Prozedur wird wie folgt aufgerufen:

```
PROC PHREG data =;
    class ... / param=ref desc;
    model ... / maxiter=25;
    random ... / method=REML pconv=0.0001 initial=1;
run;
```

#### SPGAM1

Dieses SAS-Makro realisiert den EM-Algorithmus für das shared Gamma Frailty-Modell. Angegeben werden muss der Name der SAS Datei (f\_in), die Ereigniszeitvariable (t), der Zensierungsindikator (cens) und der Name der Clustervariable (ic). Im x-Statement

werden die festen Kovariablen benannt. Wenn keine Kovariable in die Betrachtung einbezogen werden soll ist x leer. Zu beachten ist, dass eine kategoriale Variable in mehrere Dummy-Variablen überführt werden muss, da das Makro nur mit binären bzw. stetigen Kovariablen umgehen kann. Mittels titel wird der Titel der im Output-Fenster erscheinenden Ergebnisse benannt. Da das Konvergenzkriterium (epsilon) mit 0,0001 recht groß ist, sollte es in  $10^{-9}$  geändert werden. Ebenso sollte die maximale Anzahl der Iterationsschritte nach oben korrigiert werden (maxiter), da eine Konvergenz nach 50 Schritten oft noch nicht erreicht wird. Ein weiteres nützliches Argument ist ztheta. Dieser Wert steuert die Entscheidung, ob der Schätzer für die Frailtyvarianz NULL entspricht oder nicht. Sollte der Schätzer den im ztheta-Statement angegebenen Wert unterschreiten, wird der Schätzer als NULL angesehen. Bezüglich der Ergebnisausgaben können fünf Argumente angesprochen werden. Festgelegt werden kann, ob eine allgemeine Ausgabe erzeugt werden soll (itprint). Weiterhin kann der Iterationsverlauf ausgegeben werden (ithist). Die Ausgabe der Kovarianzmatrix kann mittels cov beeinflusst werden (Die Wahlmöglichkeiten hierbei sind: keine Ausgabe = 0, Ausgabe von Kovarianzmatirx von Frailty und festen Effekten = 1 und Ausgabe der Kovarianzmatrix der Hazard, der Frailty und der festen Effekte = 2.). Wenn in cov die Ausgabe einer Kovarianzmatrix erwünscht ist, wird im outcov-Statement der Name der Ausgabedatei angegeben. Die Benennung der Ausgabedatei der Schätzer erfolgt mittels outpar. Der Aufruf des Makros lautet wie folgt<sup>10</sup>:

#### SPLN3

SPLN3 ist ein Makro zur Berechnung von shared Log-normalen Frailty-Modellen mit Hilfe der EM-Methode. Die Syntax ist äquivalent zu der von SPGAM1, wird aber erweitert durch das zusätzliche Argument n. Dies ist eine positive ganze Zahl, welche die Genauigkeit der im EM-Algorithmus approximierten Integrale kontrolliert. Je größer der Wert des Argumentes n, desto präziser sind die Werte. Diese Integrale haben jedoch keine explizite Lösung im Log-normalen Modell.

 $<sup>^{10} \</sup>rm Das$  Makro kann unter http://www.mirrorservice.org/sites/lib.stat.cmu.edu/general/spgam1 her-untergeladen werden.

#### %PCFrailty

Dieses neu entwickelte Makro ist in drei Abschnitte aufgeteilt. In den ersten beiden Schritten wird der Originaldatensatz für die Auswertung mittels eines gemischten Poisson-Regressionsmodells vorbereitet. Dafür wird am Anfang die Prozedur LIFETEST verwendet, um alle beobachteten Ausfallzeiten zu ordnen und Intervallgrenzen festzulegen. Die Intervalle können unterschiedlich festgelegt werden. Das dafür zu verwendende Statement lautet pieces. Intervalle können sowohl äquidistant als auch in Abhängigkeit von der Ereignisanzahl innerhalb jedes Intervalls gewählt werden. Ebenso ist es möglich individuelle Intervallgrenzen festzulegen. Anschließend wird das %Lexis Makro von Carstensen 34 verwendet. Hierbei wird der Originaldatensatz mittels der Intervallgrenzen aus PROC LIFETEST aufgefaltet, wodurch eine Beobachtung mehrmals im Datensatz vorkommen kann. Je mehr Teilintervalle ein Individuum überlebt, desto häufiger wird es im aufgefalteten Datensatz gelistet. Das bedeutet jedoch auch, dass mit steigender Intervallanzahl der Datensatz immer größer wird. Dies kann sowohl zu Problemen bei der Speicherkapazität als auch zu einer längeren Auswertungsdauer führen. Im dritten Schritt wird der aufgefaltete Datensatz in die Prozedur PROC GLIMMIX eingelesen und ausgewertet. Diese Prozedur verwendet zur Analyse die Newton-Raphson Methode und liefert viele Funktionen, die das Modell flexibler machen. Zum einen kann PROC GLIMMIX Gauß-Quadratur zur Parameterschätzung nutzen, was zu exakten Maximum-Likelihood Schätzern führt. Zum anderen liefert das Objekt COVTEST unterschiedliche Optionen zur Hypothesentestung bei Varianz und Kovarianz des zufälligen Effektes. Auch die Standardfehler und Konfidenzintervalle für alle Varianzparameter werden angegeben. Kovarianzmatrizen können mittels der Option LDATA im RANDOM-Statement definiert werden. Ebenso können nicht-parametrische Funktionen für feste Effekte durch TYPE=PSPLINE, TYPE=RSMOOTH oder durch das neue EFFECT-Statement geschätzt werden. Zusätzlich besitzt GLIMMIX Optionen für die:

- Schätzung robuster Standardfehler von festen Effekten,
- Modellierung mehrerer hierarchischer Ebenen und
- Implementierung von zeitabhängigen Kovariablen.
Der Aufruf des eben beschriebenen Makros lautet:

Im Folgenden werden die hier aufgelisteten R-Funktionen, SAS Makros und Prozeduren mit dem Begriff **Funktionen** zusammengefasst.

### 3.1.3 Schätzalgorithmen

Der EM-Algorithmus wurde entwickelt, um den Partial-Likelihood Ansatz von Cox [1] zu erweitern [35]. Er benutzt die vollständige Likelihoodfunktion, welche aus den beobachteten Ereigniszeiten und der unbeobachteten Frailty hergeleitet wird. Im E-Schritt (Expectation) wird die erwartete Likelihoodfunktion der beobachteten Daten berechnet. Anschließend wird im M-Schritt (Maximisation) zur Schätzung der festen Effekte die Partielle Likelihoodfunktion hergeleitet. Dabei werden Profile-Likelihood Methoden verwendet [35]. Es gibt unterschiedliche Möglichkeiten den beschriebenen EM-Algorithmus zu erweitern. Vu [26] nutzte einen Hybridansatz, um die SAS Makros sowohl für shared Gamma als auch Log-normale Frailty-Modelle anzuwenden. Der Hybridansatz verwendet zur Schätzung der Varianzparameter die marginale Loglikelihoodfunktion. Um die festen Effekte zu schätzen wird ein Offset genutzt. Dieser Offset ist der Logarithmus der posteriori Erwartungen der Frailty [36].

Eine andere Möglichkeit zur Analyse ist der Penalised-Partial-Likelihood-Ansatz, welcher aus dem Bereich der gemischten Modelle stammt. Er stellt eine Alternative zum EM-Algorithmus dar und gilt als schneller konvergierend [10], [12], [29]. Die vollständige Likelihoodfunktion besteht aus zwei Teilen. Der erste Teil stellt die bedingte Likelihoodfunktion der beobachteten Daten gegeben die Frailty dar, während der zweite Teil der Dichte der Frailtyverteilung entspricht. In diesem Ansatz wird der zweite Teil als Strafterm angesehen. Wenn der aktuelle Wert des zufälligen Effektes  $\hat{w}$  weit von seinem Mittelwert entfernt geschätzt wird, ist der absolute Wert des Logarithmus der Dichtefunktion für den Schätzer des zufälligen Effektes groß. Der Strafterm liefert dann einen großen negativen Beitrag zu der vollständigen Loglikelihoodfunktion. Die Maximierung der Penalised-Partial (Log-) Likelihoodfunktion besteht je aus einer inneren und äußeren Schleife. In der inneren Schleife wird mittels Newton-Raphson Verfahren und der Zuhilfenahme eines vorläufigen Wertes  $s_{Start}^2$  maximiert [10]. Dabei entstehen vorläufige Werte für  $\beta$  und w (BLUPs - best linear unbiased predictors). In der äußeren Schleife wird der Restricted-Maximum-Likelihood Schätzer der Varianz  $\hat{s}^2$  durch die BLUPs ermittelt [10]. Dieser Prozess wird bis zur Konvergenz iteriert [12]. Für ein shared Gamma Frailty-Modell erhält man im PPL- sowie im EM-Algorithmus identische Ergebnisse [12], [29], [36].

In der Funktion frailtyPenal wird der eben beschriebene Ansatz verändert. Hier geht man von einer Penalised-Full Likelihoodfunktion (PL) aus [37]. Mittels der PL wird eine nichtparametrische Schätzung der Hazardfunktion durchgeführt. Dabei wird die Basishazardfunktion als glatt angenommen. Je weniger glatt  $\mu_0(t)$  ist, desto größer fällt der Strafterm aus [37].

Die Prozedur GLIMMIX nutzt einen weiteren Ansatz. Hier wird eine eingeschränkte (restricted) Pseudo-Likelihoodfunktion (RSPL) verwendet. Unter einer Pseudo-Likelihoodfunktion versteht man eine Likelihoodfunktion, welche höhergeordnete Abhängigkeiten ignoriert, das heißt sie stellt eine vereinfachte Version der vollständigen Likelihoodfunktion dar. Dabei wird davon ausgegangen, dass sowohl die vollständige Likelihoodfunktion als auch die Pseudo-Likelihoodfunktion das Maximum im selben Punkt erreichen. Der bevorzugte Maximierungsalgorithmus ist auch hier Newton-Raphson, mit dem die BLUPs ermittelt werden [38].

# 3.2 Validierungsmethoden zur quantitativen Bewertung der untersuchten Funktionen

Egal welche Modellanpassung und welche Funktion betrachtet werden, im Fokus der Auswertungen stehen immer die interessierenden Parameter. Für die in dieser Arbeit verwendeten Modelle (Abs.: 1.8.1, 1.8.2 und 1.9) sind diese Parameter unterteilt in feste und zufällige Effekte:

- $\hat{\beta}_i$  stellen die Punktschätzer der Regressionskoeffizienten dar und werden als normalverteilt angenommen. Die Parameterräume dieser Regressionskoeffizienten liegen im Bereich  $(-\infty, +\infty)$ .
- s<sup>2</sup> stellt den Punktschätzer der Varianz des zufälligen Effektes dar. Diese Schätzer werden als linksschief angenommen, da Schätzwerte unterhalb der 0 nicht definiert sind (Abbildung 8). Der Parameterraum für s<sup>2</sup> liegt im Bereich zwischen [0, +∞). Identische Aussagen gelten für den Parameterraum und den Punktschätzer der Frailtyvarianz σ<sup>2</sup>.

Bei der Betrachtung der Güte eines Punktschätzers  $\hat{p}$  ist ein Grundmaß für deren Einschätzung der Bias (Verzerrung) [39]. Als Bias  $\epsilon$  bezeichnet man die erwartete Abweichung des Punktschätzers  $\hat{p}$  von dem wahren Wert des Parameters p. In Simulationsstudien wird der Bias als Differenz zwischen dem arithmetischen Mittel der berechneten Punktschätzer und dem wahren Wert des Parameters verstanden. Angenommen wird, dass die Verteilung der Punktschätzer symmetrisch ist, da das arithmetische Mittel zur Berechnung des Bias verwendet wird. Ist diese Annahme verletzt, sollten Alternativen verwendet werden. Bei den Punktschätzern der Regressionskoeffizienten kann von einer symmetrischen Verteilung ausgegangen werden [40]. Eine asymmetrische Verteilung der Punktschätzer ist jedoch bei den Parametern  $\sigma^2$  und  $s^2$  zu beobachten [40]. Alternativ zum arithmetischen Mittel kann hier der Median verwendet werden. Das arithmetische Mittel und der Median sind annähernd identisch bei symmetrischen Verteilungen, der Median ist bei schiefen Verteilungen jedoch zu bevorzugen. Aus diesen Gründen wird in der vorliegenden Arbeit nicht der Bias, sondern die Abweichung des Median vom wahren Wert verwendet. Zur sprachlichen Vereinfachung wird die Abweichung des Medians vom wahren Wert durch den Begriff der "medianen Abweichung" ersetzt. Weiterhin wurden die dazugehörigen Standardfehler und empirischen Standardfehler der Regressionskoeffizienten sowie der Frailtyvarianz/ der Varianz des zufälligen Effektes angegeben. Um die Qualität der berechneten Standardfehler zu überprüfen, wurden 95% Überdeckungswahrscheinlichkeiten genutzt. Diese geben an, wie oft der wahre Parameter im 95%-igen Konfidenzintervall (KI) enthalten war. Zusätzlich wurde der mittlere quadratische Fehler (MQF) berechnet. Um die Ergebnisse zu veranschaulichen wurden diese mittels Boxplots in R dargestellt.



Abbildung 8: Verteilungen der Punktschätzer von  $s^2$  bei unterschiedlichen wahren Werten.

### 3.2.1 Mittlerer quadratischer Fehler

Der mittlere quadratische Fehler (MQF) ist ein Maß für die Genauigkeit eines Schätzers [39]. Von Vorteil ist, dass er den Bias (Verzerrung) und die Variabilität der Schätzer miteinander verbindet.

$$MQF(\hat{p}) = (\overline{\hat{p}} - p)^2 + (SF(\hat{p}))^2$$
(8)

Dabei ist  $\overline{\hat{p}}$  der mittlere Schätzer (über alle Simulationen), p der wahre Wert des Parameters und  $SF(\hat{p})$  ist der empirische Standardfehler des Schätzers (über alle Simulationen). Bei Betrachtung der Formel (8) fällt auf, dass bei einer schiefen Verteilung der Schätzer für dieses Genauigkeitsmaß problematisch ist, da das arithmetische Mittel sowie der empirische Standardfehler durch die Schiefe der Verteilung verzerrt sind. Dies ist vor allem bei der Beurteilung der Schätzer der Frailtyvarianz der Fall, da dieser Schätzer durch Null beschränkt ist.

Wenn die Punktschätzer unverzerrt sind, also die Differenz zwischen Punktschätzer und wahrem Wert gegen Null geht, ist der MQF gerade die Varianz der Schätzungen [41]. Als idealer Wert für den MQF wird also ein Wert nahe der erhaltenen Varianz angesehen. Durch die Quadratwurzel wird der MQF zurück auf die Skala des betrachteten Parameters transformiert (Root Mean Square Error (RMSE))[39]. Bei einem unverzerrten Schätzer ist der RMSE der Standardfehler des Schätzers.

### 3.2.2 Zweiseitige Konfidenzintervalle

Ein Konfidenzintervall (KI) ist ein Sicherheitsbereich für die Schätzung eines Parameters p [42]. Dabei ist die Definition eines KI's abhängig von dem Parameterraum des Parameters sowie von der Verteilung seines Punktschätzers  $\hat{p}$ . Um ein Konfidenzintervall zu berechnen ist es notwendig die gewünschte Überdeckungswahrscheinlichkeit  $(100(1-\alpha)\%)$  zu spezifizieren. Dabei steht  $\alpha$  für die sogenannte Irrtumswahrscheinlichkeit und  $1 - \alpha$  bezeichnet den Konfidenzkoeffizienten [43]. Die am häufigsten verwendeten Werte für den Konfidenzkoeffizienten sind 0,9, 0,95 und 0,99 [43]. Das Konfidenzintervall umschließt einen Wertebereich mit Endpunkten U und O,  $U \leq O$ , dass der wahre Wert des Parameters zu  $100(1-\alpha)\%$  enthalten ist. In den folgenden Abschnitten werden drei Methoden zur Konstruktion von mittleren 95% KI erläutert, welche für die Auswertungen innerhalb dieser Arbeit verwendet wurden.

## Symmetrische Wald-Konfidenzintervalle

Diese Methode zur Konstruktion eines Konfidenzintervalls ist am einfachsten [44] und gilt als Standardprozedur. Sie unterliegt der Annahme, dass der Punktschätzer  $\hat{p}$  asymptotisch normal verteilt (und damit symmetrisch) ist. Bei einer Irrtumswahrscheinlichkeit von 5% ist das symmetrische Wald-KI wie folgt definiert:

$$KI_{95\%} = (\hat{p} - z_{0,975}SF(\hat{p}), \hat{p} + z_{0,975}SF(\hat{p}))$$

Dabei sei  $z_{0,975}$  das Quantil der Normalverteilung. Problematisch wird die Anwendung eines solchen KI's, z.B wenn die Verteilung des Punktschätzers eine deutliche Schiefe aufweist [45]. Dies trifft dann zu, wenn der Parameter von Interesse eine Varianzkomponente eines gemischten Modells ist, oder wenn der Schätzer nahe der Grenze des Parameterraumes liegt (siehe Abbildung 8). In der hier vorliegenden Arbeit wird angenommen das die Punktschätzer  $\hat{\sigma}^2$  und  $\hat{s}^2$  einer asymmetrischen Verteilung folgen [40]. Entsprechend diesem Sachverhalt sollten geeignete Konfidenzintervalle verwendet werden.

### Asymmetrisches Wald-Konfidenzintervall

Diese Konstruktion eines Konfidenzintervalls wird standardmäßig für Varianzkomponenten verwendet [46]. Dabei wird angenommen, dass der Punktschätzer  $\hat{p}$  einer  $\chi^2$ -Verteilung folgt. Dabei kann der Freiheitsgrad (FG) mittels der Satterthwaite-Methode angepasst werden [47], [48], [49]:

$$FG = 2\left(\frac{\hat{p}}{SF(\hat{p})}\right).$$
(9)

Die Definition des asymmetrischen Wald-KI's lautet dementsprechend [46]:

$$KI_{95\%} = \left(\frac{FG\hat{p}}{\chi^2_{FG,0,975}}, \frac{FG\hat{p}}{\chi^2_{FG,0,025}}\right)$$
(10)

#### Profile-Likelihood-Konfidenzintervall

Das Profile-Likelihood-Konfidenzintervall (PL-KI), auch Likelihood-Ratio-Konfidenzintervall genannt, wird im Allgemeinen als genauer als das symmetrische Wald-KI angesehen, insbesondere bei kleiner Grundgesamtheit [44]. Es entsteht durch die Invertierung der Teststatistik.

$$LR = 2(lnL_1(\hat{p}) - lnL_1(p_0))$$

Das PL-KI wird von einer  $\chi^2$ -Verteilung mit einem Freiheitsgrad abgeleitet. Dabei besteht das 95%-ige PL-KI aus den Werten, für die der Likelihood-Ratio-Test nicht signifikant ist. Dies ist der Fall, wenn die Teststatistik LR den Wert 3,84 nicht überschreitet (95% Perzentil der  $\chi^2$ -Verteilung mit einem Freiheitsgrad) [45].

$$2\{lnL_1(p_0) - lnL_1(\hat{p})\} \le \chi^2_{1,1-\alpha} [47]$$

$$\frac{1}{2}LR \leq \frac{1}{2}\chi^2_{1,1-\alpha}$$

Abbildung 9 zeigt die Log-Likelihoodfunktion in Abhängigkeit vom Parameter p. Die Grenzen des Konfidenzintervalls werden mit Hilfe des Wertes der Log-Likelihoodfunktion von  $\hat{p}$  ermittelt. Von diesem Wert wird  $1/2\chi_{1,1-\alpha}^2$  abgezogen (Wert bei  $lnL_1(p_0)$ ). Aus diesen beiden Schnittpunkten mit der Log-Likelihoodfunktion ergeben sich die Grenzen des PL-KI's.

Bei Betrachtung des Spezialfalles von  $\sigma^2/s^2 = 0$  (in den beobachteten Daten liegt kein Clustereffekt vor) können bei der Berechnung von mittleren zweiseitigen 95%-igen



Abbildung 9: Profile-Likelihoodfunktion und 95%-iges Konfidenzintervall.

Konfidenzintervallen Probleme auftreten. Zu beachten ist, dass in diesem besonderen Fall die untere Grenze des Parameterraumes von  $\sigma^2/s^2$  gleich dessem wahren Wert ist. In diesem Fall ist es sehr wahrscheinlich, dass symmetrische Wald-KI den definierten Parameterraum überschreiten. Daraus folgt, dass eine Überdeckungswahrscheinlichkeit von 95% nicht gegeben ist, was die Anwendbarkeit des Wald-KI's in Frage stellt. Bei Verwendung des asymmetrischen Wald-KI's gilt es zu beachten, dass bei sehr kleinen Punktschätzern  $\hat{p}$  (nahe 0) die Berechnung der Grenzen problematisch ist. Unter Berücksichtigung von Gleichung 9 fällt auf, dass der Freiheitsgrad gegen 0 läuft, je kleiner der Punktschätzer ist. Für die Berechnung des asymmetrischen Wald-KI's durch Gleichung 10 folgt damit, dass der Zähler sehr klein ist. Jedoch nimmt die  $\chi^2$ -Verteilung einen noch kleineren Wert an, wodurch die untere und obere Grenze des Konfidenzintervalls gegen  $+\infty$  streben. Damit liegt der beobachtete Punktschätzer nicht im Konfidenzintervall und eine Verwendung dieses Intervalls ist deshalb nicht empfehlenswert. Etwas flexibler reagieren hier Profile-Likelihood-Konfidenzintervalle. Das Minimum der unteren Grenze liegt bei diesen KI genau auf der unteren Grenze des Parameterraums. Dagegen ist die obere Grenze abhängig von der Log-Likelihoodfunktion.

## 3.2.3 Überdeckungsrate

Unter der Überdeckungsrate eines Konfidenzintervalls (KI) versteht man die Häufigkeit, wie oft das KI des Punktschätzers pro Simulation den wahren Wert des Parameters beinhaltet. Ausgehend von der Betrachtung eines mittleren 95%-igen KI's sollte die Überdeckungsrate ungefähr bei 95% liegen. Wenn die Überdeckungsrate zu hoch ist, spricht dies für eine zu konservative Schätzung der angewandten Methode. Als liberal wird ein Konfidenzintervall dann bezeichnet, wenn die Überdeckungsrate geringer als die vorgeschriebenen 95% ist. Im Allgemeinen sind konservative Konfidenzintervalle liberalen Konfidenzintervallen vorzuziehen [43], [50]. Der Schwankungsbereich der Überdeckungsrate ergibt sich durch [39]:

$$KI_h = \left(h - z_{\alpha/2}SF(h), h + z_{\alpha/2}SF(h)\right).$$
(11)

Dabei ist h die nominale Überdeckungswahrscheinlichkeit  $(100(1-\alpha)\%)$  und SF(h) der Standardfehler von h, unter Berücksichtigung der Simulationsanzahl n:

$$SF(h) = \sqrt{h(1-h)/n}$$

## 3.3 Daten

Die im Abschnitt 3.1 aufgelisteten Funktionen sind frei erhältlich (ausgenommen sind direkt in SAS implementierte Prozeduren). Es gibt jedoch noch keine Validierungsstudien zu diesem Thema. Damit stellt sich für den Nutzer die Frage, welche dieser Funktionen geeignet für die zu untersuchenden Daten ist. Um eine befriedigende Antwort darauf geben zu können, wurde zunächst eine umfangreiche Simulationsstudie durchgeführt. Nach eingehender Betrachtung der Simulationsergebnisse wurden die besten Verfahren genutzt, um einen realen Datensatz zu analysieren.

### 3.3.1 Simulationen

Einbezogen in die Simulationsstudie wurden die Funktionen PROC PHREG, PROC NLMIXED, SPLN3, %PCFrailty, SPGAM1, coxph, coxme und frailtyPack. Um die Güte dieser acht Funktionen zu vergleichen, wurden unterschiedliche Situationen betrachtet. Das Ziel war es, mindestens 1000 Schätzer für alle Funktionen zu erhalten. Da nicht alle Funktionen gleich gut konvergieren, wurden jeweils 2000 Datensätze pro Szenario generiert. In die anschließenden Analysen wurden nur die Datensätze einbezogen, für die alle Funktionen konvergierten, um die Vergleichbarkeit der Simulationsergebnisse zu gewährleisten. Es wurden sechs unterschiedliche Situationen simuliert, wobei zwei Modelle die Grundlage darstellten. Der erste Teil der Daten wurde auf Grundlage eines shared Gamma Frailty-Modells (5) generiert, der zweite Teil mittels eines shared Log-normalen Frailty-Modells (7). Es wurde eine Frailtyvarianz ( $\sigma^2$ ) bzw. eine Varianz des zufälligen Effektes ( $s^2$ ) von 0,5 gewählt. Die Datensätze, welche je 1000 Beobachtungen beinhalteten, wurden entweder in 20 Cluster mit je 50 Beobachtungen, 40 Cluster mit je 25 Beobachtungen oder 100 Cluster mit je 10 Beobachtungen eingeteilt. Damit sollte der Effekt von Clusteranzahl und Clustergröße auf die Schätzung untersucht werden. Die Zensierungsrate wurde auf 70 % festgelegt, wobei die maximale Beobachtungszeit bei 70 Zeiteinheiten liegt. Sowohl die Zensierungs- als auch die Ausfallzeiten folgen einer Gompertz-Verteilung mit den Parametern  $\lambda = 0,001$ ,  $\varphi_1 = 0,1$  für die Ausfallzeiten und  $\varphi_2 = 0,11$  für die Zensierungszeiten. In die Modelle eingebunden sind zwei Kovariablen:  $X_1$  (normalverteilt) und  $X_2$  (binär). Den dazugehörigen Regressionskoeffizienten wurden die Werte  $\beta_1 = 1$  und  $\beta_2 = -0.5$  zugewiesen.

Bei der Funktion PROC NLMIXED wurden für die Analyse der Simulationsdaten die Simultationsparameter (wahre Werte) als Startwerte genutzt.

#### 3.3.2 Halluca Studie

Die Halle Lung Cancer (HALLUCA)-Studie ist eine epidemiologische Studie mit dem Ziel die Versorgungsqualität von Patienten mit Lungenkrebs im südlichen Sachsen-Anhalt einzuschätzen. Untersucht wurde ein unselektiertes Patientengut aus der Region Halle - Dessau. In Zusammenarbeit mit den regionalen Krebsregistern wurden alle auflaufenden Krebspatienten auf eine standardisierte Weise dokumentiert. Der Beobachtungszeitraum ersteckte sich von April 1996 bis September 1999 mit einem anschließenden Follow-up bis September 2000 [51]. Dabei wurden die Daten der Krebsregister mit den auflaufenden Totenscheinen verglichen [52]. Zusätzlich wurde ein jährlicher Vergleich mit den Daten des gemeinsamen Krebsregisters der neuen Bundesländer (Berlin) durchgeführt. Eingeschlossen in die Studie wurden 1696 Lungenkrebspatienten. Weiterhin sollte das Überleben von Lungenkrebspatienten in Abhängigkeit von verschiedenen prognostischen Faktoren untersucht werden. Die Überlebenszeit wurde als Zeit von Diagnosestellung bis zum Tod definiert. Der Diagnosetag wurde hauptsächlich als Tag der histologischen Befundung betrachtet. Waren diese Daten nicht abrufbar, wurde auf Zeitpunkte mit geringerer Evidenz zurückgegriffen. Diese Alternativen waren:

- Tag der zytologischen Befundung,
- Tag des Tumorstagings oder
- Tag der klinischen Diagnose.

Dabei wurde immer die Angabe mit dem höchsten Evidenzlevel bevorzugt [52]. Eingeschlossen in die Halluca-Studie wurden 56 diagnostizierende Einheiten. Bei Betrachtung der medianen Überlebenszeit innerhalb dieser diagnostizierenden Einheiten fällt eine starke Heterogenität auf. In Abbildung 10 sind die 26 Einheiten, welche mehr als 5 Beobachtungen beinhalteten, bezüglich ihres medianen Überlebens dargestellt. Hier fällt auf, dass es große Unterschiede zwischen den medianen Überlebenszeiten gibt.

Die mediane Gruppengröße betrug in der Halluca-Studie 3,5 Beobachtungen. Dabei ist die beobachtete Spannweite mit einem Probanden bis 392 Probanden pro diagnostizierender Einheit sehr breit. Die Zensierungsrate innerhalb der Halluca-Studie betrug 20,5 %. Einen speziellen Fall stellt die diagnostizierende Einheit 4 dar. Hierbei handelt es sich um das Pathologische Institut des Uniklinikums Halle (Saale). In dieser Gruppe liegt eine mediane Überlebenszeit von 0 Monaten vor.



Abbildung 10: Nach Größe geordnete mediane Überlebenszeiten (mit dazugehörigem Interquartilerange) von 26 diagnostizierenden Einheiten der HALLUCA-Studie.

Es wurden fünf Kovariaten zur Modellierung verwendet, welche als wichtige Einflussfaktoren angesehen werden. Diese sind das Alter der Patienten zum Diagnosezeitpunkt (in Jahren), das Geschlecht, der Selbstständigkeitsstatus nach ECOG - Einstufung (Eastern Cooperative Oncology Gruppe) sowie Bronchialkarzinomtyp und - stadium nach UICC (Union Internationale Contre le Cancer). In Tabelle 5 ist die Studienpopulation bezüglich dieser Einflussfaktoren beschrieben. Die in die Studie aufgenommenen Patienten waren zwischen 24,2 und 91,6 Jahre alt. Das Durchschnittsalter lag bei 64,7 Jahren und das mediane Alter bei 65,3 Jahren. Die Männer sind mit 81% am stärksten in dieser Patientengruppe vertreten. Der histologische Typ ist eingeteilt in kleinzellige (sclc - small cell lung cancer) und nicht kleinzellige (nsclc - non small cell lung cancer) Tumore. Der Anteil von nicht kleinzelligen Bronchialkarzinomen liegt mit etwa 70% weit über dem der kleinzelligen Karzinome mit 22%.

Der Selbstständigkeitsstatus nach ECOG beurteilt das Maß an Eigenständigkeit eines Patienten in 6 Ausprägungen von 0 bis 5. Dabei zeigt 0 einen sehr guten körperlichen Zustand an. Je höher die Einstufung, desto größer ist die körperliche Einschränkung. Stufe 5 kodiert den Tod des Patienten. Für die Auswertungen wurde der Selbstständigkeitsstatus dichotomisiert in "guter Zustand" = Stufe 0 bis 2 und " schlechter Zustand " = Stufe 3 und 4 (siehe Tabelle 5). Rund 69% der Patienten sind zum Zeitpunkt der Diagnose in einem guten körperlichen Zustand. Auffällig ist, dass der Anteil der fehlenden Werte in dieser Variable mit 24% am höchsten ausfällt. Die Tumorstadien wurden gemäß UICC kodiert, allerdings wurden die Stadien Ia und Ib sowie IIa und IIb auf Grund ihrer geringen Häufigkeit zusammengefasst. Dennoch bilden diese zwei Gruppen lediglich einen Anteil von 15,6%. Den größten Anteil stellen die Patienten mit Tumorstadium IV dar (36,6%). Bis zum Ende des Follow-up verstarben 1349 Patienten (79, 5%). Die mediane Überlebenszeit betrug 9,3 Monate.

Kovariable	Kategorie	Absolute Häufigkeit	relative Häufigkeit
Geschlecht	Männer	1374	81,0%
	Frauen	322	19,0%
${ m histologischer}$	$\operatorname{sclc}$	336	21,6%
Тур	$\operatorname{nsclc}$	1183	69,7%
	fehlend	147	8,7%
Lebensqualität	1 bis $2$	1166	68,8%
(nach ECOG)	3 bis $4$	123	7,2%
	fehlend	407	24,0%
$Tumorstadium^1$	Ι	185	10,9%
	II	79	4,7%
	IIIa	195	11,5%
	IIIb	280	16,5%
	IV	621	$36{,}6\%$
	fehlend	336	19,8%

Tabelle 5: Patientencharakteristik der Halluca - Studie.

<sup>1</sup> Tumorstadium nach UICC

# 4 Ergebnisse

In diesem Kapitel werden sowohl die Ergebnisse der Simulationen als auch die Ergebnisse der Analyse des realen Datensatzes (Halluca-Studie) dargestellt und erläutert. Im Folgenden werden die Funktionen, Prozeduren und Makros aus Abschnitt 3.1 unter dem Sammelbegriff **Funktion** zusammengefasst. Zunächst werden die Ergebnisse für das shared Gamma Frailty-Modell dargestellt. Im Anschluss daran folgen die Ergebnisse von Funktionen für das shared Log-normale Frailty-Modell.

# 4.1 Ergebnisse der Funktionen SPGAM1, coxph und frailty-Penal

## 4.1.1 Ergebnisse der durch ein PH Modell generierten Daten

Dieser Abschnitt befasst sich mit der Analyse der 2000 Simulationsdatensätze, welche mittels eines Gompertz-verteilten PH Modells in SAS 9.3 simuliert wurden. Eine Clustervariable ist in den Datensätzen zwar vorhanden, jedoch ist kein diesbezüglicher Effekt generiert worden. Auf diese Daten wurden die folgenden Modelle angepasst (mit dazugehörigen Funktionen):

- 1. ein Gompertz-verteiltes PH Modell (parametrischer Ansatz in PROC NLMIXED ohne RANDOM-Statement, mit Simulationsparametern als Startwerten),
- 2. ein semiparametrisches shared Gamma Frailty-Modell (realisiert in SPGAM1, coxph),
- 3. ein quasi-semiparametrisches shared Gamma Frailty-Modell (realisiert in frailtyPenal), sowie
- 4. ein Proportionales Hazards Modell (semiparametrischer Ansatz in PROC PHREG ohne RANDOM-Statement).

Die Ergebnisse sind in Tabelle 6 und Abbildung 11 aufgeführt. Auf Grund der verwendeten Modellansätze ist zu erwarten, dass die Funktion PROC NLMIXED die besten Schätzergebnisse erzielen. Damit können sie als Referenzen betrachtet werden.

Tabelle	e 6: Sin	nulatio	nsergebn	11  sse von  200	JU Wiede	erholung	en eines	Gompertz-	verteiltei	n Propoi	tionalen	Hazards M	lodells.
			Cluster	ranzani = 20			Cluste	ranzani = 40		Cluster größe = 10			
			N	$-1989^{1}$			N	$-1000^{1}$			Ofusie.	$-1002^{1}$	
			11	- 1505				- 1550			11	- 1552	
Wahre	r Wert	Median	Ges.SF E	Emp.SF MQF	$\ddot{\mathrm{U}}\mathrm{b}\mathrm{erd.}^2$	Median	Ges.SF	Emp.SF MQF	$\ddot{\mathrm{U}}\mathrm{berd.}^2$	Median	Ges.SF I	Emp.SF MQF	Überd. <sup>2</sup>
Param	etrisches	Modell:	PROC NLM	IXED									
$\beta_1$	1.0	1,006	0,067	$0,066 \ 0,004$	95,12	1,006	0,067	0,066 $0,004$	95,13	1,006	0,067	0,066 $0,004$	$95,\!13$
$\beta_2$	-0,5	-0,499	0,118	$0,\!115\ 0,\!013$	94,72	-0,499	0,118	$0,\!115\ 0,\!013$	94,72	-0,499	0,118	$0,\!115\ 0,\!013$	$94,\!73$
Semipa	arametri	sches sha	ared Gamn	na Frailty-Mod	ell: SPGAM:	L							
$\beta_1$	$1,\!0$	1,007	0,068	0,067 $0,005$	95,02	1,009	0,068	0,067  0,005	95,23	$1,\!010$	0,069	0,067 $0,005$	$95,\!33$
$\beta_2$	-0,5	-0,498	$0,\!119$	$0,\!116\ 0,\!013$	94,82	-0,499	0,119	$0,\!116\ 0,\!013$	94,87	-0,499	0,119	$0,117 \ 0,014$	$94,\!83$
$\sigma^2$	$^{0,0}$	0,000	$0,\!031$	0,013 $0,000$	94,77	0,000	0,040	0,018 $0,000$	98,69	0,000	0,057	$0,028 \ 0,001$	$99,\!65$
Quasi-	semipara	ametrisch	nes shared	Gamma Frailty	y-Modell: :	frailtyPe	nal						
$\beta_1$	$1,\!0$	1,010	0,067	0,067 $0,005$	94,97	1,011	0,068	0,067  0,005	94,92	1,012	0,068	0,068 $0,005$	94,78
$\beta_2$	-0,5	-0,483	$0,\!118$	$0,\!117\ 0,\!014$	94,42	-0,483	0,119	$0,\!117\ 0,\!014$	94,47	-0,484	0,119	$0,\!117\ 0,\!014$	$94,\!48$
$\sigma^2$	$^{0,0}$	0,000	$0,\!010$	0,013 $0,000$	$100,\!00$	0,000	0,015	0,018 $0,000$	99,80	0,000	0,023	0,027 $0,001$	$99,\!65$
Semipa	arametri	sches sha	ared Gamn	na Frailty-Mod	ell: coxph								
$\beta_1$	1,0	1,007	0,067	$0,067 \ 0,005$	95,07	1,008	0,067	0,067 $0,005$	94,92	1,009	0,067	0,067 $0,005$	$94,\!83$
$\beta_2$	-0,5	-0,498	$0,\!118$	0,116 $0,013$	94,77	-0,498	0,119	0,116 $0,013$	94,87	-0,498	0,119	0,116 $0,013$	$94,\!83$
$\sigma^2$	0,0	0,000	-	0,014 0,000	_	0,000	-	0,017 $0,000$	_	0,000	-	0,026 $0,001$	_
Propor	tionales	Hazards	Modell na	ach Cox: PROC	PHREG								
$\beta_1$	$1,\!0$	1,005	0,067	$0,066 \ 0,004$	95, 17	$1,\!005$	0,067	$0,066 \ 0,004$	95, 18	$1,\!005$	0,067	$0,066 \ 0,004$	$95,\!18$
$\beta_2$	-0,5	-0,497	$0,\!118$	$0,\!116\ 0,\!013$	$94,\!92$	-0,496	0,118	$0,\!116\ 0,\!013$	$94,\!92$	-0,497	0,118	$0,\!116\ 0,\!013$	$94,\!93$

T-1-11-6.01 1 4. . 0 1 11 . . . • • • ъ . . тт

 $^1$  Anzahl der über alle Funktionen konvergierten Datensätze $^2$  Symmetrisches 95% Wald Konfidenzintervall



Abbildung 11: Feste und zufällige Effektschätzer der 2000 Simulationsdaten, generiert durch ein proportionales Hazards Modell nach Cox; wahre Werte mittels horizontaler Linie dargestellt.

Als Erstes werden die Schätzungen der festen Effekte genauer betrachtet. Sowohl die Schätzer für  $\beta_1$  als auch für  $\beta_2$  sind nahezu identisch über alle Funktionen und die Güte der Schätzung ist unabhängig von der Clusteranzahl wie in Tabelle 6 zu erkennen ist. Die mediane Abweichung liegt zwischen 0,006 und 0,010. Etwas größere Abweichungen treten bei der Funktion frailtyPenal auf. Hier liegt die mediane Abweichung vom wahren Wert zwischen 0,010 und 0,012 für  $\hat{\beta}_1$  und bei 0,017 für  $\hat{\beta}_2$ . Bezüglich der Standardfehler sind zwischen den Funktionen keine wesentlichen Unterschiede zu erkennen. Diese liegen im Bereich von 0,067 bis 0,069 für  $\hat{\beta}_1$  und von 0,115 bis 0,119 für  $\hat{\beta}_2$ . Weiterhin gibt es keine nennenswerten Unterschiede zwischen geschätzten und empirischen Standardfehlern. Dies spricht für eine adäquate Schätzung der Standardfehler. Vergleicht man die mittleren quadratischen Fehler (MQF) von  $\hat{\beta}_1$  mit  $\hat{\beta}_2$  fällt auf, dass die MQF's unter  $\beta_2$  ungefähr halb so groß sind wie unter  $\beta_1$  (0,005 unter  $\hat{\beta}_1$ im Vergleich zu 0,013 unter  $\hat{\beta}_2$ ). Grund hierfür sind größere Standardfehler. Bereits die mediane Abweichung und der MQF sprechen für eine gute Schätzung der festen Effekte. Bestätigt wird dies zusätzlich durch Überdeckungsraten, welche zwischen 94% und 96% liegen. Auch bezüglich des Interquartile Ranges (IQR) und der Spannweite gibt es kaum Unterschiede zwischen den Funktionen. In Abbildung 11 ist zu erkennen, dass der IQR für  $\hat{\beta}_1$  über alle Funktionen und Clustergrößen bei einem Wert von 0,09 liegt. Im Fall von  $\hat{\beta}_2$  beträgt der IQR ungefähr 0,15. Lediglich unter der Funktion frailtyPenal fällt der IQR marginal größer aus.

Als Zweites ist in Tabelle 6 zu erkennen, dass die gerundeten Mediane der Schätzer für  $\sigma^2$  gleich dem Wert 0 sind (die Werte schwanken im Bereich [0; 5 \* 10<sup>-7</sup>]). Damit sind die Abweichungen vom wahren Wert sehr gering. Bezüglich der Standardfehler für  $\hat{\sigma}^2$  gibt es zwischen den Prozeduren große Unterschiede. Während keine Aussagen zur Funktion coxph getroffen werden können (Grund hierfür ist, dass keine Standardfehler für den Varianzschätzer zur Verfügung gestellt werden) beträgt bei SPGAM1 der geschätzte Standardfehler etwa das Doppelte des empirischen Standardfehlers. Auch über die Clustergrößen ist ein Trend ersichtlich. Je mehr Cluster vorhanden sind (bei sinkender Clustergröße), desto größer werden der empirische und geschätzte Standardfehler. Dahingegen sind die von der Funktion frailtyPenal zur Verfügung gestellten Standardfehler von  $\hat{\sigma}^2$  nahezu identisch mit den empirischen Standardfehlern und liegen in einem Bereich von 0,01 bis 0,027 (Tabelle 6). Auf Grund der geringen medianen Abweichung liegt der MQF hier ebenso bei 0. Bei der Bewertung der Konfidenzintervalle treten weitere Probleme auf. Die asymmetrischen Konfidenzintervalle sind nach ihrer Definition nicht für den Fall geeignet, dass der Schätzer 0 ist. Der Freiheitsgrad strebt in diesem Fall gegen 0. Dies führt dazu, dass die zugehörigen Grenzen des Konfidenzintervalls gegen  $+\infty$  laufen. Aufgrund dessen liegen die Uberdeckungsraten bei 0 und sind nicht in der Tabelle 6 angegeben. Der Einsatz von symmetrischen KI ist ebenso wenig geeignet, da sie den negativen Wertebereich mit einschließen und somit fast immer die 0 (Überdeckungsrate von 95% - 100%). Mit der Überschreitung der unteren Grenze des Parameterraumes ist die Definition eines mittleren 95%-igen KI's nicht mehr sinnvoll. Dies stellt die Interpretierbarkeit der symmetrischen Konfidenzintervalle für die Varianz der Frailtyvarianz in Frage. Betrachtet man dahingegen die IQR's und Spannweiten der Schätzer von  $\sigma^2$  in Abbildung 11, fallen Unterschiede zwischen den drei Funktionen auf. Bei der Funktion coxph bleibt der IQR der Schätzungen über alle Clustergrößen annähernd konstant. Anders verhält es sich bei der Funktion SPGAM1. Hier steigt der IQR mit sinkender Clustergröße. Etwas geringer fällt dieser Effekt bei der Funktion frailtyPenal aus. Die Spannweiten sind wiederum über alle Funktionen recht ähnlich. Eine Möglichkeit, die Robustheit der Schätzalgorithmen der Funktionen zu betrachten, bietet die Konvergenzrate (Tabelle 7). Sowohl unter coxph als auch unter frailtyPenal sind alle Datensätze konvergiert. Lediglich 5 Datensätze sind unter SPGAM1 nicht konvergiert.

	Clustergröße: 50	Clustergröße: 25	Clustergröße: 10
PROC NLMIXED	$99,\!60\%$	$99,\!60\%$	$99,\!60\%$
SPGAM1	$99,\!90\%$	$99,\!85\%$	$100,\!00\%$
frailtyPenal	$100,\!00\%$	$100,\!00\%$	$100,\!00\%$
coxph	$100,\!00\%$	$100,\!00\%$	$100,\!00\%$
PROC PHREG	$100,\!00\%$	$100,\!00\%$	$100,\!00\%$

Tabelle 7: Konvergenzraten der Analyse von 2000 Datensätzen, welche durch ein Gompertzverteiltes Proportionales Hazards Modell erzeugt wurden.

## 4.1.2 Ergebnisse der durch ein shared Gamma Frailty-Modell generierten Daten

Betrachtet werden in diesem Abschnitt die Ergebnisse der Auswertungen der Funktionen SPGAM1, frailtyPenal und coxph, welche eine Gamma-verteilte Frailty annehmen. Erzeugt wurden die 2000 Datensätze mit Hilfe eines shared Gamma Frailty-Modells (siehe Abschnitt 3.3.1). Dargestellt werden die Simulationsergebnisse in Tabelle 9 und veranschaulicht in der Abbildung 12. Die in den Analysen verwendeten Modellansätze bzw. Funktionen sind:

- 1. ein shared Gamma Frailty-Modell mit Gompertz-verteilter Basishazardfunktion (parametrischer Ansatz in PROC NLMIXED nach der Methode von Liu und Yu [53], mit Simulationsparametern als Startwerten), sowie
- 2. die in Abschnitt 4.1.1 unter den Punkten 2. 4. aufgeführten Modellansätze und Funktionen.

Betrachtet man die Konvergenzraten aus Tabelle 8 sind keine Unterschiede zwischen den interessierenden Funktionen SPGAM1, frailtyPenal und coxph zu erkennen. Die zu vergleichszwecken genutzte Funktion PROC NLMIXED schneidet in Bezug auf die Konvergenzrate deutlich schlechter ab. Bei all diesen Funktionen liefen die 2000 Simulationsdatensätze ohne Probleme durch. In Tabelle 9 sind die Simulationsergebnisse im Detail dargestellt.

Tabelle 8: Konvergenzraten der Analyse von 2000 Datensätzen, welche durch ein shared Gamma Frailty-Modell erzeugt wurden.

	Clustergröße: 50	Clustergröße: 25	Clustergröße: 10
PROC NLMIXED SPGAM1	81,40% 100,00%	$79,25\% \\ 100,00\%$	$74,20\%\\100,00\%$
frailtyPenal	$100{,}00\%$	$100{,}00\%$	$100,\!00\%$
coxph	$100,\!00\%$	$100,\!00\%$	$100,\!00\%$
PROC PHREG	$100,\!00\%$	$100,\!00\%$	$100,\!00\%$

Zunächst werden die Ergebnisse der Schätzungen der festen Effekte betrachtet. Bei den folgenden Aussagen ist das proportionale Hazards Modell nach Cox (PROC PHREG) nicht eingeschlossen. Die mediane Abweichung der  $\beta_1$ -Schätzer ist gering und liegt zwischen 0,001 und 0,006. Ähnliches gilt für die Schätzer von  $\beta_2$  (0,001 bis 0,009). Ausgenommen sind dabei die Schätzungen der Funktion frailtyPenal, bei denen es zu einer etwas stärkeren medianen Abweichung kommt. Hier liegt die mediane Abweichung zwischen 0,019 und 0,027. Allgemein ist bei der Schätzung der festen Effekte ein leichter Trend zu erkennen. Je mehr Cluster vorhanden sind beziehungsweise je geringer die Clustergröße ist, desto geringer ist die mediane Abweichung. Diese Aussage wird durch

	-	Labelle	9: Simul	ationsergeb	nisse von	2000 W	/iederhc	olungen des s	shared G	amma F	railty-M	odells.	
			$\operatorname{Cluste}$	${ m ranzahl}=20$			Cluste	${ m eranzahl}=40$			$\operatorname{Clusters}$	${ m anzahl}=100$	
			Cluste	$ m ergr{ m o}$ se $= 50$			$\operatorname{Clust}$	${ m ergr{ m ö}}$ ß ${ m e}=25$			Cluste	$ m rgr{o}$ ße = 10	
			Ν	$= 1628^{1}$			Ν	$= 1585^{1}$			Ν	$= 1484^{1}$	
Wahren	r Wert	Median	Ges.SF E	Emp.SF MQF	$\ddot{\mathrm{U}}\mathrm{berd.}^2$	Median	$\operatorname{Ges.SF}$	Emp.SF MQF	$\ddot{\mathrm{U}}\mathrm{berd.}^2$	Median	Ges.SF E	Emp.SF MQF	$\ddot{\mathrm{U}}\mathrm{berd.}^2$
Param	etrisches	shared	Gamma Fi	railty-Modell: I	ROC NLMI	KED							
$\beta_1$	1.0	1.005	0.069	$0.071 \ 0.005$	96.50	0.999	0.071	0.071 $0.005$	95.90	1.005	0.074	0.074 $0.006$	95.55
$\beta_2$	-0.5	-0.493	0.122	$0.121 \ 0.015$	96.31	-0.495	0.124	0.125 $0.016$	95.46	-0.499	0.128	$0.127 \ 0.016$	95.15
$\sigma^2$	0.5	0.460	0.162	$0.165 \ 0.028$	88.57 /	0.488	0.138	$0.138 \ 0.019$	91.92 /	0.490	0.123	$0.123 \ 0.015$	94.88 /
-	0,0	0,200	0,-0-	0,200 0,020	$97.05^3$	0,200	0,200	0,200 0,020	$96.47^3$	0,200	0,0	0,220 0,020	$95.49^3$
					01,00				00,11				00,10
Semipa	arametri	sches sha	ared Gamr	na Frailty-Mod	ell: SPGAM1								
β <sub>1</sub>	1.0	1.003	0.070	0.071 0.005	95.09	0.999	0.071	0.071 $0.005$	95.08	1.004	0.074	0.074 $0.005$	94.88
$\beta_1$	-0.5	-0.491	0.122	$0.121 \ 0.015$	95.21	-0.496	0.124	0.125 $0.016$	94.89	-0.496	0.128	$0.127 \ 0.016$	94.68
$\sigma^2$	0.5	0, 101 0 456	0 162	0.165 0.028	87 29 /	0 484	0,137	0.137 0.019	91 23 /	0,100 0.482	0,123	0.123 0.015	94.61 /
0	0,0	0,100	0,102	0,100 0,020	$96,99^3$	0,101	0,101	0,101 0,010	$96.59^3$	0,102	0,120	0,120 0,010	$95.75^3$
					00,00				00,00				00,10
Quasi-s	semipara	ametrisch	nes shared	Gamma Frailt	v-Modell: f	railtvPe	nal						
β1	1.0	1.006	0.070	0.071 0.005	95.21	0.999	0.071	0.071 $0.005$	95.08	1.004	0.074	0.074 $0.005$	94.81
Ba	-0.5	-0.473	0.122	0.123 0.016	94.41	-0.479	0.124	0.126 0.016	94.07	-0.481	0.127	0.128 0.017	93.87
$\sigma^2$	0.5	0,458	0,165	0.167 0.029	87 22 /	0,170 0,479	0,137	0.138 0.019	90.60 /	$0,101 \\ 0.473$	0,122	0.123 0.015	94.00 /
0	0,0	0,100	0,200	0,101 0,010	$97.17^3$	0,110	0,101	0,100 0,010	$96.78^3$	0,110	0,122	0,120 0,010	$96.36^3$
					01,11				00,10				00,00
Semina	arametri	sches sha	ared Gamr	na Frailty-Mod	ell: coxph								
β <sub>1</sub>	1.0	1.004	0.070	0.071 $0.005$	95.02	0.999	0.071	0.071 $0.005$	94.76	1.003	0.072	0.074 $0.005$	94.54
$\beta_{2}^{P_{1}}$	-0.5	-0.492	0.122	$0.121 \ 0.015$	95.33	-0.496	0.124	0.125 $0.016$	95.02	-0.496	0.127	$0.127 \ 0.016$	94.74
$\sigma^2$	0.5	0.484	-	$0.179 \ 0.032$	-	0.494		$0.143 \ 0.020$		0.474	-	$0.121 \ 0.015$	-
5	0,0	0,101		0,110 0,002		0,101		3,110 0,020		0,111		0,121 0,010	
Propor	tionales	Hazards	s Modell na	ach Cox: PROC	PHREG								
$\beta_1$	1,0	0,883	0,065	0,076 0,020	53,81	0,875	0,065	0,071 $0,020$	50,98	0.875	0,065	0,070 $0,020$	50,74
$\beta_2$	-0,5	-0,434	$0,\!119$	$0,\!122$ $0,\!019$	90,60	-0,436	0,119	$0,\!124$ $0,\!020$	$89,\!65$	-0,429	$0,\!119$	$0,\!123\ 0,\!020$	90,36

\*\*\*\* . . . . . ..

<sup>1</sup> Anzahl der über alle Funktionen konvergierten Datensätze
 <sup>2</sup> Symmetrisches 95% Wald Konfidenzintervall
 <sup>3</sup> Asymmetrisches 95% Wald Konfidenzintervall



Abbildung 12: Ergebnisse des shared Gamma Frailty-Modells; wahre Werte mittels horizontaler Linie dargestellt.

die Boxplots von  $\hat{\beta}_1$  und  $\hat{\beta}_2$  in Abbildung 12 bestätigt. In der Abbildung 12 ist gut zu erkennen, dass es kaum Unterschiede in den Spannweiten wie auch in den IQR's gibt. Durchschnittlich liegt die Spannweite bei den  $\beta_1$ -Schätzern bei 0,5, wohingegen sie für  $\hat{\beta}_2$  bei 0,82 liegt. Die geschätzten Standardfehler für die Schätzer der festen Effekte weisen kaum Unterschiede auf. Über alle Funktionen und Clustergrößen betrachtet liegen sie in einem Bereich von 0,069 - 0,074. Ebenfalls gering sind die Unterschiede zwischen geschätzten und empirischen Standardfehlern. Ausgehend von diesen Betrachtungen ist es nachvollziehbar, dass der MQF über alle Auswertungen gesehen nahezu identisch ist. Bei den Schätzungen von  $\beta_1$  überschreitet dieser nicht den Wert 0,006. Im Vergleich dazu sind die MQF's der  $\beta_2$ -Schätzer doppelt so groß und liegen in einem Bereich zwischen 0,015 und 0,017. Weiterhin wird die Güte der Schätzungen von  $\hat{\beta}_1$ und  $\hat{\beta}_2$  durch die adäquate Überdeckungsrate von [94, 68%; 96, 90%] unterstrichen. Abgeleitet wird dieser Schwankungsbereich der Überdeckungsrate von Gleichung 11 bei ca. 1500 Datensätzen (die tatsächlich konvergierten Datensätze) und einer nominalen Überdeckungswahrschenlichkeit von 95%. Einzig bei der Funktion frailtyPenal ist die Überdeckungsrate unter  $\hat{\beta}_2$  etwas geringer.

Auffälliger sind die Unterschiede der Schätzungen bei der Frailtyvarianz  $\sigma^2$  (Tabelle 9). Zu erkennen ist, dass bessere Ergebnisse für Schätzungen mit steigender Clusteranzahl (geringerer Clusterumfang) erzielt werden. Die mediane Abweichung ist bei 20 Clustern, was einer Clustergröße von  $n_i=50$  entspricht, mit 0,044 am größten. Eine Ausnahme stellt die Schätzung mittels der Funktion coxph dar, welche über alle Clustergrößen stabile Schätzergebnisse liefert (im Mittel 0,484). Weiterhin weichen die Schätzungen von frailtyPenal im Vergleich zu den anderen Funktionen etwas stärker ab. Bei Betrachtung der Boxplots für  $\hat{\sigma}^2$  in Abbildung 12 ist zu erkennen, dass die Spannweite wie auch der IQR über alle Funktionen annähernd identisch sind. Jedoch sinkt die Spannweite mit steigender Clusteranzahl von durchschnittlich 1,177 auf 0,752. Ein vergleichbares Verhalten ist auch für den IQR zu erkennen.

Die Nichtberücksichtigung eines vorliegenden zufälligen Effektes in der Analyse kann zu erheblichem Bias bei der Schätzung der festen Effekte führen. Veranschaulicht wird dies durch die Ergebnisse der Funktion PROC PHREG. Bewusst wurde dabei das RANDOM-Statement nicht definiert. Die Schätzer von  $\beta_1$  weichen um 0,22 vom wahren Wert ab. Geringer fällt dieser Effekt bei der Schätzung für  $\beta_2$  aus. Hier liegen die Schätzer im Bereich zwischen -0,436 und -0,429. Dies ist zusätzlich in Abbildung 12 durch die Boxplots veranschaulicht. Trotz der größeren Verzerrung der Schätzer der festen Effekte liegen die Standardfehler auf dem gleichen Niveau wie die Standardfehler der angepassten Funktionen. Dies führt dazu, dass der MQF vierfach erhöht ist für  $\hat{\beta}_1$ . Auf Grund der großen Abweichung vom wahren Wert und der verhältnismäßig kleinen Standardfehler ist die Überdeckungsrate der Schätzer für  $\beta_1$  sehr gering (ca. 52%). Im Fall von  $\hat{\beta}_2$  wird hingegen eine Überdeckungsrate von ca. 90% erreicht.

# 4.2 Ergebnisse der Funktionen PROC PHREG, %PCFrailty, SPLN3, coxme und coxph

## 4.2.1 Ergebnisse der durch ein PH Modell generierten Daten

In diesem Abschnitt sind die Analysen der Simulationsdaten dargestellt, welche durch ein Proportionales Hazards Modell erzeugt wurden. Zur Auswertung wurden folgende Modelle bzw. Funktionen verwendet:

- 1. ein Gompertz-verteilten PH Modell (parametrischer Modellansatz in PROC NLMIXED, mit Simulationsparametern als Startwerten),
- 2. ein semiparametrisches shared Log-normales Frailty-Modell (realisiert in SPLN3, PROG PHREG, coxme und coxph),
- 3. ein verallgemeinertes gemischtes Modell (realisiert in %PCFrailtymit je 4 Ereignissen pro Piece beziehungsweise 10 Ereignissen pro Piece), sowie
- 4. ein semiparametrisches Proportionales Hazards Modell (realisiert in PROC PHREG ohne Random-Statement).

Zunächst werden die Funktionen bezüglich ihrer Schätzungen der festen Effekte miteinander verglichen (Tabelle 10). Bei den Schätzern für  $\beta_1$  fällt auf, dass alle Funktionen den wahren Wert überschätzen und im Wertebereich von 1,009 bis 1,012 liegen. Dabei scheinen die Schätzungen keinem Trend bezüglich Clusteranzahl bzw. Clustergröße zu unterliegen. Gleiches gilt für die Schätzer von  $\beta_2$ . Ihre mediane Abweichung vom wahren Wert ist maximal 0,003. Die geringen Unterschiede der Schätzer  $\hat{\beta}_1$  und  $\hat{\beta}_2$  über alle Funktionen ist in Abbildung 13 sehr gut zu erkennen. Weder die Spannweiten noch die IQR's deuten auf qualitative Unterschiede hin. Auch die geschätzten Standardfehler unterscheiden sich über alle Funktionen kaum voneinander. Auffällig ist lediglich, dass

Tabelle	10: Simu Clu Clu	$\begin{array}{l} \textbf{lationse} \\ \textbf{steranzahl} = \\ \textbf{stergröße} = \\ \textbf{N}^1 = 1657 \end{array}$	<b>rgebni</b> = 20 : 50	isse von	1 2000 V	Vieder <sup>l</sup> Clust Clus	nolungen teranzahl = tergröße = $N^1 = 1713$	a eines : 40 25	Gompe	ertz-ver	teilten <sup>Clust</sup> Clus	$\begin{array}{l} \mathbf{PH} \ \mathbf{Mc} \\ \mathrm{eranzahl} = \\ \mathrm{stergröße} = \\ \mathrm{stergröße} = \\ \mathrm{N}^1 = 1714 \end{array}$	dell. $100$ $10$	
Wahrer Wert Median	Ges.SF	Emp.SF	MQF	Überd. <sup>2</sup>	Median	Ges.SF	Emp.SF	MQF	Überd. <sup>2</sup>	Median	Ges.SF	Emp.SF	MQF	Überd. <sup>2</sup>
Parametrisches PH Mod ( $\beta_1$ 1,0 1,009 $\beta_2$ -0,5 -0,499	.ll: PROC N 0,067 0,118	LMIXED 0,066 0,114	$0,004 \\ 0,013$	95,17 95,05	1,008 -0,499	0,067 0,118	0,066 0,114	$0,004 \\ 0,013$	$95,21 \\ 95,10$	1,008 -0,499	0,067 0,118	0,066 0,114	$0,004 \\ 0,013$	95,22 95,10
$\begin{array}{llllllllllllllllllllllllllllllllllll$	ed Log-norn 0,067 0,119 0,014	nales Frailty 0,067 0,115 0,015	<ul> <li>Modell:</li> <li>0,005</li> <li>0,013</li> <li>0,000</li> </ul>	PROC PH 94,99 95,17 -	REG (mit z 1,010 -0,497 1,9E-08	ufälligem 0,067 0,119 0,019	$\begin{array}{c} {\rm Effekt}) \\ 0,067 \\ 0,117 \\ 0,020 \end{array}$	$\begin{array}{c} 0,005\\ 0,014\\ 0,001\end{array}$	94,92 94,92 -	1,012 -0,502 3,2E-07	$\begin{array}{c} 0,068\\ 0,119\\ 0,030\end{array}$	$\begin{array}{c} 0,068\\ 0,116\\ 0,031\end{array}$	$\begin{array}{c} 0,005\\ 0,013\\ 0,001\end{array}$	94,69 95,33 -
$\begin{array}{llllllllllllllllllllllllllllllllllll$	PCFrailt, 0,067 0,119 0,029	y 0,067 0,115 0,015	$0,005 \\ 0,013 \\ 0,000$	94,27 95,17	1,013 -0,499 2,7E-20	$\begin{array}{c} 0,067\\ 0,119\\ 0,038\end{array}$	$\begin{array}{c} 0,068\\ 0,115\\ 0,020\end{array}$	$\begin{array}{c} 0,005\\ 0,013\\ 0,001\end{array}$	94,22 95,39	1,015 - 0,502 0,002	$\begin{array}{c} 0,068\\ 0,119\\ 0,057\end{array}$	$\begin{array}{c} 0,068\\ 0,116\\ 0,032\end{array}$	$\begin{array}{c} 0,005\\ 0,013\\ 0,001\end{array}$	94,34 95,38 -
$ \begin{array}{c} 10 \   {\rm Ereignisse} \   {\rm pro} \   {\rm Piece} \\ \beta_1 \\ \beta_2 \\ \beta_2 \\ \beta_2 \\ \beta_2 \\ 0,0 \   6,6 E-20 \\ \end{array} $	<b>\$PCFrailt</b> 0,067 0,118 0,029	5 <b>y</b> 0,067 0,115 0,015	$\begin{array}{c} 0,005\\ 0,013\\ 0,000\end{array}$	94,45 95,11	1,013 -0,499 1,1E-19	$\begin{array}{c} 0,067\\ 0,119\\ 0,038\end{array}$	$\begin{array}{c} 0,068\\ 0,115\\ 0,020\end{array}$	$\begin{array}{c} 0,005\\ 0,013\\ 0,001\end{array}$	94,22 95,21	1,015 - 0,502 0,001	$\begin{array}{c} 0,068\\ 0,119\\ 0,057\end{array}$	$\begin{array}{c} 0,068\\ 0,115\\ 0,031\end{array}$	$\begin{array}{c} 0,005\\ 0,013\\ 0,001\end{array}$	94,17 95,20 -
Semiparametrisches shar $\beta_1$ 1,0 1,010 $\beta_2$ -0,5 -0,498 $s^2$ 0,0 0,000	ed Log-norn 0,068 0,119 0,031	nales Frailty 0,067 0,115 0,013	Modell: 0,005 0,013 0,000	<b>SPLN3</b> 94,99 95,35	1,010 - 0,498 0,000	$0,068 \\ 0,119 \\ 0,040$	$\begin{array}{c} 0,067\\ 0,115\\ 0,018\end{array}$	$0,005 \\ 0,013 \\ 0,000$	95,33 95,39	1,012 - 0,500 0,000	$\begin{array}{c} 0,069\\ 0,120\\ 0,059\end{array}$	0,068 0,115 0,029	$\begin{array}{c} 0,005\\ 0,013\\ 0,001\end{array}$	95,27 95,33 -
Semiparametrisches shar $\beta_1$ 1,0 1,011 $\beta_2$ -0,5 -0,499 $s^2$ 0,0 0,000	ed Log-norn 0,067 0,119 -	nales Frailty 0,067 0,115 0,015	Modell: 0,005 0,013 0,000	<b>coxme</b> 94,99 95,17	1,011 -0,499 0,000	0,067 0,119 -	$\begin{array}{c} 0,067\\ 0,115\\ 0,020\end{array}$	$\begin{array}{c} 0,005\\ 0,013\\ 0,001\end{array}$	94,80 95,45	1,012 - 0,500 0,000	0,067 0,119 -	$\begin{array}{c} 0,068\\ 0,115\\ 0,030\end{array}$	$\begin{array}{c} 0,005\\ 0,013\\ 0,001\end{array}$	94,75 95,33 -
$\begin{array}{llllllllllllllllllllllllllllllllllll$	ed Log-norn 0,067 0,119 -	nales Frailty 0,067 0,115 0,014	Modell: 0,005 0,013 0,000	<b>coxph</b> 94,99 95,17	1,011 -0,499 0,000	0,067 0,119 -	$\begin{array}{c} 0,067\\ 0,115\\ 0,020\end{array}$	$\begin{array}{c} 0,005\\ 0,013\\ 0,001\end{array}$	94,75 95,45	1,012 - 0,500 0,001	0,068 0,119 -	$\begin{array}{c} 0,068\\ 0,115\\ 0,031\end{array}$	$\begin{array}{c} 0,005\\ 0,013\\ 0,001\end{array}$	94,69 95,33 -
Proportionales Hazards I $\beta_1$ 1,007 $\beta_2$ -0,5 -0,497	Aodell nach 0,067 0,118	Cox: PRDC 0,067 0,114	: PHREG 0,005 0,013	95,11 95,35	1,007 -0,497	$0,067 \\ 0,118$	$0,067 \\ 0,114$	0,005 0,013	95,15 95,39	1,007 -0,497	$0,067 \\ 0,118$	$0,067 \\ 0,114$	$0,005 \\ 0,013$	95,16 95,39
<ol> <li>Anzahl der über alle F<sup>1</sup></li> <li>Symmetrisches 95% Wi</li> </ol>	ınktionen k ald Konfideı	onvergierter nzintervall	ı Datensä	itze										



Abbildung 13: Ergebnisse des shared Log-normalen Frailty-Modells; wahre Werte mittels horizontaler Linie dargestellt.

die Standardfehler für  $\hat{\beta}_1$  nur halb so groß wie für  $\hat{\beta}_2$  sind. Die Unterschiede zwischen den geschätzten und empirischen Standardfehlern sind sehr gering. Aufgrund dessen ist es nachvollziehbar, dass der MQF unter  $\hat{\beta}_2$  mit durchschnittlich 0,013 ungefähr dreimal so groß ist wie unter  $\hat{\beta}_1$ . Auch die Überdeckungsrate ist über alle Funktionen vergleichbar. Bei  $\hat{\beta}_1$  liegt sie im Mittel einen Prozentpunkt unter den Überdeckungsraten von  $\hat{\beta}_2$ . Alle Überdeckungsraten liegen jedoch im erwarteten Bereich von 94% - 96%.

Für die Betrachtung der Varianzschätzer ist folgende Fragestellung zu beachten: Wie gut erkennen die Funktionen PROC PHREG, %PCFrailty, SPLN3, coxme sowie coxph, dass die Varianz  $s^2$  des zufälligen Effektes in den zu analysierenden Datensätzen Null ist?

Einen ersten Aufschluss ermöglicht die Betrachtung der medianen Abweichung von  $\hat{s^2}$  in Tabelle 10. Alle zuvor genannten Funktionen weisen Schätzer auf, welche sehr nahe bei Null liegen. Die mediane Abweichung ist sehr gering (nahe Null). Dabei gilt für %PCFrailty, dass es keine nennenswerten Unterschiede zwischen den Schätzern der Konfigurationen (4 Ereignisse pro Piece und 10 Ereignisse pro Piece) gibt. Bei der Betrachtung der Boxplots für  $\hat{s^2}$  in Abbildung 13 lässt sich deutlich ein Trend von Spannweite und IQR erkennen. Diese steigen mit wachsender Clusteranzahl. Genauso auffällig ist die geringere Streuung von  $\hat{s^2}$  unter SPLN3, während die anderen Funktionen recht identische IQR's aufweisen. Ebenso interessant sind die geschätzten Standardfehler, welche mit steigender Clusteranzahl ebenfalls ansteigen (Tabelle 10). PROC PHREG zeigt geschätzte Standardfehler von 0,014 bis 0,030. Damit liegen sie um die Hälfte unter den Werten der anderen Funktionen. Weiterhin gibt es kaum Unterschiede zwischen den geschätzten und empirischen Standardfehlern unter PROC PHREG. Anders verhält es sich bei den Standardfehlern für SPLN3 und %PCFrailty. Hier werden die geschätzten Standardfehler scheinbar zu hoch angegeben, da die empirischen Standardfehler deutlich geringer sind. Über alle Funktionen liegen die empirischen Standardfehler in einem Bereich von 0,013 bis 0,032. Da der Bias sehr gering über alle Funktionen ist und der MQF abhängig vom empirischen SF ist, liegen die Werte für alle MQF's nahe bei dem Wert Null. Unterschiede zwischen den Funktionen sind bei diesem Gütekriterium demnach nicht zu erkennen. Bei der Betrachtung der Überdeckungsraten liegen die gleichen Probleme vor wie bereits in Abschnitt 4.1.1 erläutert.

Zur Bewertung der Robustheit der Funktionen wird die Konvergenzrate genutzt (vgl. Tabelle 11). Keine Konvergenzprobleme zeigen sich bei den Funktionen coxph und coxme. Geringe Konvergenzraten von 88,55% - 95,90% liegen dahingegen bei der Funktion %PCFrailty vor. Wobei die Ergebnisse darauf schließen lassen, dass mit höherer Pieceanzahl eine geringere Robustheit vorliegt. SPLN3 sowie PROC PHREG zeigen recht stabile Konvergenzraten von ca. 100%. Eine Abweichung ist einzig bei der Funktion PROC PHREG zuerkennen. Diese scheint leichte Schwächen für Clustergrößen unterhalb von 50 zu besitzen.

Tabelle 11: Konvergenzraten de	r Analyse	von 2000	Datensätzen,	welche	$\mathbf{durch}$	ein
Proportionales Hazards Model	l erzeugt <sup>.</sup>	wurden.				

	(	Clustergröf	se
	50	25	10
PROC NLMIXED	$99,\!60\%$	$99{,}60\%$	$99{,}60\%$
PROC PHREG	$96,\!80\%$	$99,\!95\%$	$99,\!95\%$
%PCFrailty (4 Ereignisse pro Piece)	88,55%	$88,\!65\%$	$88,\!65\%$
%PCFrailty (10 Ereignisse pro Piece)	$95{,}80\%$	$95{,}90\%$	$95{,}90\%$
SPLN3	99,75%	$99,\!95\%$	$100,\!00\%$
coxme	$100,\!00\%$	$100,\!00\%$	$100,\!00\%$
coxph	$100{,}00\%$	$100,\!00\%$	$100,\!00\%$
PROC PHREG $(ohne \ Random)$	$100{,}00\%$	$100,\!00\%$	$100,\!00\%$

## 4.2.2 Ergebnisse der durch ein shared Log-normales Frailty-Modell generierten Daten

In diesem Abschnitt werden die Ergebnisse betrachtet, welche durch die Auswertungen von Simulationsdaten entstanden sind, die mit einem shared Log-normalen Frailty-Modell erzeugt wurden. Für die Auswertung wurden:

- 1. ein parametrisches shared Log-nomales Frailty-Modell (realisiert in PROC NLMIXED, mit Simulationsparametern als Startwerten), sowie
- 2. die in Abschnitt 4.2.1 unter den Punkten 2. 4. aufgeführten Modellansätze und Funktionen verwendet.

Alle Ergebnisse sind in Tabelle 12 und Abbildung 14 dargestellt. Zunächst werden die Ergebnisse der Schätzungen der festen Effekte  $\beta_1$  und  $\beta_2$  aus Tabelle 12 betrachtet. Die Schätzer für  $\beta_1$  weisen eine leichte Überschätzung für alle Funktionen und Clusteranzahlen auf. Die mediane Abweichung ist jedoch mit 0,007 sehr gering. Ähnliches

Tabelle 1	2: Simul	ationsergeb lusteranzahl = $2$ llustergröße = $5$ C N <sup>1</sup> = 1688	nisse von 20	00 Wie	derholu Clu Clu	ingen des a steranzahl = 4. istergröße = $2$ : $N^1 = 1737$	shared Log-r	iormalei	n Frailt <sup>Clus</sup>	y-Modelle. teranzahl = 10 stergröße = 10 $N^1 = 1711$	0
Wahrer Wert Media	n Ges.SF ]	Emp.SF MQF	Überd. <sup>2</sup>	Median	Ges.SF E	Imp.SF MQF	Überd. <sup>2</sup>	Median	Ges.SF E	mp.SF MQF	Überd. <sup>2</sup>
Parametrisches shared	Log-normales	s Frailty-Modell:	PROC NLMIXED								
$\beta_1$ 1,0 1,00	5 0,070	0,069 $0,005$	96,15	1,006	0,072	0,070 $0,005$	96, 37	1,005	0,076	0,076 $0,006$	94,92
$\beta_2$ -0,5 -0,45	6 0,250	0,126 $0,016$	100,00	-0,496	0,200	0,131 $0,017$	99,77	-0,504	0,162	0,128 $0,016$	98,54
$s^2$ 0,5 0,45	9 $0,149$	0,183 $0,034$	$84,54\/$ 93 07 <sup>3</sup>	0,470	0,135	$0,160 \ 0,026$	87,85 / 94.533	0,486	0,144	$0,155 \ 0,024$	$92,\!46\;/$ 94 04 $^3$
Semiparametrisches shi	wed Log-norr	males Frailty-Mo	dell: PROC PHRE	G (mit zuf.	älligem Efi	Fekt)	00/10				1 0(1 0
$\beta_1$ 1,0 1,00	4 0,069	0,067 $0,005$	95, 32	,1,003	0,070	$0,069 \ 0,005$	95,51	0,997	0,071	0,073 $0,005$	94,62
$\beta_2$ -0,5 -0,45	8 0,121	0,120 $0,014$	95,08	-0,497	0,122	0,127 $0,016$	94,30	-0,501	0,126	0,125 $0,016$	94,92
$s^2$ 0,5 0,48	1 0,186	0,183 $0,033$	89,87 / 96.80 <sup>3</sup>	0,463	0,142	0,147 $0,022$	$89.47$ / $96.49^3$	0,457	0,117	0,121 $0,016$	$90,18\ /$
4 Ereignisse pro Piece:	%PCFrailt	Λ					0+600				0
$\beta_1$ 1.0 1.00	7 0,069	0.068 0.005	95.20	1.006	0.070	0.069 $0.005$	95.45	1,001	0.071	0.073 $0.005$	94.51
$\beta_2$ -0,5 -0,49	9 $0,121$	$0,121 \ 0,015$	95,14	-0,498	0,122	0,127 $0,016$	94,19	-0,504	0,126	0,126 $0,016$	94,62
$s^2$ 0,5 0,48	5 0,192	0,184 $0,034$	$\frac{90,28}{22}/\frac{90}{22}$	0,468	0,147	0,149 $0,023$	90,27 / $72.00$	0,462	0,123	0,122 $0,016$	$\frac{91,70}{22}$
10 Fraignissa pro Piaca	"VPCErail	tv	97,10° / 83,59⁺				91,06° / 11,61*				98,U1~ / 70,95 <sup>±</sup>
$R_{i}$	- Mr OL EULE		08 90	1 005	0.060	0 060 0 005	08 30	1 000	0.071	0.072.0.005	04 69
$\beta_1$ 1,0 1,0 1,0 1,0 1,0 1,0 1,0 1,0 1,0 1,0	9 0,003	0,000 $0,000$ $0,015$ $0,015$	95,02	-0,497	0,003	0,127 $0,016$	94.13	-0.503	0,0.126	0.125 $0.016$	94.68
$s^2$ 0,5 0,48	$\frac{1}{4}$ 0,192	0,184 $0,034$	90,34 /	0,466	0,147	0,148 $0,022$	90,39 /	0,459	0,122	0,121 $0,016$	91,70 /
			$97,27^3$ / $83,41^4$				$96,95^3 \ / \ 77,49^4$				$97,95^3$ / $71,65^4$
Semiparametrisches shi	ared Log-nor	males Frailty-Mo	dell: SPLN3								
$\beta_1$ 1,0 1,00	4 0,070	0,067 $0,005$	95,56	1,003	0,070	0,069 $0,005$	95,68	1,003	0,073	0,073 $0,005$	94,86
$\beta_2$ -0.5 -0.45	$\frac{8}{2}$ 0,122	0,120 $0,014$	95,14	-0,497	0,123	0,127 $0,016$	94,30	-0.504	0,126	$0,126\ 0,016$	94,97 25.15 /
<i>s</i> <sup>2</sup> 0,5 0,40	i/ 0,183	0,187 0,035	$86,97$ / $96,56^3$	0,471	0,151	0,156 0,025	$90,39\ /\ 96,43^3$	0,482	0,135	0,138 0,019	$^{93,40}_{96,03^3}$
Semi parametrisches shi	pred Log-nor	males Frailty-Mo	dell: coxme								
$\beta_1$ 1,0 1,00	5 0,069	0,067 $0,005$	95,26	1,004	0,070	0,069 $0,005$	95,51	0,999	0,071	0,073 $0,005$	94,51
$\beta_2$ -0.5 -0.45	9 0,121	0,120 $0,014$	95,08	-0,498	0,122	0,127 $0,016$	94, 30	-0,502	0,126	0,125 $0,016$	94,92
$s^{z}$ 0,5 0,45		0,193 0,037	-	$0,\!484$		$0,160 \ 0,026$	ı	0,469	0,133	- 0,018	ļ
Semi parametrisches sh	ared Log-nor	males Franty-Mo	dell: coxpn				7				
β1 1,0 1,00 β1 β1	13 U,U09	0,007 0,000	95,32 or oc	1,002 0,405	0,0,0	0,009 0,005 0 197 0 016	10,09	1997	170,0 0.196	0,073 0,005 0 197 0 016	94,08 04.09
72 -0,0 -0,42 -2 0,7	0 0,121 7	0,120 0,014 0 173 0 031	80,UO	-0,4%0 0.470	0,144	010,0 141,0	94, <del>5</del> 0	100,0-	0,120	0,120 0,010 0.016	34,34
s - 0,0 - 0,4 Pronortionales Hazards	. Modell nach	U, LO U, LO U, UOL	RF.G.	U,4JU	1	U,140 U,U22	ı	0,±02	0,12U	- U,UIU	ı
$\beta_1$ 1.0 0.88	7 0.065	0.073 0.017	58.00	0.888	0.064	$0.069 \ 0.017$	57.34	0.882	0.064	0.067 $0.018$	54.12
$\beta_2$ -0,5 -0,44	0 $0,118$	0,123 $0,019$	90,46	-0,440	0,118	0,125 $0,020$	89,87	-0,441	0,118	0,118 $0,017$	91,58
<sup>1</sup> Anzahl der über alle	Funktionen k	convergierten Da	tensätze								
<sup>2</sup> Symmetrisches 95% <sup>3</sup>	Vald Konfide	inzintervall									
<sup>4</sup> 95% Profile-Likelihoo	wang wonne d-Konfidenzi	lenzhnervan ntervall									



Abbildung 14: Ergebnisse des shared Log-normalen Frailty-Modells; wahre Werte mittels horizontaler Linie dargestellt.

gilt auch für die Schätzungen bezüglich  $\beta_2$ . Dies ist auch in Abbildung 14 zu sehen. Die abgetragenen Mediane weichen kaum von den eingezeichneten wahren Werten ab. Ebenso gibt es keine Unterschiede zwischen den IRQ's bzw. den Spannweiten der einzelnen Funktionen. Bei Betrachtung der geschätzten Standardfehler ist ein leichter Trend zu erkennen (Tabelle 12). Mit steigender Clusteranzahl (fallender Clustergröße) kommt es zu einer geringen Erhöhung der Standardfehler um 0,005. Wie bereits in den vorhergehenden Abschnitten (4.1.1, 4.1.2 und 4.2.1) erwähnt, sind auch bei diesen Auswertungen die geschätzten Standardfehler unter  $\hat{\beta}_2$  (0,121 - 0,126) doppelt so hoch wie unter  $\hat{\beta}_1$ (0,069 - 0,072). Unerwartet ist jedoch, dass unter dem parametrischen Modell (hier als wahres Modell verstanden) die Standardfehler für  $\hat{\beta}_2$  sehr groß geschätzt werden. Diese Standardfehler liegen im Bereich von 0,16 - 0,25 und sind zum Teil doppelt so groß wie ihre empirischen Standardfehler, was zu einer großen Überdeckungsrate führt. Bei Betrachtung der anderen Funktionen ist diese große Differenz zwischen geschätzten und empirischen Standardfehlern nicht zu erkennen. Alle erhaltenen empirischen Standardfehler liegen in identischen Wertebereichen für die untersuchten Funktionen. Der mittlere quadratische Fehler (MQF), wie in Tabelle 12 zu erkennen, ist über alle Funktionen sehr ähnlich. Für die Schätzer von  $\beta_1$  nimmt der MQF einen Wert von durchschnittlich 0,005 an und für die Schätzer von  $\beta_2$  einen Wert von 0,016. Auch bei diesen Auswertungen liegen die erhaltenen Überdeckungsraten im erwarteten Bereich von 94% - 96%.

Beachtet man im Gegensatz dazu die Schätzungen unter PROC PHREG ohne RANDOM-Statement, fallen als erstes die starken Unterschätzungen auf. Besonders gilt dies für die Schätzungen von  $\beta_1$  welche eine mediane Abweichung zwischen 0,113 -0,118 aufweisen. Um die Hälfte geringer fallen die Schätzfehler bei  $\hat{\beta}_2$  aus (0,06). Das Verhalten der Standardfehler ist ähnlich dem im oberen Absatz besprochenem. Lediglich der clusterspezifische Trend spielt hier keine Rolle. Sowohl die Standardfehler als auch die erhaltenen MQF's sind größer als bei den für die Problemstellung angepassten Funktionen (Funktionen die einen zufälligen Effekt berücksichtigen). Weitere Anzeichen für die ungenauen Schätzungen sind die Überdeckungsraten. Im Fall von  $\hat{\beta}_1$ , liegen diese zwischen 54 % - 58 %. Geringer ist die Abweichung unter  $\hat{\beta}_2$ . Veranschaulicht werden diese Aussagen durch die dazugehörigen Boxplots in Abbildung 14.

Unterschiedliche Beobachtungen bezüglich der Schätzgüte treten bei Betrachtung der Analyseergebnisse von  $s^2$  auf. Je nach Funktion weisen die Schätzer unterschiedliche Tendenzen auf. Während unter PROC NLMIXED und SPLN3 die mediane Abweichung mit steigender Clusteranzahl geringer wird, ist bei den Funktionen PROC PHREG, %PCFrailty,

coxme und coxph ein gegenläufiger Trend zu erkennen (Tabelle 12). Dabei gibt es bei %PCFrailtykeine nennenswerten Unterschiede zwischen den Schätzern der Konfigurationen 4 Ereignisse pro Piece und 10 Ereignisse pro Piece. Über alle Funktionen betrachtet liegt ein Bias zwischen 0,003 bis 0,050 vor, wobei alle Funktionen zur Unterschätzung neigen. Diese Tendenz ist ebenfalls in den  $s^2$ -Boxplots von Abbildung 14 zu erkennen. Sowohl die IQR's als auch die Spannweiten sind annähernd identisch. Jedoch schwanken die Werte des unteren Quartils  $(Q_1)$  und oberen Quartils  $(Q_3)$  zwischen den Funktionen leicht. Etwas stärkere Variabilität zeigte sich bei PROC NLMIXED. Eine weitere Tendenz über alle Funktionen ist in Tabelle 12 bei den geschätzten Standardfehlern zu erkennen. Mit steigender Clusteranzahl (sinkende Clustergröße) fällt der Betrag der geschätzten Standardfehler. Dieser Effekt ist ebenso bei den empirischen Standardfehlern zu erkennen. Die Differenz zwischen den geschätzten und empirischen Standardfehlern ist damit sehr gering. Anders stellt sich die Situation unter der Funktion PROC NLMIXED dar. Die geschätzten Standardfehler unterliegen bezüglich der Clusteranzahl keinem erkennbaren Trend. Weiterhin weichen sie deutlich von den empirischen Standardfehlern ab. Dabei handelt es sich um Differenzen im Bereich von -0,034 bis -0,011. Da die empirischen Standardfehler der Funktion im selben Wertebereich liegen, wie die der anderen Funktionen, ist davon auszugehen, dass die Standardfehler unter PROC NLMIXED zu klein geschätzt werden. Aus den recht ähnlichen empirischen Standardfehlern resultieren auch die MQF's, welche in etwa identisch sind und demselben fallenden Trend wie die empirischen Standardfehler folgen. Die mittleren quadratischen Fehler liegen über alle Funktionen in einem Wertebereich von 0,016 - 0,037.

Bei der Begutachtung der Konfidenzintervalle für die Varianz  $s^2$  des zufälligen Effektes muss man drei Fälle separat betrachten. Die erste Möglichkeit der Angabe des Konfidenzbereiches ist das symmetrische Konfidenzintervall. Die Überdeckungsraten der Konfidenzintervalle liegen für alle Funktionen im Mittel unter dem erwarteten Bereich von 94% – 96%. Dabei zeigt die Funktion PROC NLMIXED die geringsten Überdeckungsraten. Bei den asymmetrischen Konfidenzintervallen liegen die Überdeckungsraten von PROC NLMIXED im Idealbereich, wärend die der anderen Funktionen mit einem Prozentpunkt darüber liegen. Die Funktion %PCFrailty liefert neben diesen beiden Typen auch das Profile-Likelihood-Konfidenzintervall. Dessen Überdeckungsraten sind jedoch sehr gering mit Werten zwischen 70,95% und 83,59%.

Weiterhin lässt sich über die interessierenden Funktionen sagen, dass sie zu 100% konvergieren. Ausnahme bildet wie schon in Abschnitt 4.2.1 die Funktion %PCFrailty. Je nach Piece-Anzahl zeigt sie eine Konvergenzrate zwischen 88% - 95,1%. Tabelle 13: Konvergenzraten der Analyse von 2000 Datensätzen, welche durch ein shared Log-normales Frailty-Modell erzeugt wurden.

	(	Clustergröf	se
	50	25	10
PROC NLMIXED	99,25%	$99,\!95\%$	99,70%
PROC PHREG	$100,\!00\%$	$100,\!00\%$	$100,\!00\%$
%PCFrailty (4 Ereignisse pro Piece)	$88,\!00\%$	$89,\!40\%$	$87,\!90\%$
%PCFrailty (10 Ereignisse pro Piece)	94,55%	$94{,}80\%$	$95,\!10\%$
SPLN3	99,20%	$100,\!00\%$	$100,\!00\%$
coxme	$100,\!00\%$	$100,\!00\%$	$100,\!00\%$
coxph	$100,\!00\%$	$100,\!00\%$	$100,\!00\%$
PROC PHREG (ohne Random)	$100,\!00\%$	$100,\!00\%$	$100,\!00\%$

## 4.3 Ergebnisse der Halluca-Studie

Zunächst wurden die Daten der Halluca-Studie durch ein Proportionales Hazards Modell nach Cox analysiert. Verwendet wurde hierfür die Funktion PROC PHREG. Die erhaltenen Ergebnisse sind in Tabelle 14 dargestellt.

		PROC	PHREG
Parameter	Schätzer	SF	KI1
Alter	0,009	0,003	(0,004;0,015)
Geschlecht	-0,162	$0,\!071$	(-0,302;-0,022)
$\operatorname{nsclc}$	$-0,\!125$	$0,\!067$	(-0,256 ; 0,006)
ECOG $3\&4$	$0,\!610$	0,104	$(0,408\ ;\ 0,813)$
TNM II	$0,\!455$	0,169	$(0,124\ ;\ 0,785)$
TNM IIIa	$0,\!594$	0,131	$(0,337\ ;\ 0,850)$
TNM IIIb	$0,\!947$	0,120	$(0,712 \ ; \ 1,183)$
TNM IV	$1,\!345$	0,110	$(1,130\ ;\ 1,561)$

Tabelle 14: Auswertung der Halluca Daten mittels des Cox-Modells.

 $^1$  Symmetrisches 95% Wald Konfidenzintervall

Der einzige stetige Einflussfaktor in der Auswertung ist die Kovariable Alter (in Jahren). Für die Patienten der Halluca-Studie gilt, dass pro zusätzlichem Lebensjahr das Risiko an Lungenkrebs zu versterben um den Faktor 1,009 ( $e^{0,009}$ ) steigt. Die Frauen haben im Vergleich zu den Männern ein auf das 0,850-fach reduzierte Risiko. In der

Halluca-Studie weisen die nicht kleinzelligen Karzinome ein auf das 0,883-fach verringerte Risiko auf, im Vergleich zu kleinzelligen Karzinomen. Jedoch deutet das dazugehörige Konfidenzintervall auf einen nicht eindeutigen Effekt hin. Einen sehr großen Einfluss auf das Überleben des Patienten hat der Selbstständigkeitsstatus. Patienten mit einem ECOG Status von über 2 haben ein 1,841-faches Risiko an Lungenkrebs zu versterben im Vergleich zu Patienten, welche einen Selbstständigkeitsstatus von 0 bis 2 aufweisen. Erhöhte Risiken sind auch bei den angegebenen Tumorstadien zu erkennen. In Tabelle 14 ist zu erkennen, dass mit steigendem Tumorstadium das Sterberisiko eines Patienten wächst. Dabei ist das Sterberisiko bei einem Patienten mit TNM II-Stadium das 1,576-fache ( $e^{0,455}$ ) im Vergleich zu einem Patienten in Stadium TNM I. Im höchsten Stadium (TNM IV) weisen die Patienten das 3,839-fache des Risikos von Patienten der Referenzgruppe TNM I auf.

Im Weiteren wurde bei der Analyse der Halluca-Daten ein Effekt der Diagnoseeinheiten angenommen. Diese Annahme stützt sich auf Abbildung 10, durch welche die unterschiedlichen Lebensdauerverteilungen in Abhängigkeit von der Clusterzugehörigkeit dargestellt sind. In den folgenden Abschnitten wurden zwei Modelle zur Analyse der Halluca-Daten verwendet. Zum einen wurde der Clustereffekt mittels einer Gammaverteilten Frailty modelliert. Dabei wurden die Funktionen coxph und SPGAM1 genutzt. Zum anderen wurde für die Anpassung an die Daten ein shared Log-normales Frailty-Modell verwendet. Für diese Auswertung wurden die Funktionen PROC PHREG, coxme und %PCFrailty (mit 10 Ereignissen pro Piece) verwendet.

## 4.3.1 Ergebnisse des shared Gamma Frailty-Modells

Betrachtet werden in diesem Abschnitt die Ergebnisse aus der Halluca-Studie, welche durch die Funktionen coxph und SPGAM1 erhalten wurden. Die Ergebnisse werden in Tabelle 15 dargestellt. Sowohl der PPL- als auch der EM-Algorithmus sind für die Halluca-Daten konvergiert. Bei erster Betrachtung fällt auf, dass alle angegebenen Schätzer nahezu identisch sind, wobei es zu kleinen Unterschieden zwischen den Schätzern der Clustervarianz  $\hat{\sigma}^2$  kommt. Die Bewertungen der Schätzer für die festen Effekte sind über beide Funktionen gleich. Bei der Interpretation von festen Effekten in einem shared Frailty-Modell ist zu beachten, dass die Regressionsparameter für die Korrelation innerhalb der Cluster adjustiert sind.

Wie in Tabellen 14 und 15 zu erkennen ist, sind die Ergebnisse der festen Effekt-

schätzer leicht verschieden. Im Allgemeinen tendieren die Schätzer des shared Gamma Frailty-Modells zur Null. Die Schätzer der festen Effekte sind in diesem Modell dem Betrag nach kleiner im Vergleich zum Cox-Modell (Tabelle 14).

Für die Auswertung wurden die diagnostizierenden Einheiten als Clustervariable verwendet. Man geht also davon aus, dass es Unterschiede im Patientengut zwischen den einzelnen Diagnoseeinrichtungen gibt. Dieser Clustereffekt wird mit Hilfe der Frailtyvarianz  $\sigma^2$  charakterisiert. Unter coxph besitzt  $\hat{\sigma}^2$  den Wert 0,311 mit einem Profile-Likelihood-Konfidenzintervall von 0, 133 – 0, 588. Diese Frailtyvarianz von 0,311 ist ein Maß für die Korrelation der Lebensdauern der Lungenkrebspatienten innerhalb der einzelnen Diagnoseeinheiten. Bei der Betrachtung der Ergebnisse von SPGAM1 fällt auf, dass die Heterogenität der Diagnoseeinheiten hier einen Wert von 0,294 annimmt, also etwas kleiner ist als bei der Schätzung mittels coxph. Das dazugehörige asymmetrische Wald-Konfidenzintervall reicht von 0,158 bis 0,722 und ist damit etwas beiter als das Profile-Likelihood-Konfidenzintervall der Funktion coxph.

Bei Betrachtung beider Analysemöglichkeiten ist zu erkennen, dass ein signifikant von Null verschiedener Clustereffekt, in den Daten der Halluca-Studie vorhanden ist. Zwischen den Diagnoseeinheiten tritt eine Frailtyvarianz von 0,3 auf.

		со	xph		SPG	AM1
Parameter	Schätzer	SF	$\mathrm{KI}^1$	Schätzer	SF	KI <sup>1</sup>
Alter	0,009	0,003	(0,003;0,014)	0,009	0,003	(0,003;0,014)
Geschlecht	-0,158	$0,\!073$	(-0,301;-0,016)	-0,159	$0,\!073$	(-0,301;-0,016)
$\operatorname{nsclc}$	-0,097	0,068	(-0,230;0,036)	-0,097	0,068	$(-0,230 \ ; \ 0,036)$
ECOG 3&4	$0,\!600$	0,108	$(0,\!389;0,\!812)$	$0,\!600$	$0,\!108$	$(0,\!388\ ;\ 0,\!811)$
TNM II	$0,\!450$	0,170	$(0,\!116~;~0,\!783)$	$0,\!450$	$0,\!170$	$(0,116\ ;\ 0,783)$
TNM IIIa	$0,\!600$	$0,\!133$	$(0,\!339\ ;\ 0,\!861)$	$0,\!600$	$0,\!133$	$(0,\!339~;0,\!861)$
TNM IIIb	$0,\!957$	0,123	$(0,715\ ;\ 1,198)$	$0,\!956$	$0,\!123$	$(0,715\ ;\ 1,197)$
TNM IV	$1,\!314$	0,113	$(1,\!093~;~1,\!535)$	$1,\!314$	$0,\!112$	$(1,\!093\ ;\ 1,\!534)$
$\sigma^2$	0,311	-	$(0,133;0,588^2)$	0,294	$0,\!110$	$(0,158;0.722^3)$

 

 Tabelle 15: Auswertung der Halluca Daten mittels des shared Gamma Frailty-Modells.

 $^1$  Symmetrisches 95% Wald Konfidenzintervall

 $^2$ 95% Profile-Likelihood-Konfidenzintervall

<sup>3</sup> Asymmetrisches 95% Wald Konfidenzintervall

### 4.3.2 Ergebnisse des shared Log-normalen Frailty-Modells

In diesem Abschnitt wird die Auswertung der Halluca-Daten mittels eines shared Lognormalen Frailty-Modells betrachtet. Dazu wurden die Funktionen PROC PHREG, %PCFrailty mit 10 Ereignissen pro Piece und coxme verwendet. Bei allen drei Funktionen konvergierten die verwendeten Algorithmen. Wie auch bei der Auswertung durch ein shared Gamma Frailty-Modell wurden Alter, Geschlecht, Tumortyp, Selbständigkeitsstaus und Tumorstadium als feste Effekte in die Analyse einbezogen. Die Clustervariable bildete wie im vorherigen Abschnitt die diagnostizierende Einheit.

In Tabelle 16 ist zu erkennen, dass die Parameterschätzer mit denen aus Tabelle 15 übereinstimmen. Dieselbe Aussage lässt sich auch auf die Standardfehler und die daraus resultierenden Konfidenzintervalle übertragen. Geringe Abweichungen treten bei den Schätzern der Funktion %PCFrailty auf, wohingegen die Standardfehler identisch sind.

Die Schätzer für die Varianz des zufälligen Effekts  $s^2$  liegen im Bereich von 0,293 bis 0,303 und überdecken damit denselben Wertebereich wie die Schätzer  $\hat{\sigma}^2$  des shared Gamma Frailty-Modells. Durch die Funktion PROC PHREG ergibt sich bei einem symmetrischen Konfidenzintervall von 0,169 - 0,628 ein Wert für die Varianz des zufälligen Effektes von 0,293. Dieses Konfidenzintervall ist das Breiteste der hier dargestellten. Einen etwas größeren Schätzer mit schmalerem asymmetrischen Konfidenzintervall liefert die Funktion %PCFrailty. Diese gibt im Gegensatz zu den anderen hier angesprochenen Funktionen neben dem Standardfehler auch asymmetrische und Profile-Likelihood-Konfidenzintervalle an. Die R-Funktion coxme liefert einen Schätzwert für die Varianz des zufälligen Effekts von 0,295, gibt jedoch keinen Standardfehler an. Nach der von Therneau und Grambsch [21] beschriebenen Methode ist es jedoch möglich, sich das Profile-Likelihood-Konfidenzintervall ausgeben zu lassen, welches den Bereich von 0,151 bis 0,594 abdeckt. Alle Konfidenzintervalle schließen die Null nicht ein, was für einen signifikanten Effekt spricht.

Tabelle 16: Auswertung der Halluca Daten mittels des shared Log-normalen Frailty-Modells.									
	PROC PHREG			%PCFrailty			coxme		
		(10 Ereignisse pro Piece)							
Parameter	Schätzer	SF	$\mathrm{KI}^{1}$	Schätzer	SF	$\mathrm{KI}^{1}$	Schätzer	SF	$\mathrm{KI}^{1}$
Alter	0,009	0,003	(0,003;0,014)	0,009	0,003	(0,003 ; 0,015)	0,009	0,003	(0,003;0,014)
Geschlecht	-0,159	$0,\!073$	(-0,302;-0,017)	-0,162	$0,\!073$	(-0,304;-0,020)	-0,159	$0,\!073$	(-0,301;-0,017)
$\operatorname{nsclc}$	-0,097	0,068	(-0,230; 0,036)	-0,100	0,068	$(-0,233 \ ; \ 0,033)$	-0,097	0,068	$(-0,230 \ ; \ 0,036)$
ECOG 3&4	$0,\!605$	0,108	$(0,394\ ;\ 0,816)$	$0,\!606$	0,108	$(0,395\ ;\ 0,818)$	$0,\!605$	0,108	$(0,\!394\ ;\ 0,\!816)$
TNM II	$0,\!450$	$0,\!170$	$(0,116 \ ; \ 0,783)$	$0,\!446$	$0,\!170$	$(0,113\ ;\ 0,779)$	$0,\!450$	$0,\!170$	$(0,\!116~;0,\!783)$
TNM IIIa	0,599	$0,\!133$	$(0,\!338\ ;\ 0,\!860)$	$0,\!601$	$0,\!133$	$(0,340 \ ; \ 0,862)$	0,599	$0,\!133$	$(0,\!338\ ;\ 0,\!860)$
TNM IIIb	$0,\!956$	$0,\!123$	$(0,715\ ;\ 1,197)$	$0,\!961$	0,123	$(0,720 \ ; \ 1,201)$	$0,\!956$	$0,\!123$	$(0,715\ ;\ 1,197)$
TNM IV	$1,\!314$	$0,\!113$	$(1,\!093~;1,\!535)$	$1,\!316$	0,113	$(1,095\ ;\ 1,537)$	$1,\!314$	$0,\!113$	$(1,\!093~;1,\!535)$
$s^2$	0,293	$0,\!096$	$(0,169;0,628^2)$	$0,\!303$	$0,\!103$	$(0,171;0,673^2)$	$0,\!295$	-	$(0,151;0,594^3)$
						$(0,150;0,600^3)$			

<sup>1</sup> Symmetrisches 95% Wald Konfidenzintervall
 <sup>2</sup> Asymmetrisches 95% Wald Konfidenzintervall
 <sup>3</sup> 95% Profile-Likelihood-Konfidenzintervall

# 5 Diskussion

Diese Arbeit beinhaltet umfassende Simulationen, um die Güte der unterschiedlichen Funktionen, Prozeduren und Makros zur Analyse von shared Frailty-Modellen miteinander zu vergleichen. Im Folgenden werden diese Funktionen, Prozeduren und Makros unter dem Begriff **Funktionen** zusammengefasst. Für den Vergleich wurden unterschiedliche Validierungsmethoden verwendet. Nach Untersuchung aller Ergebnisse und genauerer Betrachtung der verwendeten Funktionen wurden die Besten auf die Daten der Halluca-Studie angewandt.

Für die vorliegende Arbeit können folgende Aussagen getroffen werden. Alle Funktionen sind für die Analyse von shared Gamma und Log-normalen Frailty-Modellen geeignet. Die Handhabung ist einfach und erfordert keine speziellen Kenntnisse. Für beide Modelle gilt, dass die Schätzungen der festen Effekte für alle Funktionen nah beieinander liegen, bei minimalen Abweichungen. Dies trifft auch bei unterschiedlicher Clustergröße zu. Bei den Schätzungen der festen Effekte weisen die Punktschätzer  $\hat{\beta}_2$ eine höhere Variabilität als  $\hat{\beta}_1$  auf, was in den Abbildungen 11, 12, 13 und 14 zu sehen ist. Erklärt werden kann dies durch die unterschiedlichen Skalenniveaus der Variablen  $x_1$  (metrisch) und  $x_2$  (binär). Da metrische Variablen mehr Informationen beinhalten als binäre Variablen und damit deren Schätzung präziser ist, sind die in der Arbeit auftretenden Ergebnisse plausibel. Weiterhin deuten die geringen Unterschiede zwischen geschätzten und empirischen Standardfehlern darauf hin, dass die Funktionen die Standardfehler gut schätzen.

Unterschiede zwischen den Funktionen konnten bei der Schätzung der Varianz der Frailty bzw. des zufälligen Effektes beobachtet werden. Dabei ist es egal, ob man die Ergebnisse aus dem Cox-Modell (Varianz gleich 0) oder dem shared Frailty-Modell (Varianz gleich 0,5) untersucht. Bei der Analyse der Daten, welche durch ein Cox-Modell generiert wurden, liegen  $\hat{\sigma}^2$  beziehungsweise  $\hat{s}^2$  nahe dem wahren Wert Null und demzufolge auch nahe der unteren Grenze des definierten Parameterraums von  $\sigma^2$ beziehungsweise  $s^2$ . Die Ergebnisse aus den Tabellen 6 und 10 zeigen, dass die untersuchten Funktionen PROC PHREG, SPGAM1, SPLN1, %PCFrailty, frailtyPenal und coxph eine nicht vorhandene Frailtyvarianz detektieren können. Auffällig bei den Ergebnissen war ein Anwachsen der Variabilität der Varianzschätzer mit steigender Clusteranzahl. Der gegenläufige Effekt ist im Fall des shared Frailty-Modells zu beobachten. Dieser Effekt ist nachvollziehbar, da im Cox-Modell alle Beobachtungen zu einem großen Cluster, mit
der Frailty eins und der Varianz null, gehören. Demzufolge nimmt die Variabilität mit höherer Clusteranzahl und den damit berechneten Frailty-Werten zu. Im Frailty-Modell ist die Situation eine andere. Die Frailtyvarianz bzw. die Varianz des zufälligen Effektes kann mit größerer Clusteranzahl präziser geschätzt werden. Vergleicht man die Simulationsergebnisse der unterschiedlichen Funktionen miteinander, sollte beachtet werden, dass der EM- und PPL-Algorithmus im shared Gamma Frailty-Modell zu denselben Ergebnissen führen [12], [36]. In anderen Frailty-Modellen gilt dies im Allgemeinen jedoch nicht. Zu beachten ist, dass die Schätzgüte für die Frailtyvarianz  $\hat{\sigma}^2$  und die Varianz des zufälligen Effektes  $\hat{s}^2$  sich bei diesen Daten schwerer bewerten lässt, da der wahre Wert der Varianz  $\sigma^2 = 0$  beziehungsweise  $s^2 = 0$  ist und damit genau die Grenze des Parameterraums darstellt. Die Bewertung des MQF's sowie die Verwendung von geeigneten Konfidenzintervallen sind dadurch problematisch.

Als beste Funktionen zur Schätzung der Parameter des shared Gamma Frailty-Modells haben sich coxph und SPGAM1 erwiesen. Die Funktion coxph ist sehr schnell und konvergiert im Großteil der Fälle. Sie weist einen minimalen Fehler bei unterschiedlicher Clusteranzahl auf. Jedoch müssen die Kontrollparameter für die Genauigkeit und die maximale Iterationsanzahl berücksichtigt werden (siehe Tabelle 3). Ein großer Nachteil für die praktische Verwendung dieser Funktion ist, dass sie keine Schätzung des Standardfehlers der Frailtyvarianz zur Verfügung stellt. Es besteht jedoch die Möglichkeit nach der Methode von Therneau und Grambsch [29] ein Profile-Likelihood-Konfidenzintervall zu berechnen. Damit ist die Einschätzung des Varianzschätzers  $\hat{\sigma^2}$ möglich. Als Alternative zur Funktion coxph kann das Makro SPGAM1 angesehen werden. Im Gegensatz zu coxph stellt es den Standardfehler für den Schätzer der Frailtyvarianz bereit. Sowohl SPGAM1 als auch die Funktion frailtyPenal benötigen mehr Zeit für die Datenanalyse. Die Funktion frailtyPenal bietet die Möglichkeit die Rechenzeit etwas zu variieren, indem man die Anzahl der Knoten in n.knots verringert. Laut Rondeau [31] verschlechtern sich die Maximum-Penalized-Likelihood Schätzer  $\hat{\lambda}_0(t), \hat{\beta}, \hat{\theta}$  mit sinkender Knotenzahl nicht. Weiterhin muss der Glättungsparameter (kappa1) angegeben werden. In jedem Durchlauf ist es notwendig die beste Einstellung zu finden, um eine Konvergenz zu erreichen. Dies macht frailtyPenal zu einer schwerer zu handhabenden Funktion. Bei einer Simulationsstudie ist es notwendig, für jeden Datensatz das beste Stetting von kappa1 zu verwenden. Dies wird mit dem Argument cross.validation erreicht. Dieses Argument schätzt den Glättungsparameter separat für jeden Datensatz. Der Vorteil der Funktion frailtyPenal ist die Berechnung des Standardfehlers für die Frailtyvarianz. Ein weiterer positiver Aspekt ist die Fähigkeit, mit komplizierteren

Frailty-Modellen umzugehen (bspw. nested- und joint-Frailty-Modelle [32]), sowie die Ausgabe der geglätteten Basishazardfunktion.

Für das shared Log-normale Frailty-Modell liefert die Funktion coxme die Schätzungen mit dem geringsten Bias. Sie ist der Funktion coxph vorzuziehen, da sie eine größere Klasse an Modellen neben dem shared Log-normalen Frailty-Modell analysieren kann. Für beide R-Funktionen ist es ratsam, die Anzahl der Iterationen und die Genauigkeit vorher festzulegen. Die Funktion PROC PHREG ist eine Standardprozedur innerhalb von SAS (seit SAS 9.3). Durch das Random-Statement ist es möglich shared Log-normale Frailty-Modelle zu analysieren. Neben einem Punktschätzer für die Varianz des zufälligen Effektes liefert sie ebenfalls den dazugehörigen Standardfehler. Gleiches gilt für das von Vu [26] entwickelte Makro SPLN3. Jedoch zeigt die Abbildung 14 das SPLN3 leichte Schwächen bei einer geringen Clusteranzahl (Clustergröße 50) hat. Diese Aussage deckt sich mit denen der Simulationsstudie von Vu [26]. Darin weist die Autorin darauf hin, dass der ML-EM-Algorithmus bei der Schätzung von  $s^2$  unterschiedlich auf den wahren Wert dieser Varianz, den Stichprobenumfang und die Zensierungswahrscheinlichkeit reagiert. Dahingegen zeigen PROC PHREG und %PCFrailty diese Schwächen bei einer großen Clusteranzahl (Clustergröße 10). Die Ergebnisse von %PCFrailty konnten die Ergebnisse von Feng et al. [24] nicht vollständig replizieren. Dies könnte an der Struktur der Simulationsdaten beider Studien liegen. Während Feng et al. [24] eine exponentialverteilte Basishazardfunktion bevorzugten, wurde in dieser Arbeit eine für praktische Anwendungen realistischere Gompertz-Verteilung zur Simulation der Daten gewählt. Zudem benötigen die Funktionen PROC PHREG, coxph und coxme nur wenige Sekunden pro Durchlauf, wohingegen die Funktionen %PCFrailty und SPLN3 wesentlich mehr Berechnungszeit benötigen. Die Ergebnisse der Funktion %PCFrailty zeigen, dass ein Verallgemeinertes gemischtes Modell mit binären Response, einer Poisson Likelihood und einem spezifischen Offset geeignet ist, um gruppierte Lebensdauerdaten zu analysieren. Dabei liefern Auswertungen mit 10 Ereignissen pro Piece bereits adäquate Ergebnisse. Für den Nutzer ergeben sich folgende Hinweise bei der Anwendung von %PCFrailty: Es ist nicht erforderlich eine möglichst große Anzahl an Pieces für die Modellierung der Basishazardfunktion zu verwenden. Die Ergebnisse der Simulationsstudie zeigten, dass es keinen Schätzvorteil von %PCFrailty mit 4 Ereignissen pro Piece im Vergleich zu %PCFrailty mit 10 Ereignissen pro Piece gibt. Empfehlenswert ist es, genügend Ereignisse pro Piece zu definieren (npieces = CoxXX (XX = Ereignisse pro Piece)).

Da alle Funktionen robuste und gute Schätzungen für shared Gamma- bzw. Lognormale Frailty-Modelle liefern, ist es erforderlich für eine Empfehlung weitere Vor- und Nachteile der Funktionen zu berücksichtigen. Für ein shared Gamma Frailty-Modell ist es sinnvoll die Funktionen coxph bzw. SPGAM1 zu verwenden. Die Funktion coxph weist keine Konvergenzprobleme auf und lieferte die Ergebnisse mit der geringsten medianen Abweichung. Dennoch liefert sie keine Schätzung des Standardfehlers der Frailtyvarianz. Jedoch ist es möglich die PL-Konfidenzintervalle zu berechnen [21]. Die Funktion SPLN3 weist eine große Variabilität in den Schätzungen auf, wenn die Frailtyvarianz Null ist. Jedoch liefert sie einen Schätzer für den Standardfehler. Beide Funktionen sind frailtyPenal vorzuziehen, da es bei dieser Funktion zu einer Verzerrung bei der Schätzung der festen Effekte kommt. Bei den Funktionen, die der Auswertung eines Log-normalen Frailty-Modells dienen, fällt die Entscheidung schwerer. Auch wenn coxme keine Schätzung des Standardfehlers liefert, weist es im Mittel die geringste Verzerrung der Schätzung von  $s^2$  auf. Weiterhin besteht in coxme, genau wie in coxph. die Möglichkeit Profile-Likelihood-Konfidenzintervalle zu berechnen [21]. Ebenso ist die Funktion PROC PHREG einzuschätzen. Da sie eine Standardfunktion in SAS ist und eine robuste, gute Schätzung liefert, ist sie für SAS-Nutzer vorzuziehen. Als Alternative mit zahlreichen Auswertungsmöglichkeiten ist %PCFrailty zu sehen. Der Modellansatz erlaubt eine variable Gestaltung der Basishazardfunktion durch verschiedene Definitionen der Pieces. Sie bietet weiterhin die Möglichkeit unterschiedliche Schätzverfahren zu verwenden, wie Gauß-Quadratur, Laplace-Approximation oder verschiedene Pseudo-Likelihood-Verfahren. Außerdem können unterschiedliche Strukturen des zufälligen Effektes mit Hilfe der Kovarianzmatrix modelliert werden. Verschiedene Hypothesentests bezüglich Varianz und Kovarianz des zufälligen Effektes sind ebenso enthalten. Im Zuge dessen beinhaltet die Funktion %PCFrailty drei Konfidenzintervalltypen; das asymmetrische Wald-Konfidenzintervall, das Profile-Likelihood-Konfidenzintervall und das Geschätzte-Likelihood-Konfidenzintervall. Auf letzteres wurde in dieser Arbeit nicht eingegangen.

Im ersten Teil der Arbeit wurden Simulationsdaten erstellt, um mit deren Hilfe die Güte der Funktionen zu bewerten. Zu beachten ist hier, dass die Simulationsdaten lediglich als Abbild der Realität verstanden werden können. Die zu analysierenden Daten sind also idealisiert und die betrachteten Funktionen erzielten unter diesen Umständen bessere Ergebnisse als bei einem realen Datensatz zu erwarten wären. Jedoch bieten Simulationsstudien die Möglichkeit, unterschiedliche Szenarien (Clusteranzahl, Clustergröße, Zensierungsrate, unterschiedliche feste Effekte und Größe der Frailtyvarianz) zu betrachten und somit Limitationen von Funktionen zu erkennen. Da in einer Simulationsstudie die wahren Werte der interessierenden Parameter bekannt sind, lassen sich in einer Simulationsstudie Rückschlüsse über die Schätzgüte treffen, was bei alleiniger Betrachtung von realen Datensätzen nicht möglich ist. Eine Limitation der vorliegenden Simulationsstudie liegt darin, dass ausschließlich SAS und R Funktionen zur Analyse verwendet wurden. STATA und WinBUGS wurden ausgeschlossen, da sie parametrisch arbeiten und der Fokus der vorliegenden Arbeit auf semiparametrischen Modellen liegt. Anhand der Ergebnisse zeigte sich, dass keine Unterschiede zwischen den Punktschätzern und deren Standardfehlern bei parametrischen (PROC NLMIXED) und semiparametrischen Modellansätzen zu erkennen sind. Dies stellt den Schätzvorteil, im Falle einer bekannten Basishazardfunktion, von parametrischen Modellansätzen gegenüber semiparametrischen Modellansätzen in Frage. Nicht alle Prozeduren die shared Gamma und Log-normale Frailty-Modelle analysieren können, wurden hier besprochen. Beispielsweise ist bei dem Makro Gamfrail die Rechenzeit zu lang, um in eine Simulationsstudie integriert zu werden. Eine andere Möglichkeit, Frailty-Modelle zu bearbeiten, liefert die R-Funktion phmm. Jedoch muss hier der Anwender selbst über das Erreichen der Konvergenz entscheiden [40]. Zu beachten ist, dass R-Pakete und deren Funktionen häufigen Änderungen unterliegen können. Zu Änderungen innerhalb von Paketen kommt es entweder durch die Behebung von Fehlern oder durch die Erweiterung von Funktionen.

Kritisch betrachtet werden muss auch der mittlere quadratische Fehler (MQF) für die Varianz der Frailty  $\sigma^2$  beziehungsweise des zufälligen Effektes  $s^2$ . Er unterliegt der Annahme, dass die Punktschätzer einer symmetrischen Verteilung folgen. Für die Auswertung der Daten die durch ein shared Frailty-Modell generiert wurden (Abschnitte 4.1.2 und 4.2.2) scheint diese Annahme erfüllt zu sein. Jedoch sind bei den Ergebnissen des Cox-Modells rechtsschiefe Verteilungen der Punktschätzer  $\hat{\sigma}^2$  und  $\hat{s}^2$  zu erkennen, wodurch der MQF weniger für die Bewertung der Schätzgüte geeignet ist.

Die Daten der Halluca-Studie wurden mit Hilfe der Funktionen coxph, SPGAM1 sowie PROC PHREG, %PCFrailty und coxme ausgewertet. Auf Grund der unterschiedlichen medianen Überlebenszeiten innerhalb der diagnostizierenden Einheiten, zu sehen in Abbildung 10, wurde bei den Daten der Halluca-Studie ein Clustereffekt angenommen. Wie im Artikel von Kuß et al. [51] erwähnt, ist eine dieser diagnostizierenden Einheiten die Pathologie der Uniklinik Halle. Die dort erhobenen Daten waren Daten von Totenscheinen (21 DCO's), wodurch sich vermuten lässt, dass die in diesem Zentrum beobachteten Überlebenszeiten sich stark von den anderen 55 Zentren unterscheiden. Wenn dieses Zentrum aus der Auswertung ausgeschlossen wird, sinkt die Frailtyvarianz  $\hat{\sigma^2}$  der Halluca-Studie von 0,302 auf ca. 0,1. Ähnlich verhält es sich bei Annahme eines shared Log-normalen Frailty-Modells. Die Varianz des zufälligen Effektes  $\hat{s^2}$  verringert sich von 0,297 auf ca. 0,080.

Die Vergleichbarkeit der Werte der Log-Likelihoodfunktion ist nur bedingt gegeben. Grund hierfür sind zum einen die unterschiedlichen Schätzalgorithmen (Abschnitt 3.1.3) und zum anderen die variierenden Datensätze. Bei den Funktionen coxme und coxph sowie der Prozedur PROC PHREG und dem Makro SPGAM wird der Originaldatensatz in die Analyse aufgenommen. Anders verhält es sich bei dem Makro %PCFrailty. Aufgrund der Auffaltung entstehen je nach Piecedefinition unterschiedlich große Datensätze, wodurch eine Betrachtung der Log-Likelihoodfunktion über alle verwendeten Funktionen nicht möglich ist.

Die in den Tabellen 15 und 16 dargestellten festen Effekte weichen nur leicht voneinander ab. Beide Modellansätze, Gamma- und Log-normale shared Frailty, führen zu gleichen Schätzungen der festen Effekte. Wenn die Schätzung der festen Effekte im Fokus der Auswertung steht, ist folglich die Verteilung der Frailty nicht von primärer Bedeutung. Auf die Modellierung des Clustereffektes sollte jedoch nicht verzichtet werden, da es sonst zu einer Unterschätzung der festen Effekte kommt (vgl. Abbildungen 12 und 14).

# 6 Zusammenfassung

Ist es das Ziel den Einfluss von prognostischen Faktoren bzw. Risikofaktoren auf die Ereigniszeit zu ermitteln, so sind Methoden aus dem Bereich der Lebensdaueranalyse erforderlich. Die Modelle aus diesem Bereich der Statistik berücksichtigen das Auftreten von Zensierungen (unvollständige Beobachtungen). Ein Basismodell stellt dabei das von Cox entwickelte Proportionale Hazardsmodell dar [1]. Wie jedes mathematische Modell unterliegt auch das Cox-Modell Annahmen. Eine dieser Annahmen ist die Proportionalität der Hazards. Doch diese Annahme kann verletzt sein. Ein möglicher Grund ist das Vorhandensein von Gruppeneffekten. Ein solcher Effekt liegt vor, wenn die Ereigniszeiten innerhalb einer Gruppe sich ähnlicher sind als die Zeiten unterschiedlicher Gruppen. Zur Analyse von gruppierten Lebensdauerdaten liegen unterschiedliche Realisierungen in Softwarepaketen vor. Unklar ist jedoch wie genau die Schätzungen der Funktionen sind, da noch keine vergleichenden Validierungsstudien durchgeführt wurden. Das Ziel der Arbeit ist es, diese methodische Lücke zu schließen. Es gibt in der Lebensdaueranalyse verschiedene Modellansätze, um gruppierte Ereigniszeitdaten zu analysieren (Copula-Modelle, marginale Modelle, Verallgemeinerte lineare gemischte Modelle, stratifizierte Modelle und Frailty-Modelle). Soll die Frage beantwortet werden, wie groß die Variabilität zwischen den Gruppen ist, können shared Frailty-Modelle (Gamma und Log-normal) und alternativ gemischte lineare Modelle verwendet werden. Umgesetzt wurden diese Modelle in vielen bekannten Statistikpaketen. In dieser Arbeit wird auf die Statistikpakete SAS und R eingegangen. Genutzt wurden die Funktionen:

- SAS: SPGAM1, SPLN3, PROC PHREG, %PCFrailty
- R: coxph, coxme, frailtyPenal

Zur quantitativen Betrachtung der Schätzungen wurden die mediane Abweichung, die geschätzten und empirischen Standardfehler, der mittlere quadratische Fehler, die Überdeckungsrate (aus 95% Konfidenzintervallen) und die Konvergenzrate verwendet.

Für die durchgeführte Validierung wurden Simulationsdaten generiert. Mithilfe dieser Daten und den daraus resultierenden Ergebnissen wurde eine Auswahl der Funktionen getroffen. Diese Auswahl wurde verwendet um einen realen Datensatz zu analysieren. Bei diesem Datensatz handelt es sich um Daten der Halluca-Studie, einer populationsbasierten prospektiven Studie von 1696 Lungenkrebspatienten. Durchgeführt wurde sie in den Regionen Halle und Dessau von 1996 bis 2002. Das Ziel der Studie war es die Versorgungsqualität von Patienten mit Lungenkrebs im südlichen Sachsen-Anhalt einzuschätzen. Dabei war ein interessierender Outcome die Lebensdauer der Lungenkrebspatienten.

Alle zuvor genannten Funktionen sind geeignet um gruppierte Lebensdauerdaten zu analysieren. Ebenso sind sie in der Lage zu erkennen, wenn kein Gruppeneffekt in den Daten vorliegt (Cox-Modell). Vergleicht man die Ergebnisse von parametrischen mit semiparametrischen Modellansätzen fällt auf, dass keine nennenswerten Unterschiede zwischen den Schätzungen zu erkennen sind. Obwohl die parametrischen Funktionen die zugrundeliegende Verteilung der Daten benutzen, konnten keine Verbesserungen bei den Punktschätzern und den dazugehörigen Standardfehlern im Vergleich zu semiparametrischen und quasi-semiparametrischen Funktionen verzeichnet werden. Dadurch wird die Verwendung von parametrischen Modellansätzen im Bereich der shared Frailty-Modelle erheblich in Frage gestellt. Weiterhin zeigen die Ergebnisse der Simulationen, dass verallgemeinerte gemischte Modelle mit einer Poisson Likelihood, einem spezifischen Offset und einem binären Response geeignet sind um gruppierte Ereigniszeitdaten zu analysieren. Diese Modellklasse bietet dem Nutzer weiterführende Auswertungsmöglichkeiten. Der Modellansatz der verallgemeinerten gemischten Modelle wird im SAS Makro %PCFrailty genutzt. Bei der Verwendung des Makros ist es vorteilhaft in jedem Piece eine gleichgroße Anzahl an Ereignissen zu definieren, damit der Informationsgehalt für die Schätzung über alle Pieces konstant bleibt. Bei Betrachtung der Schätzer der festen Effekte ergeben sich über alle Funktionen hinweg nur marginale Unterschiede in der Güte. Eine Ausnahme bildet die Funktion frailtyPenal, bei der die Schätzer für die dichotome Kovariable stärker vom wahren Wert abgewichen sind. Über alle Funktion ist zu erkennen, dass metrisch erfasste Kovariablen mit geringerer Variabilität geschätzt werden als dichotome Kovariaben. Dieser Effekt ist auf den geringeren Informationsgehalt der dichotomen Kovariablen zurückzuführen. Bei den Schätzungen der Frailtyvarianz bzw. der Varianz des zufälligen Effektes treten Unterschiede auf. Wird ein gammaverteilter zufälliger Effekt angenommen, haben sich die Funktionen coxph und SPGAM1 als besonders geeignet herausgestellt. Über alle Auswertungen hinweg betrachtet wiesen sie die geringste Verzerrung sowie gute Überdeckungs- und Konvergenzraten auf. Im zweiten Fall, der log-normal-verteilten Frailty, stellten sich drei Funktionen als besonders geeignet heraus. PROC PHREG, %PCFrailty und coxme zeigten eine geringe mediane Abweichung sowie gute Überdeckungs- und Konvergenzenraten. Diese fünf Funktionen wurden genutzt, um die Halluca-Daten zu analysieren. Die Auswertungen unterstreichen, dass die Schätzungen der festen Effekte robust bezüglich der Frailtyverteilung sind.

# Literatur

- Cox DR (1972) Regression Models and Life-Tables. Journal of the Royal Statistical Society. Series B (Methodological) 34:187–220.
- [2] Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE: Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models. Springer, 2005, pp. 211–212.
- [3] Wienke A: Frailty Models in Survival Analysis. Chapman & Hall/CRC, Boca Raton, 2010.
- [4] Hougaard P: Analysis of Multivariate Survival Data. Springer, New York, 2000.
- [5] Clayton DG (1978) A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika 65:141–151.
- [6] Laird N, Olivier D (1981) Covariance analysis of censored survival data using loglinear analysis techniques. Journal of the American Statistical Association 76:231– 240.
- [7] Klein JP, Moeschberger ML: Survival Analysis Techniques for Censored and Truncated Data. Springer, New York, 2003, pp. 63–70.
- [8] Lawless JF: Statistical Models and Methods for Lifetime Data. Wiley and Sons, 2003, pp. 30–33.
- Kaplan EL, Meier P (1958) Nonparametric Estimation from Incomplete Observations. Journal of the American Statistical Association 53:457 - 481.
- [10] Legrand C (2005) Assessing heterogeneity in multicenter clinical trials using the frailty model. Universiteit Hasselt; Centrum voor Statistiek, Dissertation.
- [11] Cook RJ, Lawless JF: The Statistical Analysis of Recurrent Events. Springer, New York, 2007, pp. 65–67.
- [12] Duchateau L, Janssen P: The Frailty Model. Springer, New York, 2008.
- [13] Nelsen RB: An Introduction to Copulas. Springer, New York, 2006.

- [14] Wei LJ, Lin DY, Weissfeld L (1989) Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. Journal of the American Statistical Association 84:1065–1073.
- [15] Feng S, Wolfe RA, Port FK (2005) Frailty Survival Model Analysis of the National Deceased Donor Kidney Transplant Dataset Using Poisson Variance Structures. Journal of the American Statistical Association 100:728–735.
- [16] Rockwood K (2005) Frailty and its definition: a worthy challenge. Journal of the American Geriatric Society 53:1069–1070.
- [17] Gillick M (2001) Guest Editorial: Pinning Down Frailty. Journal of Gerontology: Medical Sciences 56A:M134–M135.
- [18] Hirsch K, Wienke A, Kuss O (2011) Log-normal frailty models fitted as Poisson generalized linear mixed models: New evidence from simulations. 56. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (gmds), 6. Jahrestagung der Deutschen Gesellschaft für Epidemiologie (DGEpi). Mainz, 26.-29.09.2011.
- [19] Vaupel J, Manton K, Stallard E (1979) The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality. Demography 16:439–454.
- [20] Hutchinson TP, Lai CD: The Engineering Statistician's Guide to Continuous Bivariate Distributions. Rumsby, Adelaide, 1991.
- [21] Therneau TM, Grambsch PM: Modeling Survival Data: Extending the Cox Model. Springer, New York, 2000.
- [22] Aitkin M, Clayton D (1980) The fitting of exponential, weibull and extreme value distributions to complex censored survival data using GLIM. Applied Statistics 29:156-163.
- [23] Holford TR (1976) Life Tables with Concomitant Information. Biometrics 32:587– 597.
- [24] Feng S, Nie L, Wolfe RA (2009) Laplace's approximation for relative risk frailty models. Lifetime Data Analysis 15:343–356.
- [25] Vu HTV (2002) SAS macros for parametric and semiparametric conditional shared gamma and log-normal frailty models. Computational Statistics and Data Analysis 40:657–658.

- [26] Vu HTV, Knuiman MW (2002) A hybrid ML-EM algorithm for calculation of maximum likelihood estimates in semiparametric shared frailty models. Computational Statistics and Data Analysis 40:173–187.
- [27] Vu HTV, Knuiman MW (2002) Estimation in semiparametric marginal shared gamma frailty models. Australian and New Zeeland Journal of Statistics 44:489– 501(13).
- [28] Vaida F, Xu R (2000) Proportional hazards model with random effects. Statistics in Medicine 19:3309-3324.
- [29] Therneau TM, Grambsch PM, Pankratz VS (2003) Penalized Survival Models and Frailty. Journal of Computational and Graphical Statistics 12:156–175.
- [30] Ripatti S, Palmgren J (2000) Estimation of Multivariate Frailty Models Using Penalized Partial Likelihood. Biometrics 56:1016–1022.
- [31] Rondeau V, Gonzalez JR (2005) frailtypack: A computer program for the analysis of correlated failure time data using penalized likelihood estimation. Computer Methods and Programs in Biomedicine 80:154–164.
- [32] Rondeau V, Mazroui Y, Gonzalez JR (2012) frailtypack: An R package for the analysis of correlated survival data with frailty models using penalized likelihood estimation or parametrical estimation. Journal of Statistical Software 47:1–28.
- [33] Ramsay JO (1988) Monotone Regression Splines in Action. Statistical Science 3:425–461.
- [34] Carstensen B (2003) %Lexis: a SAS-macro for splitting follow-up time. http://BendixCarstensen.com/Lexis/Lexis.sas, aufgerufen: 08.09.2012.
- [35] Klein JP (1992) Semiparametric Estimation of Random Effects Using the Cox Model Based on the EM Algorithm. Biometrics 48:795–806.
- [36] Vu HTV, Segal MR, Knuiman MW, James IR (2001) Asymptotic and small sample statistical properties of random frailty variance estimates for shared gamma frailty models. Communications in Statistics - Simulation and Computation 30 (3):581– 595.
- [37] Rondeau V, Commenges D, Joly P (2003) Maximum penalized likelihood estimation in a gamma-frailty model. Lifetime Data Analysis 9:139–153.

- [38] Wolfinger R, O'Connell M (1993) Generalized Linear Mixed Models: A Pseudolikelihood Approach. Journal of Statistical Computation and Simulation 48:233– 243.
- [39] Burton A, Altman DG, Royston P, Holder RL (2006) The design of simulation studies in medical statistics. Statistics in Medicine 25:4279 - 4292.
- [40] Hirsch K, Wienke A (2012) Software for semiparametric shared gamma and lognormal frailty models: An overview. Computer Methods and Programs in Biomedicine 107:582–597.
- [41] Everitt BS: The Cambridge Dictionary of Statistics. Cambridge University Press, 2006, pp. 254.
- [42] Bender R, Lange S (2001) Was ist ein Konfidenzintervall? Deutsche Medizinische Wochenzeitschrift 126:T41–T41.
- [43] Burdick RK, Graybill FA: Confidence intervals on variance components. Marcel Dekker, INC, 1992, pp. 14–16.
- [44] Weiss J (2010) Lecture 8. http://www.unc.edu/courses/2010fall/ecol/563/001/docs/ lectures/lecture8.htm, aufgerufen: 09.02.2013.
- [45] Stryhn H, Christensen J Confidence intervals by profile likelihood mehtod, with applications in veterinary epidemiology. http://people.upei.ca/hstryhn/ stryhn208.pdf, aufgerufen: 09.02.2013.
- [46] Niwitpong S (2011) Coverage probability of confidence intervals for the normal mean and variance with restricted parameter space. World Academy of Science, Engineering and Technology 56:540–545.
- [47] SAS/STAT(R) 9.2 User's Guide, Second Edition, The Glimmix Procedure COV-TEST Statement. http://support.sas.com/documentation/cdl/en/statug/63033/ HTML/default/viewer.htm#statug\_glimmix\_sect001.htm, aufgerufen: 09.08.2012.
- [48] Brown H, Prescott R: Applied Mixed Models in Medicine. Wiley, Chichester, 1999, pp. 74–76.
- [49] Satterthwaite FE (1946) An Approximate Distribution of Estimates of Variance Components. Biometrics Bulletin 2:110–114.

- [50] Smithson M: Confidence intervals, Sage university papers 140. Quantitative applications in the social sciences no. 07/140; Sage Publications, Inc., 2003, pp. 10–18.
- [51] Kuss O, Blankenburg T, Haerting J (2008) A relative survival model for clustered responses. Biometrical Journal 50:408 – 418.
- [52] Bollmann A, Blankenburg T, Haerting J, Kuss O, Schütte W, Dunst J, Neef H (2004) Survival of Patients in Clinical Stages IIIIb of Non-Small-Cell Lung Cancer Treated with Radiation Therapy Alone: Results of a Population-Based Study in Southern Saxony-Anhalt. Strahlentherapie und Onkologie 180:488–496(9)
- [53] Liu L, Yu Z (2008) A likelihood reformulation method in non-normal random effects models. Statistics in Medicine 27:3105–3124.

# 8 Thesen

- 1. Shared Frailty-Modelle sind eine Möglichkeit um gruppierte Ereigniszeiten zu analysieren. Dabei werden insbesondere das shared Gamma Frailty-Modell und das shared Log-normale Frailty-Modell häufig verwendet.
- 2. Shared Frailty-Modelle sind komplexe Modelle welche nicht standardmäßig in etablierten Statistikpaketen enthalten sind. Damit ist es erforderlich auf Alternativen wie R-Funktionen oder SAS Macros zurückzugreifen. Um zu ermitteln welche Funktionen die besten Schätzungen bezüglich fester und zufälliger Effekte liefern, können Simulationsstudien genutzt werden.
- 3. Die Funktionen SPGAM1, coxph und frailtyPenal liefern für das shared Gamma Frailty-Modell adäquate Ergebnisse. Während SPLN3, PROC PHREG, %PCFrailty, coxph und coxme für das shared Log-normale Frailty-Modell vorzuziehen sind.
- 4. Die Funktion frailtyPenal zeigt Schwächen in der Schätzung von dichotomen festen Effekten bei der Analyse von Ereigniszeiten. Ansonsten sind die Unterschiede zwischen den betrachteten Funktionen marginal.
- 5. Parametrische Modellansätze liefern trotz verwendeter wahrer Parameter keine höhere Schätzgüte im Vergleich zu semiparametrischen Modellansätzen.
- 6. Verallgemeinerte gemischte Modelle mit einem binären Response, einer Poisson Likelihood und einem spezifischen Offset liefern adäquate Ergebnisse bei der Analyse von gruppierten Überlebenszeiten.
- 7. Bei der Anwendung des Makros %PCFrailty wird empfohlen, eine konstante Anzahl von Ereignissen über alle Pieces zu verwenden. Dadurch wird für die Schätzung der Basishazardfunktion in den Pieces der Informationsgehalt stabil.
- 8. Unter der Voraussetzung, dass ein Clustereffekt vorliegt (shared Frailty-Modell), ist mit steigender Clusteranzahl die Variabilität der Punktschätzer der Varianz der Frailty bzw. des zufälligen Effektes geringer.

- 9. Unter der Voraussetzung, dass kein Clustereffekt vorliegt (Cox-Modell), ist mit steigender Clusteranzahl die Variabilität der Punktschätzer der Varianz der Frailty bzw. des zufälligen Effektes größer.
- Bei der Anwendung der Funktionen auf die Halluca-Studie (Halle Lung Cancer Study) zeigt sich, dass die Schätzung der festen Effekte robust gegenüber der Frailtyverteilung ist.

# Lebenslauf

#### Persönliche Daten

Name:	Katharina Hirsch
Geburtsdatum:	2. November 1985, in Görlitz
Familienstand:	ledig
Staatsangehörigkeit:	deutsch

# Schulbildung

1992 - 1996	$Melanchthon\mathcharged Görlitz$
1996 - 1998	Melanchthon-Mittelschule Görlitz
1998 - 2004	Joliot-Curie-Gymnasium Görlitz

# Studium und Beruf

2004 - 2008	Diplomstudium der Biomathematik an der Fachhochschule
	Zittau/Görlitz
	Abschluss als Diplom Biomathematikerin
2008 - heute	Wissenschaftliche Mitarbeiterin,
	Martin-Luther-Universität Halle-Wittenberg,
	Institut für Medizinische Epidemiologie,
	Biometrie und Informatik

# ${\it Selbstst} \ddot{a}n digkeitserk l\ddot{a}rung$

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Quellen und Hilfsmittel angefertigt habe.

Ort, Datum

Katharina Hirsch

# Erklärung über frühere Promotionsversuche

Hiermit erkläre ich, dass ich keine früheren Promotionsversuche unternommen habe.

Ort, Datum

Katharina Hirsch