



**Medizinische Fakultät**  
Institut für Epidemiologie und  
Medizinische Biometrie

Leitung: Prof. Dr. D. Rothenbacher

Schwabstraße 13  
89075 Ulm, Germany

Prof. Dr. Rainer Mücke  
Tel: +49 731 50-26903  
Fax: +49 731 50-26902  
rainer.muecke@uni-ulm.de  
<http://www.uni-ulm.de/biometrie>

# Consulting Class

## Kolonkarzinom

**Datensatzbezeichnung:** Colon

**Fragestellung:**

1. Gibt es bei der Überlebenszeitanalyse einen Zusammenhang zwischen der Behandlungsart des Tumors und der Überlebenszeit?
2. Welche Faktoren beeinflussen darüber hinaus das Sterben an Darmkrebs?

**Zielgröße:** Überlebenszeit  
Tod durch Kolonkarzinom

**Auswertungshinweis:**  
Führen Sie eine Überlebenszeitanalyse durch!

# Inhaltsverzeichnis

A Datensatzbeschreibung .....	1
1 Medizinischer Hintergrund .....	1
2 Variablenbeschreibung und Kodierung.....	1
3 Sonstiges .....	2
B Auswertung .....	3
1 Univariate Deskription .....	3
1.1 Status .....	3
1.2 Therapie .....	3
1.3 Stetige Variablen .....	3
1.4 Diskrete Variablen .....	4
1.5 Deskriptive Auswertung, Stratifiziert nach Behandlungsart.....	6
2 Kaplan-Meier-Schätzungen .....	9
2.1 Gesamtüberleben .....	10
2.2 Überlebenszeiten in Abhängigkeit von Behandlungsart .....	13
2.3 Überlebenszeiten in Abhängigkeit von Geschlecht .....	14
2.4 Überlebenszeiten in Abhängigkeit von Altersklassen.....	15
2.5 Überlebenszeiten in Abhängigkeit von der Differenzierbarkeit der Tumorzellen .....	16
2.6 Überlebenszeiten in Abhängigkeit von der Ausdehnung des Tumors.....	17
2.7 Überlebenszeiten in Abhängigkeit von der Dauer der Behandlung .....	18
2.8 Überlebenszeiten in Abhängigkeit von Lymphknotenklassen.....	19
2.9 Überlebenszeiten in Abhängigkeit vom Tumorbefall anliegender Organe .....	20
2.10 Überlebenszeiten in Abhängigkeit von dem Darmdurchbruch.....	21
2.11 Überlebenszeiten in Abhängigkeit von der Darmverstopfung.....	22
3 Cox-Regression .....	23
3.1 Prüfung der Modellannahmen .....	24
3.2 Ergebnis des Cox-Modells .....	29
3.2.1 Volles Modell .....	30
3.2.2 Modell mit Backward-Selektion (slstay 10%) .....	34
3.2.3 Endmodell ohne Variable aus 3.2.2.....	39
4 Diskussion .....	40
SAS-Code .....	42

# A Datensatzbeschreibung

## 1 Medizinischer Hintergrund

Die Daten stammen aus Untersuchungen, die bei Patienten mit einem Kolonkarzinom im Rahmen einer Studie durchgeführt wurden. Es handelt sich bei diesen Studiendaten um eine nicht bekannten Datensatz aus dem R-Paket. Die Operation umfasste das vollständige Entfernen des Tumors zusammen mit 2,5 cm umliegendem Gewebe. Zu den Untersuchungen gehörte die Feststellung der Ausdehnung des Tumors im Darm, ob der Tumor eine Verstopfung des Darms hervorgerufen hat, ob andere Organe vom Tumor befallen wurden oder ob Lymphknoten von Metastasen befallen wurden. Es wurde davon ausgegangen, dass dies wichtige prognostische Parameter dafür sind, dass Patienten mit einem großen Tumor und / oder einem metastasierendem Tumor ein erhöhtes Risiko haben an Kolonkarzinom zu sterben.

Das Kolonkarzinom oder Dickdarmkrebs genannt, ist ein hochgradig bösartiger Tumor der Darmschleimhautzellen. Er neigt dazu, früh Metastasen über Lymph- und Blutbahnen zu streuen und ist die am häufigsten tödlich verlaufende Krebserkrankung mit weltweit stark steigender Anzahl an Neuerkrankungen.

Metastasen im engeren Sinne bezeichnen Absiedlungen eines Tumors in entferntem Gewebe.

Die tatsächlichen Heilungschancen hängen von der Art und der Lokalisation des Tumors ab.

Metastasen entstehen, indem sich Krebszellen vom ursprünglichen Tumor ablösen, mit dem Blut oder mit der Lymphe wandern und sich in anderen Körperteilen wieder ansiedeln und vermehren. Je nach dem Ausbreitungsweg heißen sie hämatogene (Blut) oder lymphogene (Lymphe) Metastasen.

Es wurden folgende Behandlungen miteinander verglichen. Eine Patientengruppe bekam nur das Medikament Levamisol. Die andere Gruppe bekam das Medikament Levamisol und eine Chemotherapie mit 5-FU. Die letzte Gruppe wurde nur beobachtet.

## 2 Variablenbeschreibung und Kodierung

Variable	Beschreibung / Kodierung
id	Patientenidentifikationsnummer
study	Gilt für Patienten die an der Studie teilnehmen: 1 = für alle teilnehmenden Patienten
rx	Behandlungsart: Obs = Patient wird nur beobachtet Lev = Patient bekommt Levamisol Lev+5-Fu = Patient bekommt Levamisol und Chemotherapie
sex	Geschlecht: 0 = weiblich 1 = männlich
age [y]	Das Alter des Patienten in dem der Tumor festgestellt wurde
obstruct	Verstopfung des Darmes durch den Tumor: 0 = keine Verstopfung des Darmes 1 = Verstopfung des Darmes
perfor	Durchbruch des Darmes durch den Tumor: 0 = keinen Darmdurchbruch 1 = Darmdurchbruch
adhere	Befall der anliegenden Organe durch den Tumor: 0 = keinen Befall anliegender Organe

	1 = Befall der anliegenden Organe
nodes	Anzahl der Befallenen Lymphknoten: . = als Missing gesetzt
time [d]	Anzahl der Tage bis zum Eintritt eines Ereignisses oder der Zensierung
status	Hat eine Zensierung der Daten stattgefunden: 0 = es fand eine Zensierung der Daten statt 1 = keine Zensierung
differ	Differenzierbarkeit der Tumorzellen: . = als Missing gesetzt 1 = gute Differenzierbarkeit 2 = mäßige Differenzierbarkeit 3 = geringe Differenzierbarkeit
extent	Lokale Ausdehnung des Tumors: 1 = Darmschleimhaut 2 = Darmmuskulatur 3 = Äußere Darmwand 4 = Darmdurchbruch
surg	Zeit von der Behandlung bis zum Eintritt des Ereignisses (auftreten eines Rezidivs oder Tod des Patienten): 0 = kurzer Zeitraum 1 = langer Zeitraum
Node4	Mehr als 4 Lymphknoten sind befallen: 0 = weniger als 4 Lymphknoten befallen 1 = mehr als 4 Lymphknoten befallen

### 3 Sonstiges

Die Tabelle der Colon-Daten hat 930 Zeilen und 16 Spalten, wobei in der ersten Zeile die Labels der Variablen stehen. Die Fallzahl gesamt der Colon-Daten beträgt n=929 und der Pfad lautet:

[T:\Projekte\Consulting\\_Class\Datensaetze\colon\daten\\_colon.xls](T:\Projekte\Consulting_Class\Datensaetze\colon\daten_colon.xls)

Die Daten stammen aus der R-Library „Boot“ und der Datensatz heißt dort colon.

Hintergrundliteratur:

- HOSMER, David W; LEMESHOW, Stanley: Applied Survival Analysis: Regression Modeling of Time to Event Data. New York; John Wiley & Sons; 1. Auflage 1999; ISBN 978-0-471-15410-5
- KLEINBAUM, David G: Survival Analysis: A Self-Learning Text. New York; Springer-Verlag; 1. Auflage 1996; ISBN 978-0-387-94543-1

## B Auswertung

Bei der Auswertung soll untersucht werden, welche Einflüsse das Sterben und die Dauer bis dahin (Überlebenszeit) an einem Darmkrebs beeinflussen. Hierfür werden zunächst die Daten deskriptiv ausgewertet um einen Überblick der Colon-Daten zu erhalten. Anschließend führt man eine Überlebenszeitanalyse mit Kaplan-Meier durch. Mit Hilfe der Cox-Regression können die Daten multivariat ausgewertet werden um den Einfluss der anderen Faktoren berücksichtigen zu können. Hier muss geachtet werden, dass die Modellvoraussetzungen erfüllt sind. Die Fallzahl gesamt der Colon-Daten beträgt n=929.

### 1 Univariate Deskription

#### 1.1 Status

Status am Ende der Studie			
	status	Frequency	Percent
Zensierung der Daten		477	51.35
Tod		452	48.65

Bei dieser Tabelle sieht man deutlich, dass 51.35% zensiert wurden und es 48.65% Todesfälle gab. Die Überlebenszeit wird nicht direkt deskriptiv ausgewertet, da es Zensierungen gibt. Daher macht es keinen Sinn einen Mittelwert auszurechnen. Daher ist es besser die Kaplan-Meier Analyse in 2.1 anzuschauen.

#### 1.2 Therapie

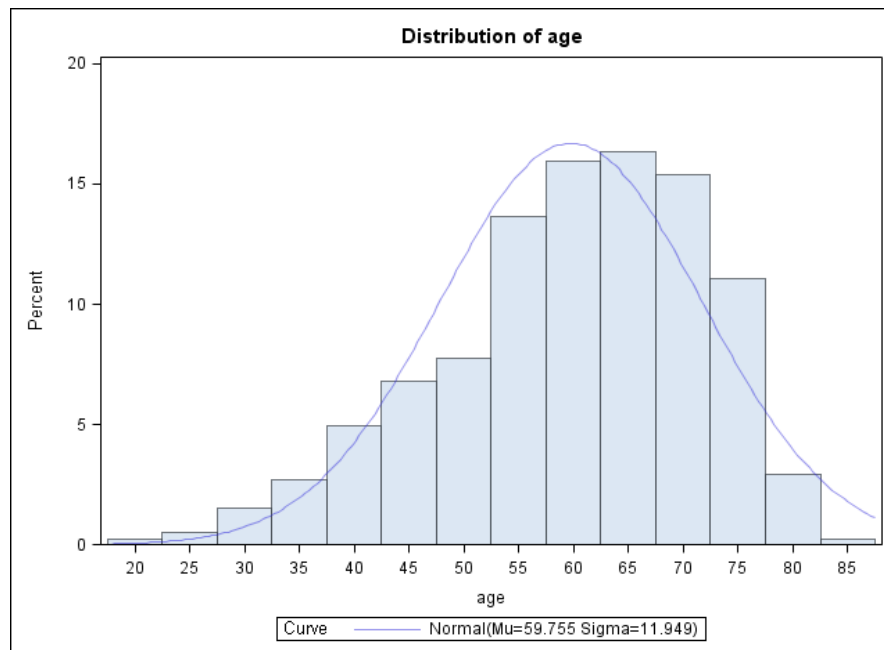
rx	Behandlungsart	
	Frequency	Percent
Lev	310	33.37
Lev+5FU	304	32.72
Obs	315	33.91

Es befindet sich in jeder Gruppe in etwa die gleiche Fallzahl an Patienten. Daher kann man davon ausgehen, dass es sich bei dieser Studie um eine randomisierte, geschichtete und balancierte Studie handelt.

#### 1.3 Stetige Variablen

Variable	Label	N		Mean	Std Dev	Median	Lower Quartile
		N	Miss				
age	Alter in Jahren	929	0	59.75	11.95	61.00	53.00

Variable	Label	Upper Quartile	Minimum	Maximum
age	Alter in Jahren	69.00	18.00	85.00



**Abbildung 1** Histogramm Alter

Das Alter der Patienten der Colon Daten ist fast normal verteilt und leicht rechtsgipflig.

## 1.4 Diskrete Variablen

Geschlecht		
sex	Frequency	Percent
w	445	47.90
m	484	52.10

Alter		
alter	Frequency	Percent
< 65	595	64.05
> 65	334	35.95

Das Alter ist in 2 Klassen eingeteilt, da die meisten Fälle in diesem Bereich liegen und Dickdarmkrebs meistens im hohen Alter auftritt.

### Differenzierung der Tumorzellen

differ	Frequency	Percent
gute Differenzierbarkeit	93	10.26
mäßige Differenzierbarkeit	663	73.18
geringe Differenzierbarkeit	150	16.56

Frequency missing = 23

Es gibt 23 Missings wo es keinen Eintrag gibt. Dies ist später für die weitere Statistik wichtig (Cox-Regression) wegen der benutzten Fallzahl.

#### Tumorausdehnung

extent	Frequency	Percent
Darmschleimhaut	21	2.26
Darmmuskulatur	106	11.41
Äußere Darmwand	759	81.70
Darmdurchbruch	43	4.63

#### Behandlungszeitraum

surg	Frequency	Percent
kurzer Behandlungszeitraum	682	73.41
langer Behandlungszeitraum	247	26.59

#### Mehr als 4 Lymphknoten befallen

node4	Frequency	Percent
weniger als 4 Lymphknoten befallen	647	72.55
mehr als 4 Lymphknoten befallen	255	27.45

#### Tumorbefall der anliegenden Organe

adhere	Frequency	Percent
kein Tumorbefall der anliegenden Organe	794	85.47
Tumorbefall der anliegenden Organe	135	14.53

#### Darmdurchbruch durch den Tumor

perfor	Frequency	Percent
keinen Darmdurchbruch durch Tumor	902	97.09
Darmdurchbruch durch Tumor	27	2.91

#### Darmverstopfung durch den Tumor

Obstruct	Frequency	Percent
keine Darmverstopfung durch den Tumor	749	80.62
Darmverstopfung durch den Tumor	180	19.38

## 1.5 Deskriptive Auswertung, stratifiziert nach Behandlungsart

**Tabelle 1** Häufigkeitstabelle Status und Behandlungsart

Frequency Percent Row Pct Col Pct	Lev	Lev+5FU	Obs	Total
Zensierte Daten	149 16.04 31.24 48.06	181 19.48 37.95 59.54	147 15.82 30.82 46.67	477 51.35
Tod	161 17.33 35.62 51.94	123 13.24 27.21 40.46	168 18.08 37.17 53.33	452 48.65
Total	310 33.37	304 32.72	315 33.91	929 100

Bei der Behandlung mit Lev+5FU gibt es mehr Zensierungen in den Daten gegenüber den anderen Behandlungsarten. Weniger Todesfälle 27% gegenüber 35% und 37%.

Statistics for Table of status by rx

Statistic	DF	Value	Prob
Chi-Square	2	12.2664	<.0022
Likelihood Ratio Chi-Square	2	12.3267	<.0021
Continuity Adj. Chi-Square	1	10.7528	<.0001
Mantel-Haenszel Chi-Square	1	0.1335	<.7149
Phi Coefficient		0.1149	
Contingency Coefficient		0.1142	
Cramer's V		0.1149	

Sample Size = 929

Es gibt einen Zusammenhang zwischen der Behandlungsart des Tumors und des Status des Patienten, da der Test signifikant ausfällt ( $p=0.0022$ ).

**Tabelle 2** Häufigkeitstabelle Geschlecht und Behandlungsart

Frequency Percent Row Pct Col Pct	Lev	Lev+5FU	Obs	Total
w	133 14.32 29.89 42.90	163 17.55 36.63 53.62	149 16.04 33.48 47.30	445 47.90
m	177 19.05 36.57 57.10	141 15.18 29.13 46.38	166 17.87 34.30 52.70	484 52.10
Total	310 33.37	304 32.72	315 33.91	929 100

Es haben mehr Frauen die Behandlung Lev+5FU bekommen als Männer.



**Tabelle 3** Häufigkeitstabelle Alter und Behandlungsart

Frequency Percent Row Pct Col Pct	Lev	Lev+5FU	Obs	Total
< 65	200 21.53 33.61 64.52	190 20.45 31.93 62.50	205 22.07 34.45 65.08	595 64.05
> 65	110 11.84 32.93 35.48	114 12.27 34.13 37.50	110 11.84 32.93 34.92	334 35.95
Total	310 33.37	304 32.72	315 33.91	929 100

Es gibt keine bevorzugte Behandlungsart ob nun der Patient über 65 oder unter 65 Jahre alt ist.

**Tabelle 4** Häufigkeitstabelle Differenzierbarkeit der Tumorzellen und der Behandlungsart

Frequency Percent Row Pct Col Pct	Lev	Lev+5FU	Obs	Total
Gute Differenzierbarkeit	37 4.08 39.78 12.33	29 3.20 31.18 9.73	27 2.98 29.03 8.77	93 10.26
Mäßige Differenzierbarkeit	219 24.17 33.03 73.00	215 23.73 32.43 72.15	229 25.28 34.54 74.35	663 73.18
Geringe Differenzierbarkeit	44 4.86 29.33 14.67	54 5.96 36.00 18.12	52 5.74 34.54 74.35	150 16.56
Total	300 33.11	298 32.89	308 34.00	906 100

Es gibt keine bevorzugte Behandlungsart ob nun die Tumorzellen Differenzierbar sind oder nicht.

**Tabelle 5** Häufigkeitstabelle Lokale Ausdehnung des Tumors und der Behandlungsart

Frequency Percent Row Pct Col Pct	Lev	Lev+5FU	Obs	Total
Darmschleimhaut	3 0.32 14.29 0.97	10 1.08 47.62 3.29	8 0.86 38.10 2.54	21 2.26
Darmmuskulatur	36 3.88 33.96 11.61	32 3.44 30.19 10.53	38 4.09 35.85 12.06	106 11.41
Äußere Darmwand	259 27.88 34.12 83.55	251 27.02 33.07 82.57	249 26.80 32.81 79.05	759 81.70
Darmdurchbruch	12 1.29 27.91 3.87	11 1.18 25.58 3.62	20 2.15 46.51 6.35	43 4.63
Total	310 33.37	304 32.72	315 33.91	929 100

Es gibt keine bevorzugte Behandlungsart ob nun der Tumor sich weit im Darm ausgebreitet hat. Bei der Gruppe die nur beobachtet wurde findet man häufiger einen Darmdurchbruch durch den Tumor vor. Bei der Gruppe mit der Behandlung Levamisol und 5FU hat sich der Tumor nur bis zur Darmschleimhaut ausgedehnt.

**Tabelle 6** Häufigkeitstabelle Behandlungszeitraum des Tumors und der Behandlungsart

Frequency Percent Row Pct Col Pct	Lev	Lev+5FU	Obs	Total
Kurzer Behandlungszeitraum	230 24.76 33.72 74.19	228 24.54 33.43 75.00	224 24.11 32.84 71.11	93 10.26
Langer Behandlungszeitraum	80 8.61 32.39 25.81	76 8.18 30.77 25.00	91 9.80 36.84 28.89	663 73.18
Total	310 33.37	304 32.72	315 33.91	929 100

Der Behandlungszeitraum spielt bei der Behandlungsart keine große Rolle.

**Tabelle 7** Häufigkeitstabelle Tumorbefall der Lymphknoten und der Behandlungsart

Frequency Percent Row Pct Col Pct	Lev	Lev+5FU	Obs	Total
Weniger als 4 Lymphkno- ten befallen	221 23.79 32.79 71.29	225 24.22 33.38 74.01	228 24.54 33.83 72.38	674 72.55
Mehr als 4 Lymphknoten befallen	89 9.58 34.90 28.71	79 8.50 30.98 25.99	87 9.36 34.12 27.62	255 27.45
Total	310 33.37	304 32.72	315 33.91	929 100

Es gibt keine bevorzugte Behandlungsart ob nun mehr als 4 Lymphknoten befallen sind oder weniger.

Frequency Missing = 23

**Tabelle 8** Häufigkeitstabelle Tumorbefall der anliegenden Organe und der Behandlungsart

Frequency Percent Row Pct Col Pct	Lev	Lev+5FU	Obs	Total
Kein Tumorbefall der anliegenden Organe	261 28.09 32.87 84.19	265 28.53 33.38 87.17	268 28.85 33.75 85.08	794 85.47
Tumorbefall der anliegenden Organe	49 5.27 36.30 15.81	39 4.20 28.89 12.83	47 5.06 34.81 14.92	135 14.53
Total	310 33.37	304 32.72	315 33.91	929 100

Es gibt keine bevorzugte Behandlungsart ob nun ein Befall der anliegenden Organe mit dem Tumor vorhanden ist oder nicht.

**Tabelle 9** Häufigkeitstabelle Darmdurchbruch und Behandlungsart

Frequency Percent Row Pct Col Pct	Lev	Lev+5FU	Obs	Total
Keinen Darmdurchbruch durch den Tumor	300 32.29 33.26 96.77	296 31.86 32.82 97.37	306 32.94 33.92 97.14	902 97.09
Darmdurchbruch durch den Tumor	10 1.08 37.04 3.23	8 0.86 29.63 2.63	9 0.97 33.33 2.86	27 2.91
Total	310 33.37	304 32.72	315 33.91	929 100

Es gibt keine bevorzugte Behandlungsart ob nun ein Darmdurchbruch vorhanden ist oder nicht.

**Tabelle 10** Häufigkeitstabelle Verstopfung und Behandlungsart

Frequency Percent Row Pct Col Pct	Lev	Lev+5FU	Obs	Total
Keine Darmverstopfung	247 26.59 32.98 79.68	250 26.91 33.38 82.24	252 27.13 33.64 80.00	749 80.62
Darmverstopfung	63 6.78 35.00 20.32	54 5.81 30.00 17.76	63 6.78 35.00 20.00	180 19.38
Total	310 33.37	304 32.72	315 33.91	929 100

Es gibt keine bevorzugte Behandlung bei einer Darmverstopfung.

## 2 Kaplan-Meier-Schätzungen

Der Kaplan-Meier-Schätzer (auch Produkt-Limit-Schätzer) dient zum Schätzen der Wahrscheinlichkeit, dass bei einem Versuchsobjekt ein bestimmtes Ereignis innerhalb eines Zeitintervalls nicht eintritt. Er wird häufig für das Schätzen der Überlebenswahrscheinlichkeit genutzt (bestimmtes Ereignis: Tod). Bei der Kaplan-Meier-Methode gibt es sogenannte Zensierungen, d.h. Patienten die entweder die Studie vorher verlassen, an anderer Ursache gestorben sind oder die Studie bis zum Ende überlebt haben.

Einfaches Kaplan-Meier Beispiel in SAS:

```
PROC LIFETEST CONFBAND=EP ALPHAQT=0.05 data=colon.colon PLOT=(S);
    TIME time * status(0);
RUN;
```

Durch die Option PLOT=(S) wird eine Grafik mit der Überlebensfunktion gezeichnet. Im Time-Statement stehen die Variable ‚Überlebenszeit‘ (hier time) und die Zensierungsvariable (hier status) mit den Zensierungszeichen (0), d. h. Patienten, die noch leben oder an anderer Ursache gestorben sind, werden zensiert. Im STRATA-Statement kann angegeben werden, dass nach einer bestimmten Variable stratifiziert werden soll, d.h. für jeden Zustand der Variable wird eine Überlebenskurve erstellt. In diesem Beispiel:

```
PROC LIFETEST CONFBAND=EP ALPHAQT=0.05 data=colon.colon PLOT=(S (CL),LLS);
    TIME time * status(0);
    STRATA rx;
```

**RUN;**

wird nach der Variable Therapie stratifiziert, die angibt ob eine Therapieart eine bessere Überlebenszeitwahrscheinlichkeit hat als eine andere.

Output besteht aus folgenden Abschnitten:

- **Product Limit Survival Estimates:** Hier wird nach dem klassischen Verfahren die Überlebenszeiten ausgerechnet. Es hat den gleichen Aufbau wie wenn man das Kaplan-Meier-Verfahren von Hand rechnen würde.
- **Quartile Estimates:** Hier werden die einzelnen Quartile (75%, 50%, 25%) angezeigt. Fehlende Angaben entstehen dadurch, dass die Überlebenswahrscheinlichkeit nicht unter diesen Wert fällt.
- **Summary of the Number of Censored and Uncensored Values:** Hier wird die Anzahl der Zensierten und Unzensierten Werte angezeigt, sowie im prozentualen Verhältnis dargestellt.
- **Testing Homogeneity of Survival Curves for time over Strata:** Hier werden die wichtigsten Tests gerechnet (speziell der Log-Rank-Test) und die Ergebnisse dargestellt.

## 2.1 Gesamtüberleben

Product-Limit Survival Estimates

time	Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.00	1.0000	0	0	0	929
23.00	0.9989	0.00108	0.00108	1	928
24.00	0.9978	0.00215	0.00152	2	927
34.00	0.9968	0.00323	0.00186	3	926
45.00	0.9957	0.00431	0.00215	4	925
52.00	0.9946	0.00538	0.00240	5	924
56.00	0.9935	0.00646	0.00263	6	923
79.00	0.9925	0.00753	0.00284	7	922
93.00	0.9914	0.00861	0.00303	8	921
113.00	0.9903	0.00969	0.00321	9	920
122.00	0.9892	0.0108	0.00339	10	919
125.00	0.9882	0.0118	0.00355	11	918
127.00	0.9871	0.0129	0.00370	12	917
129.00	0.9860	0.0140	0.00385	13	916
133.00	0.9849	0.0151	0.00400	14	915
138.00	0.9839	0.0161	0.00414	15	914
141.00	0.9828	0.0172	0.00427	16	913
144.00	0.9817	0.0183	0.00440	17	912
145.00	0.9806	0.0194	0.00452	18	911
150.00	0.9795	0.0205	0.00464	19	910
164.00	0.9785	0.0215	0.00476	20	909
165.00	0.9774	0.0226	0.00488	21	908
166.00	0.9763	0.0237	0.00499	22	907
171.00	.	.	.	23	906
171.00	0.9742	0.0258	0.00520	24	905
186.00	0.9731	0.0269	0.00531	25	904
.					
.					
.					
.					
709.00	0.7835	0.2165	0.0135	201	727
712.00	0.7824	0.2176	0.0135	202	726

716.00	0.7814	0.2186	0.0136	203	725
717.00	0.7803	0.2197	0.0136	204	724
718.00	0.7792	0.2208	0.0136	205	723
720.00	0.7781	0.2219	0.0136	206	722
721.00	0.7771	0.2229	0.0137	207	721
723.00	0.7760	0.2240	0.0137	208	720
729.00	0.7749	0.2251	0.0137	209	719
730.00	0.7738	0.2262	0.0137	210	718
736.00	0.7727	0.2273	0.0138	211	717
739.00	0.7717	0.2283	0.0138	212	716
743.00	.	.	.	213	715
743.00	0.7695	0.2305	0.0138	214	714
753.00	0.7684	0.2316	0.0138	215	713
755.00	0.7674	0.2326	0.0139	216	712
758.00	0.7663	0.2337	0.0139	217	711
.					
.					
.					
.					
.					
1798.00	0.5666	0.4334	0.0163	402	521
1800.00*	.	.	.	402	520
1803.00*	.	.	.	402	519
1812.00	0.5655	0.4345	0.0163	403	518
1814.00*	.	.	.	403	517
1814.00*	.	.	.	403	516
1818.00	0.5644	0.4356	0.0163	404	515
1819.00*	.	.	.	404	514
1820.00*	.	.	.	404	513
1823.00*	.	.	.	404	512
1823.00*	.	.	.	404	511
1827.00*	.	.	.	404	510
1827.00*	.	.	.	404	509
1828.00*	.	.	.	404	508
1829.00	0.5633	0.4367	0.0163	405	507
1831.00	0.5622	0.4378	0.0163	406	506
.					
.					
.					
.					
3017.00*	.	.	.	452	15
3019.00*	.	.	.	452	14
3024.00*	.	.	.	452	13
3030.00*	.	.	.	452	12
3078.00*	.	.	.	452	11
3085.00*	.	.	.	452	10
3087.00*	.	.	.	452	9
3173.00*	.	.	.	452	8
3185.00*	.	.	.	452	7
3192.00*	.	.	.	452	6
3214.00*	.	.	.	452	5
3238.00*	.	.	.	452	4
3308.00*	.	.	.	452	3
3309.00*	.	.	.	452	2
3325.00*	.	.	.	452	1
3329.00*	0.4551	.	.	452	0

2 Jahres (730  
Tage)  
Überlebenswsk  
=77.38%

5 Jahres (1825  
Tage)  
Überlebenswsk  
=56.44%

NOTE: The marked survival times are censored observations.

# Summary Statistics for Time Variable time

## Quartile Estimates

Percent	Point Estimate	95% Confidence Interval Transform	[Lower	Upper)
75	.	LOGLOG	.	.
50	2552.00	LOGLOG	2152.00	.
25	806.00	LOGLOG	721.00	905.00

Mediane  
Überlebenszeit=  
2552 Tage = 6.9  
Jahre

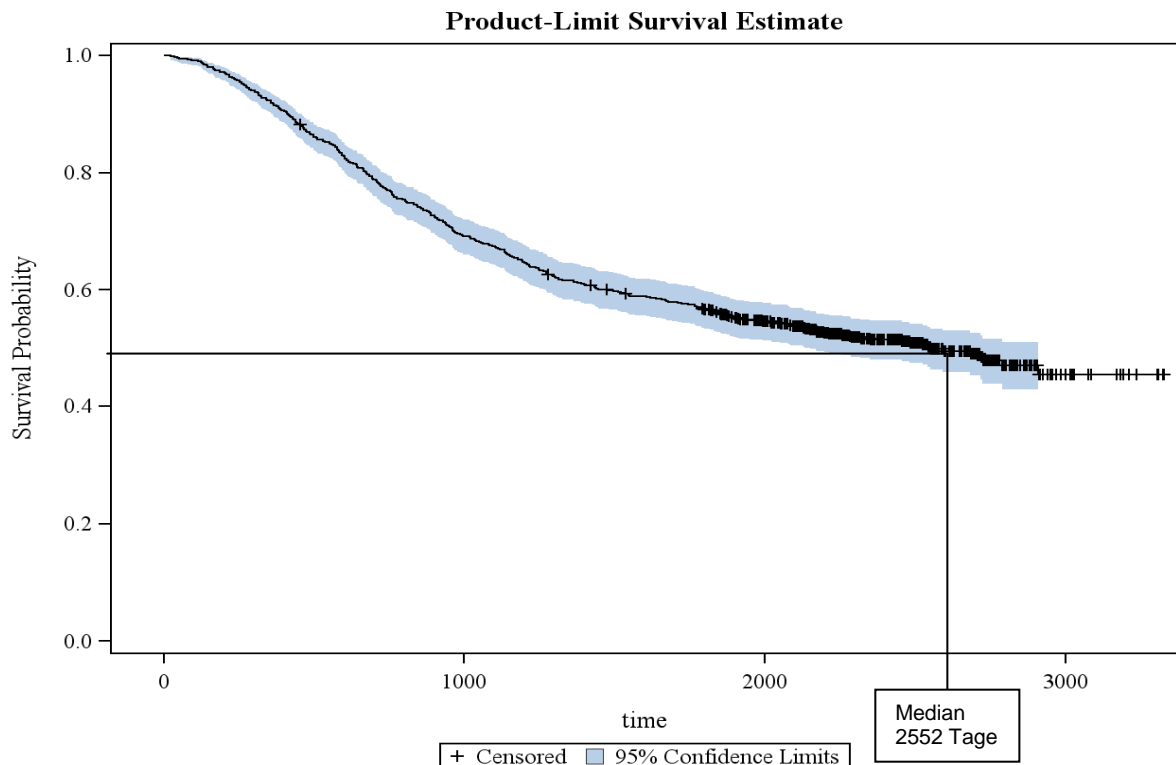
Mean	Standard Error
1928.67	35.27

NOTE: The mean survival time and its standard error were underestimated because the largest observation was censored and the estimation was restricted to the largest event time.

Bemerkung: Mittelwert ist hier kein geeignetes Maß aufgrund der Zensierungen.

## Summary of the Number of Censored and Uncensored Values

Total	Failed	Censored	Percent Censored
929	452	477	51.35



**Abbildung 2** Kaplan-Meier, Gesamtüberleben mit 95%-Konfidenzintervall

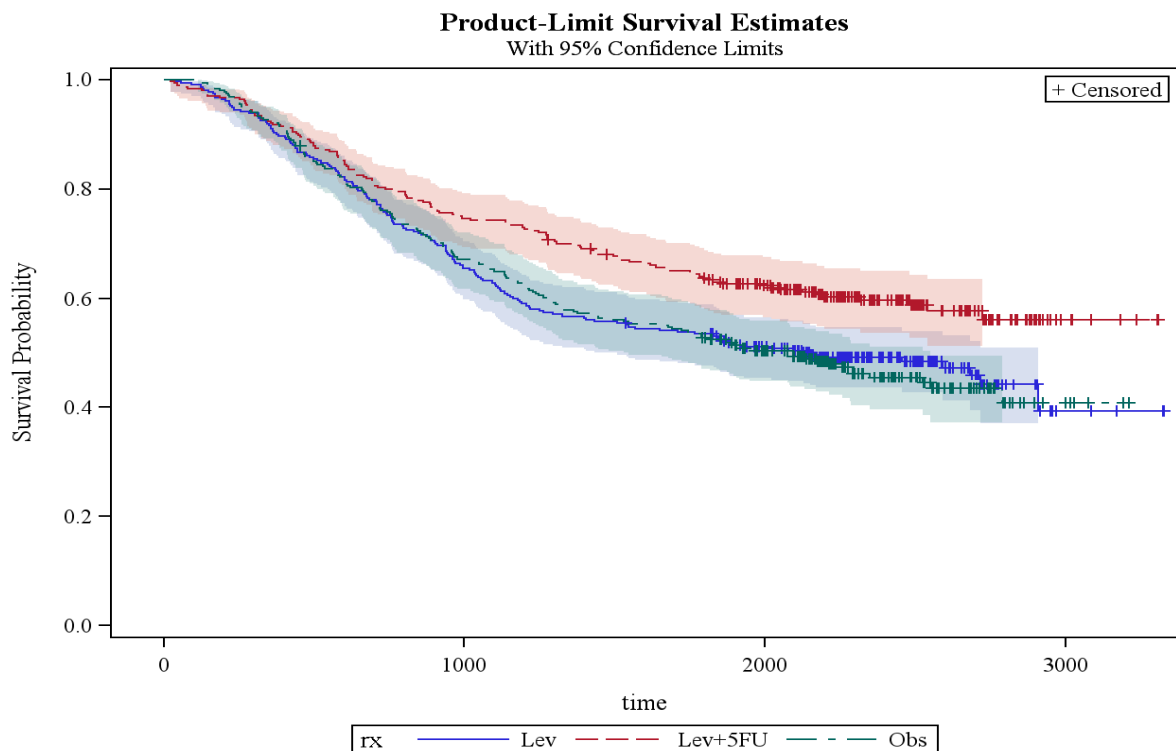
5 Jahres Überlebenswahrscheinlichkeit:

Gesamtüberleben: 0.5644

## 2.2 Überlebenszeiten in Abhängigkeit von Behandlungsart

**Tabelle 11** Zusammenfassung der Zensierten und Unzensierten Werte von Behandlungsart

Stratum	rx	Total	Failed	Censored	Percent Censored
1	Levamisol	310	161	149	48.06
2	Levamisol+5FU	304	123	181	59.54
3	Observation	315	168	147	46.67
Total		929	452	477	51.35



**Abbildung 3** Kaplan-Meier Behandlungsart mit 95%-Konfidenzintervall

Test of Equality over Strata

Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	11.6831	2	0.0029
Wilcoxon	31.3799	1	<.0001
-2Log(LR)	28.5207	1	<.0001

Log-Rank-Test:  
0.0029 < 0.05  
→ signifikant

An Hand des Wertes des Log-Ranks ( $p < 0.0029$ ) kann man einen signifikanten Einfluss der Behandlungsart erkennen.

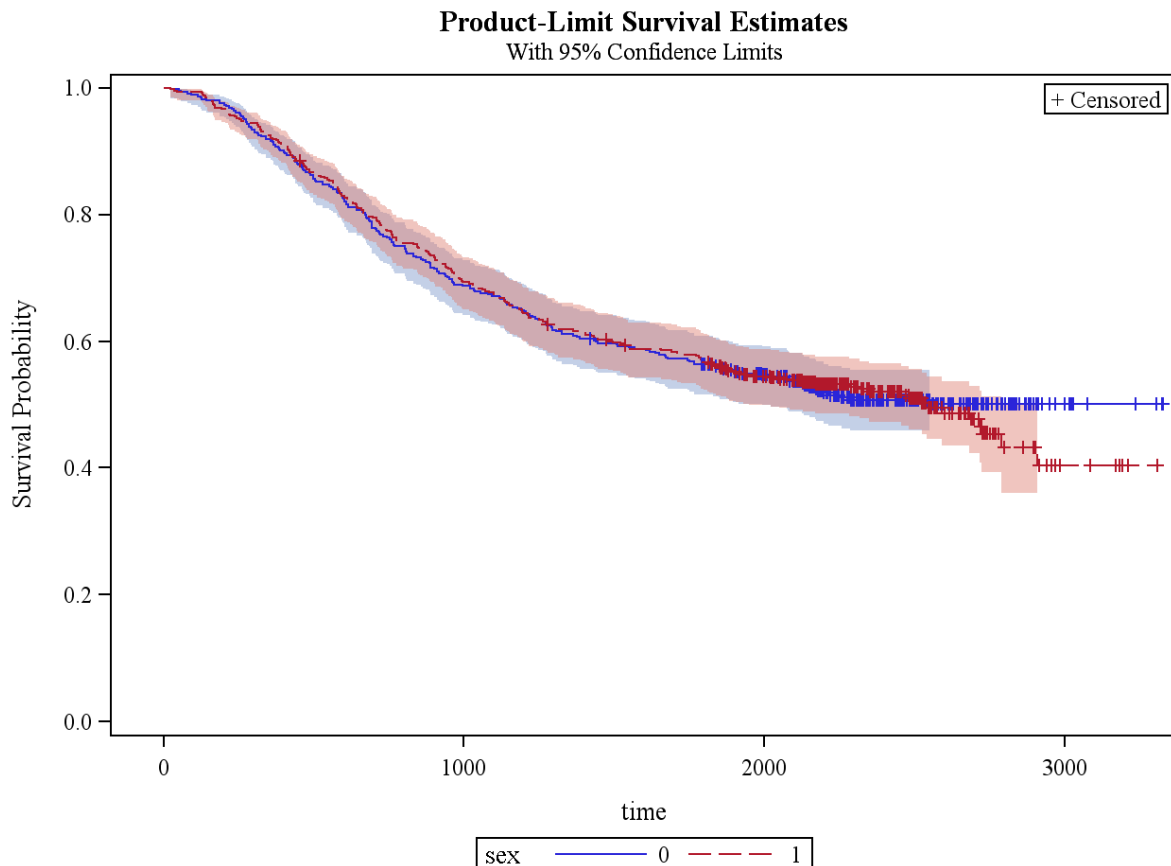
### 5 Jahres Überlebenswahrscheinlichkeit:

Lev: 0.5321  
Lev+5FU: 0.6340  
OBS: 0.5257

## 2.3 Überlebenszeiten in Abhängigkeit von Geschlecht

**Tabelle 12** Zusammenfassung der Zensierten und Unzensierten Werte von Geschlecht

Stratum	sex	Total	Failed	Censored	Percent Censored
1	w	445	215	230	51.69
2	m	484	237	247	51.03
Total		929	452	477	51.35



**Abbildung 4** Kaplan-Meier bei Frauen und Männern mit 95%-Konfidenzintervall

Test of Equality over Strata

Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	0.0202	1	0.8871
Wilcoxon	0.0464	1	0.8295
-2Log(LR)	0.0379	1	0.8456

Log-Rank-Test:  
0.8871 > 0.05  
→ nicht signifikant

An Hand des Wertes des Log-Ranks ( $p=0.8871$ ) kann man keinen signifikanten Einfluss des Geschlechts erkennen.

### 5 Jahres Überlebenswahrscheinlichkeit:

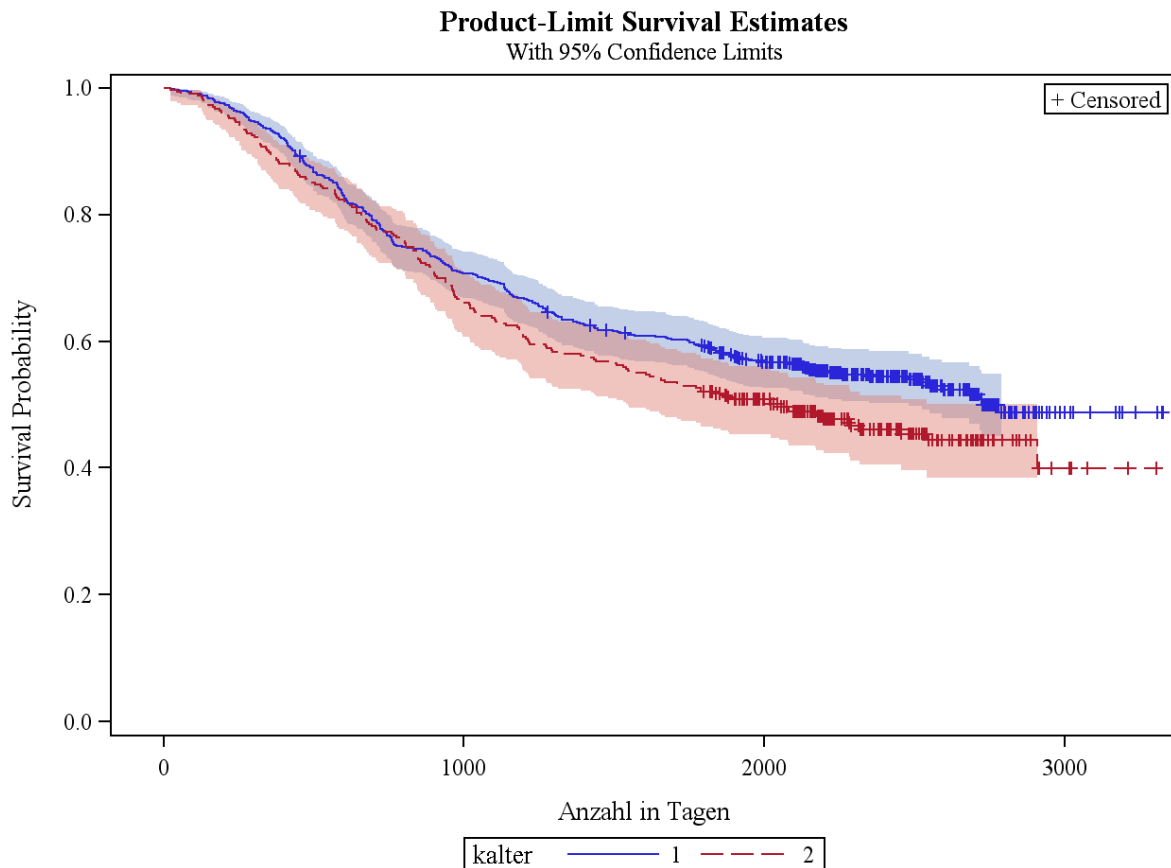
weiblich: 0.5616  
männlich: 0.5649



## 2.4 Überlebenszeiten in Abhängigkeit von Altersklassen

**Tabelle 13** Zusammenfassung der Zensierten und Unzensierten Werte von Altersklassen

Stratum	alter	Total	Failed	Censored	Percent Censored
1	< 65	595	274	321	53.95
2	> 65	334	178	156	46.71
Total		929	452	477	51.35



**Abbildung 5** Kaplan-Meier bei Altersklassen mit 95%-Konfidenzintervall

Test of Equality over Strata

Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	4.4108	1	0.0357
Wilcoxon	3.6415	1	0.0564
-2Log(LR)	4.9904	1	0.0255

Log-Rank-Test:  
0.0357 < 0.05  
→ signifikant

An Hand des Wertes des Log-Ranks ( $p=0.0357$ ) kann man einen signifikanten Einfluss des Alters erkennen.

### 5 Jahres Überlebenswahrscheinlichkeit:

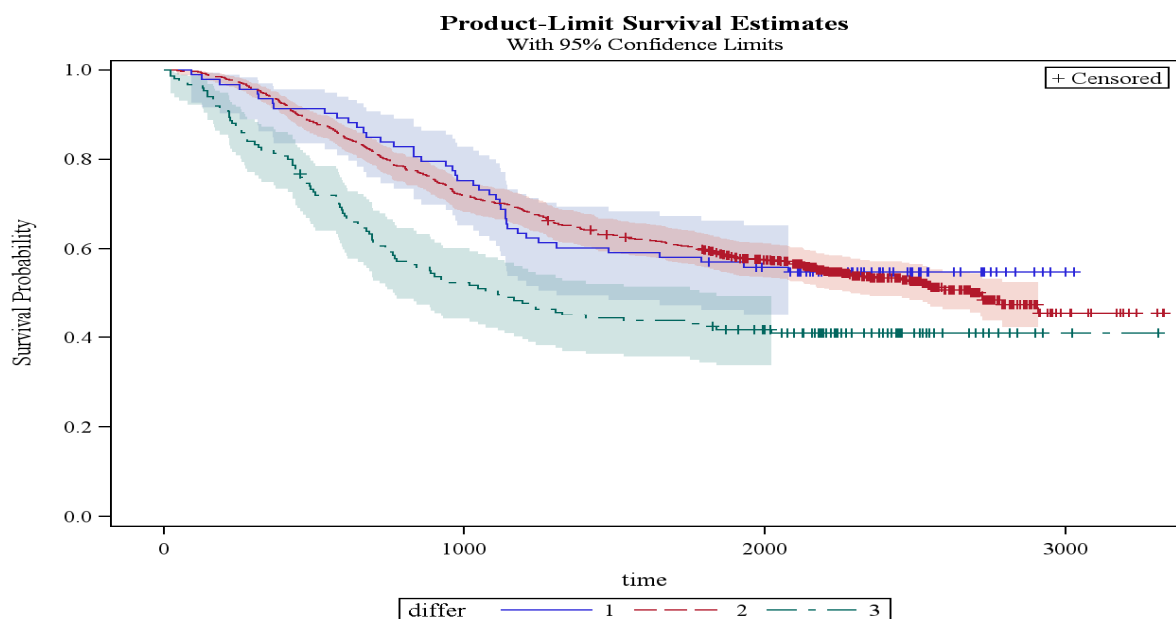
< 65 Jahre: 0.5872  
> 65 Jahre: 0.5210

## 2.5 Überlebenszeiten in Abhängigkeit von der Differenzierbarkeit der Tumorzellen

**Tabelle 14** Zusammenfassung der Zensierten und Unzensierten Werte von der Differenzierbarkeit der Tumorzellen

Stratum	differ	Total	Failed	Censored	Percent Censored
1	gute Differenzierbarkeit	93	42	51	54.84
2	mäßige Differenzierbarkeit	663	311	352	53.09
3	geringe Differenzierbarkeit	150	88	62	41.33
Total		906	441	465	51.32

**Note:** 23 observations with invalid time, censoring, or strata values were deleted.



**Abbildung 6** Kaplan-Meier nach Differenzierbarkeit der Tumorzellen mit 95%-Konfidenzintervall

Test of Equality over Strata

Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	17.1891	2	0.0002
Wilcoxon	26.9525	2	<.0001
-2Log(LR)	15.1054	2	0.0005

Log-Rank-Test:  
0.0002 < 0.05  
→ signifikant

An Hand des Wertes des Log-Ranks ( $p < 0.0002$ ) kann man einen signifikanten Einfluss der Differenzierbarkeit der Tumorzellen erkennen.

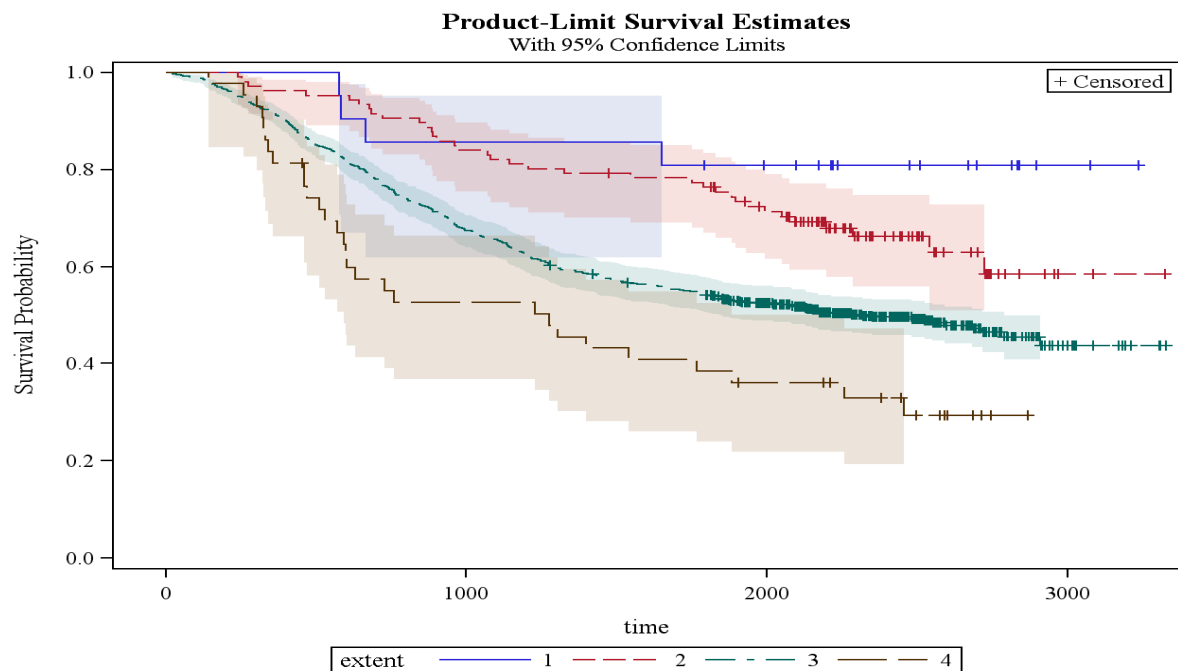
### 5 Jahres Überlebenswahrscheinlichkeit:

Gute Differenzierbarkeit:	0.5699
Mäßige Differenzierbarkeit:	0.5938
Geringe Differenzierbarkeit:	0.4237

## 2.6 Überlebenszeiten in Abhängigkeit von der Ausdehnung des Tumors

**Tabelle 15** Zusammenfassung der Zensierten und Unzensierten Werte von der Ausdehnung des Tumors

Stratum	extent	Total	Failed	Censored	Percent Censored
1	Darmschleimhaut	21	4	17	80.95
2	Darmmuskulatur	106	36	70	66.04
3	Äußere Darmwand	759	383	376	49.54
4	Darmdurchbruch	43	29	14	32.56
Total		929	452	477	51.35



**Abbildung 7** Kaplan-Meier nach Tumorausdehnung 95%-Konfidenzintervall

### Test of Equality over Strata

Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	26.9418	3	<.0001
Wilcoxon	29.0646	3	<.0001
-2Log(LR)	30.6748	3	<.0001

Log-Rank-Test:  
.0001 < 0.05  
→ signifikant

An Hand des Wertes des Log-Ranks ( $p < .0001$ ) kann man einen signifikanten Einfluss der Tumorausdehnung erkennen.

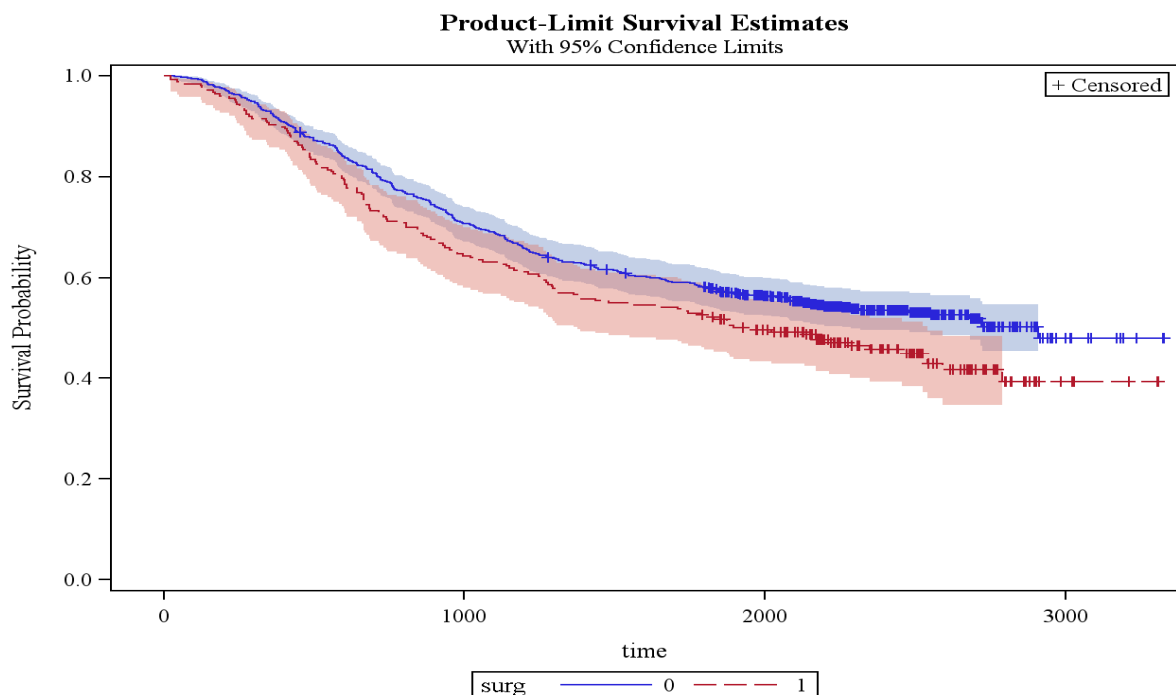
### 5 Jahres Überlebenswahrscheinlichkeit:

Darmschleimhaut: 0.8095  
Darmmuskulatur: 0.7638  
Äußere Darmwand: 0.5399  
Darmdurchbruch: 0.3830

## 2.7 Überlebenszeiten in Abhängigkeit von der Dauer der Behandlung

**Tabelle 16** Zusammenfassung der Zensierten und Unzensierten Werte von der Dauer der Behandlung

Stratum	surg	Total	Failed	Censored	Percent Censored
1	kurzer Zeitraum	682	316	366	53.67
2	langer Zeitraum	247	136	111	44.94
<hr/>					
Total		929	452	477	51.35



**Abbildung 8** Kaplan-Meier nach Dauer der Behandlung mit 95%-Konfidenzintervall

Test of Equality over Strata

Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	5.1961	1	0.0226
Wilcoxon	4.6288	1	0.0314
-2Log(LR)	4.7884	1	0.0287

Log-Rank-Test:  
0.0226 < 0.05  
→ signifikant

An Hand des Wertes des Log-Ranks ( $p < 0.0226$ ) kann man einen signifikanten Einfluss der Dauer der Behandlungszeit erkennen.

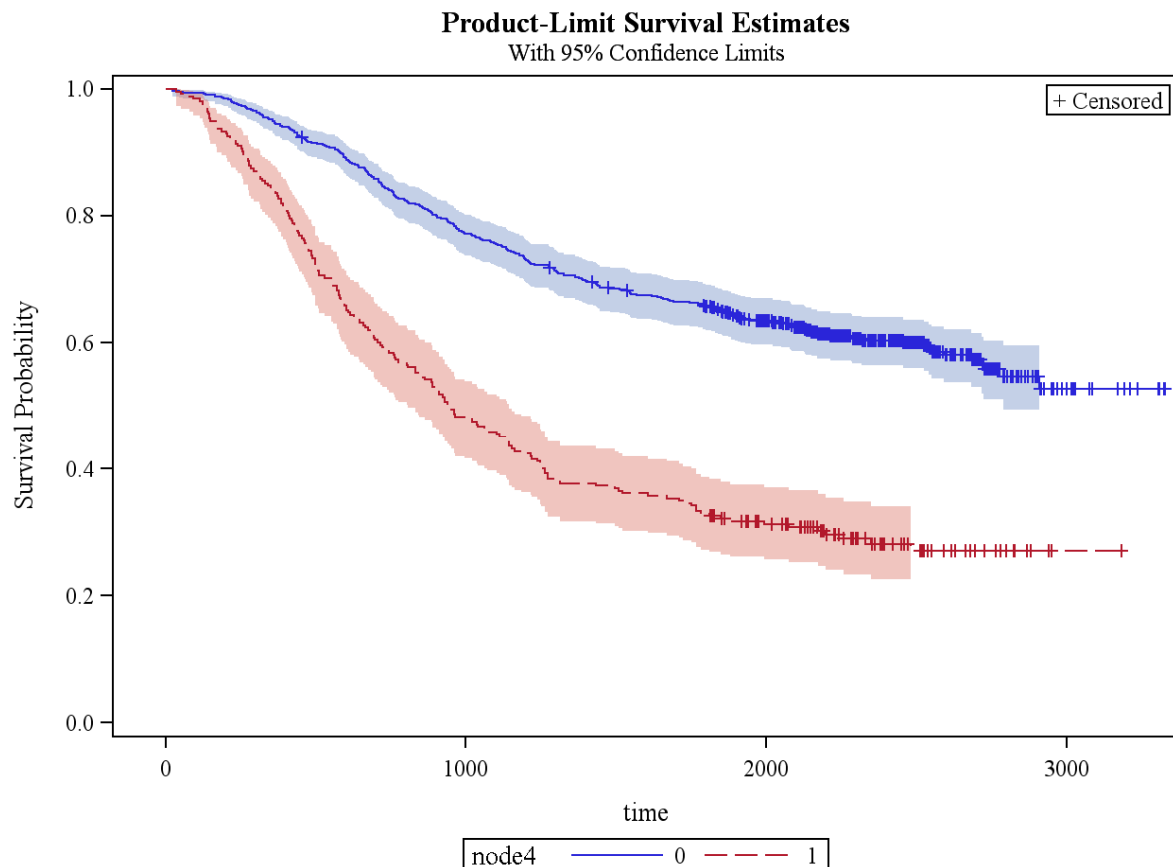
### 5 Jahres Überlebenswahrscheinlichkeit:

Kurzer Zeitraum: 0.5797  
Langer Zeitraum: 0.5222

## 2.8 Überlebenszeiten in Abhängigkeit von Lymphknotenklassen

**Tabelle 17** Zusammenfassung der Zensierten und Unzensierten Werte von Lymphknotenklassen

Stratum	node4	Total	Failed	Censored	Percent Censored
1	<=4 Lymphknoten	674	271	403	59.79
2	>4 Lymphknoten	255	181	74	29.02
Total		929	452	477	51.35



**Abbildung 9** Kaplan-Meier für Lymphknotenklassen mit 95%-Konfidenzintervall

Test of Equality over Strata

Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	104.9156	1	<.0001
Wilcoxon	111.9857	1	<.0001
-2Log(LR)	93.8439	1	<.0001

Log-Rank-Test:  
0.0001 < 0.05  
→ signifikant

An Hand des Wertes des Log-Ranks ( $p < 0.0001$ ) kann man einen signifikanten Einfluss der Lymphknoten erkennen.

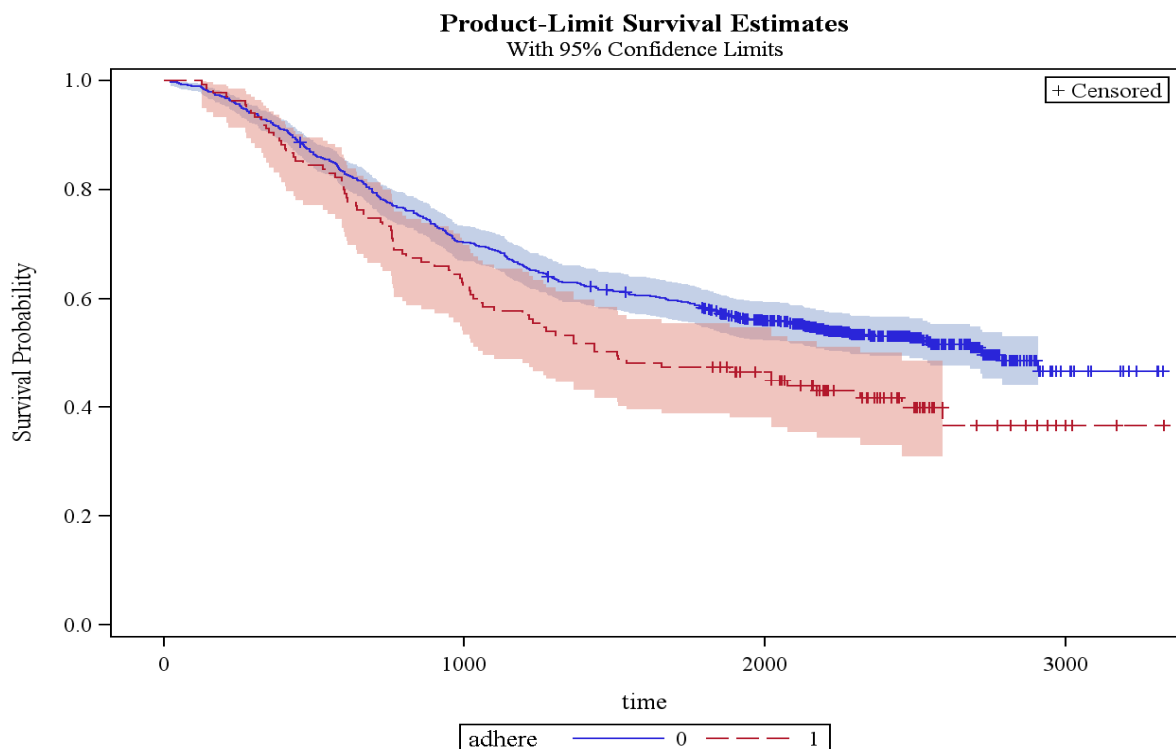
### 5 Jahres Überlebenswahrscheinlichkeit:

≤4 Lymphknoten: 0.6550  
> 4 Lymphknoten: 0.3255

## 2.9 Überlebenszeiten in Abhängigkeit vom Tumorbefall anliegender Organe

**Tabelle 18** Zusammenfassung der Zensierten und Unzensierten Werte von dem Tumorbefall anliegender Organe

Stratum	adhere	Total	Failed	Censored	Percent Censored
1	kein Befall der Organe	794	373	421	53.02
2	Befall der Organe	135	79	56	41.48
Total		929	452	477	51.35



**Abbildung 10** Kaplan-Meier nach Tumorbefall der anliegenden Organe mit 95%-Konfidenzintervall

Test of Equality over Strata

Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	6.5213	1	0.0106
Wilcoxon	5.6107	1	0.0179
-2Log(LR)	6.6238	1	0.0101

Log-Rank-Test:  
0.0106 < 0.05  
→ signifikant

An Hand des Wertes des Log-Ranks ( $p < 0.0106$ ) kann man einen signifikanten Einfluss des Tumorbefalls der anliegenden Organe erkennen.

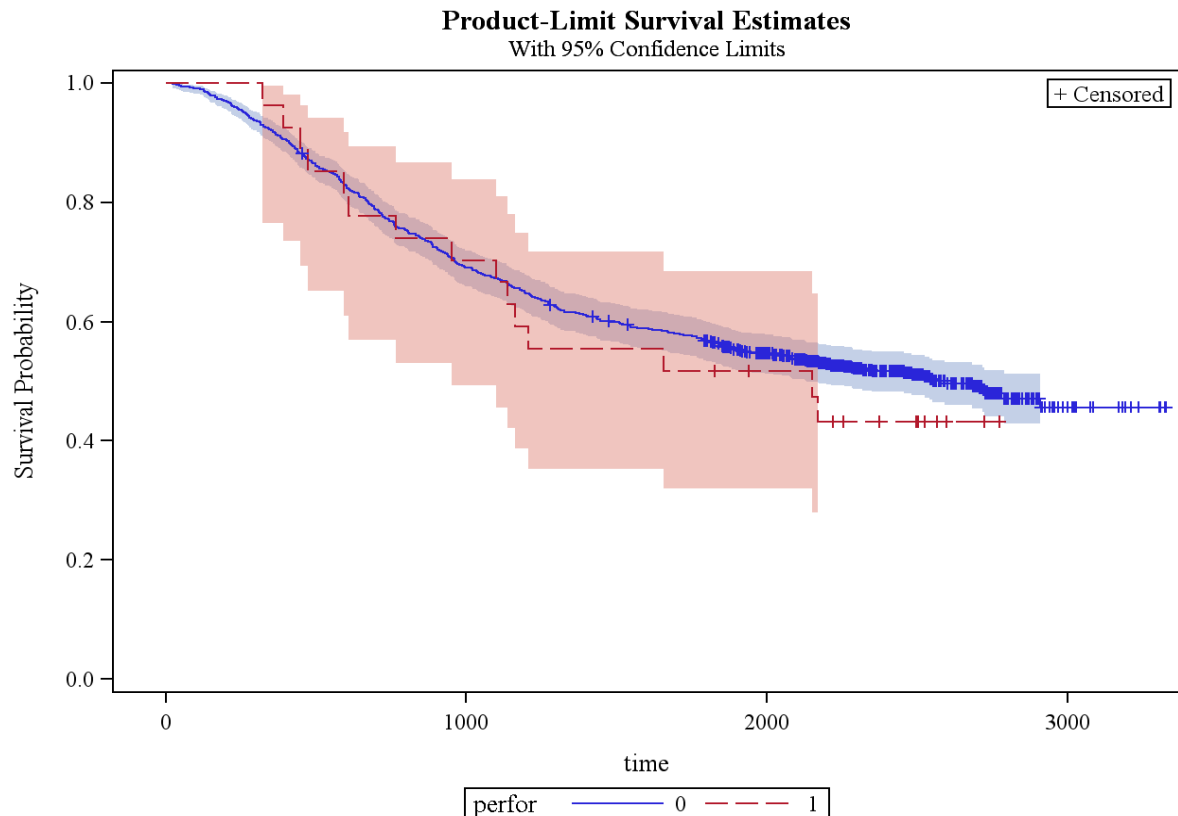
### 5 Jahres Überlebenswahrscheinlichkeit:

Keine Tumorbefall der Organe: 0.5798  
Tumorbefall der Organe: 0.4741

## 2.10 Überlebenszeiten in Abhängigkeit von dem Darmdurchbruch

**Tabelle 13** Zusammenfassung der Zensierten und Unzensierten Werte von dem Darmdurchbruch

Stratum	perfore	Total	Failed	Censored	Percent Censored
1	kein Darmdurchbruch	902	437	465	51.55
2	Darmdurchbruch	27	15	12	44.44
<hr/>					
Total		929	452	477	51.35



**Abbildung 11** Kaplan-Meier nach Darmdurchbruch mit 95%-Konfidenzintervall

Test of Equality over Strata

Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	0.3485	1	0.5550
Wilcoxon	0.1560	1	0.6929
-2Log(LR)	0.3681	1	0.5441

Log-Rank-Test:  
0.5550 > 0.05  
→ nicht signifikant

An Hand des Wertes des Log-Ranks ( $p > 0.5550$ ) kann man keinen signifikanten Einfluss des Darmdurchbruchs erkennen.

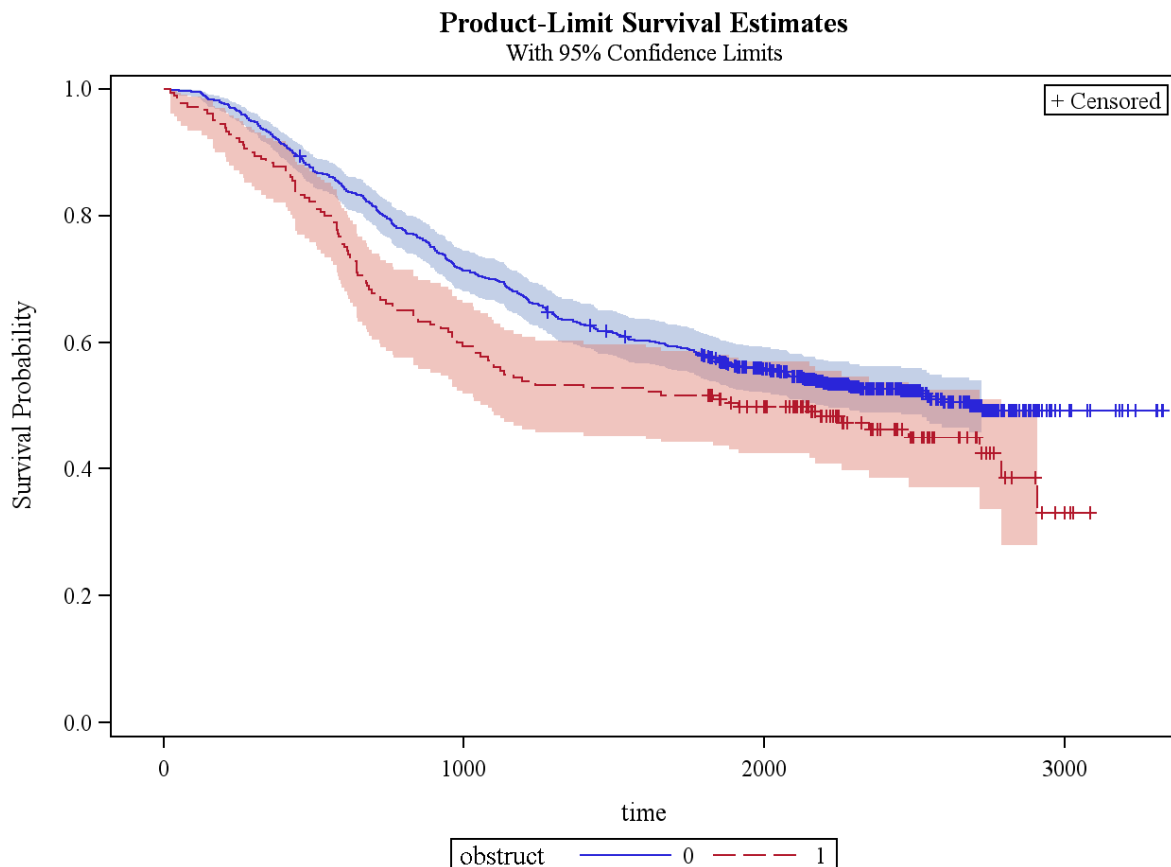
### 5 Jahres Überlebenswahrscheinlichkeit:

Keine Darmverstopfung: 0.5658  
Darmverstopfung: 0.5185

## 2.11 Überlebenszeiten in Abhängigkeit von der Darmverstopfung

**Tabelle 14** Zusammenfassung der Zensierten und Unzensierten Werte von der Darmverstopfung

Stratum	obstruct	Total	Failed	Censored	Percent Censored
1	keine Darmverstopfung	749	354	395	52.74
2	Darmverstopfung	180	98	82	45.56
Total		929	452	477	51.35



**Abbildung 12** Kaplan-Meier nach Darmverstopfung mit 95%-Konfidenzintervall

Test of Equality over Strata

Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	5.3493	1	0.0207
Wilcoxon	6.9034	1	0.0086
-2Log(LR)	4.8314	1	0.0279

Log-Rank-Test:  
0.0207 < 0.05  
→ signifikant

An Hand des Wertes des Log-Ranks ( $p < 0.0207$ ) kann man einen signifikanten Einfluss der Darmverstopfung erkennen.

### 5 Jahres Überlebenswahrscheinlichkeit:

Keine Darmverstopfung: 0.5758  
Darmverstopfung: 0.5167



### Zusammenfassung:

Nach Kaplan-Meier scheinen folgende Faktoren Einfluss auf die Überlebenswahrscheinlichkeit zu haben. (Signifikanzniveau  $\alpha=0.05$ )

Einflussgröße	Log-Rank-Wert
Behandlungsart	0.0029
Altersklasse < 65 >	0.0357
Differenzierbarkeit der Tumorzellen	0.0002
Ausdehnung des Tumors	<0.0001
Dauer der Behandlung	0.0226
Lymphknotenklasse < 4 >	<0.0001
Tumorbefall der anliegenden Organe	0.0106
Darmverstopfung durch den Tumor	0.0207

Nach Kaplan-Meier scheinen folgende Faktoren keinen Einfluss auf die Überlebenswahrscheinlichkeit zu haben. (Signifikanzniveau  $\alpha=0.05$ )

Einflussgröße	Log-Rank-Wert
Geschlecht	0.8871
Darmdurchbruch durch den Tumor	0.5550

## 3 Cox-Regression

Häufig wird versucht aus dem Vergleich von Überlebensfunktionen eine Aussage zu machen wie „Gruppe A unterliegt einem 1,5 mal höheren Risiko an der Krankheit zu erkranken als Gruppe B“. Ein besonderes Verfahren stellt die Cox-Regression dar, da sie in der Lage ist mehrere Faktoren gleichzeitig in ihrem Einfluss auf die Überlebenszeit zu prüfen. Eine Erhöhung des Risikos um einen bestimmten Faktor  $c$  (hier 1,5) bedeutet folgendes für die Überlebensfunktionen beider Gruppen:  $S_B(t)=[S_A(t)]^c$

Es wird deutlich, dass der Unterschied der beiden Funktionen nicht von der Zeit  $t$  abhängt, sondern von anderen Einflussgrößen. Um diesen Zusammenhang zu prüfen kann die sogenannte Ausfallrate  $h(t)$  (hazard function) verwendet werden.

Mit Hilfe der Überlebensfunktion lässt sich die Ausfallrate bestimmen. Diese gibt an, dass zum Zeitpunkt  $t$  ein Patient, der den Zeitpunkt  $t$  überlebt hat, kurz nach  $t$  ausfällt. Sie ist mathematisch gesehen die Ableitung der negativ logarithmierten Überlebensfunktion:  $-(\log(S(t)))' = -S'(t)/S(t) = h(t)$

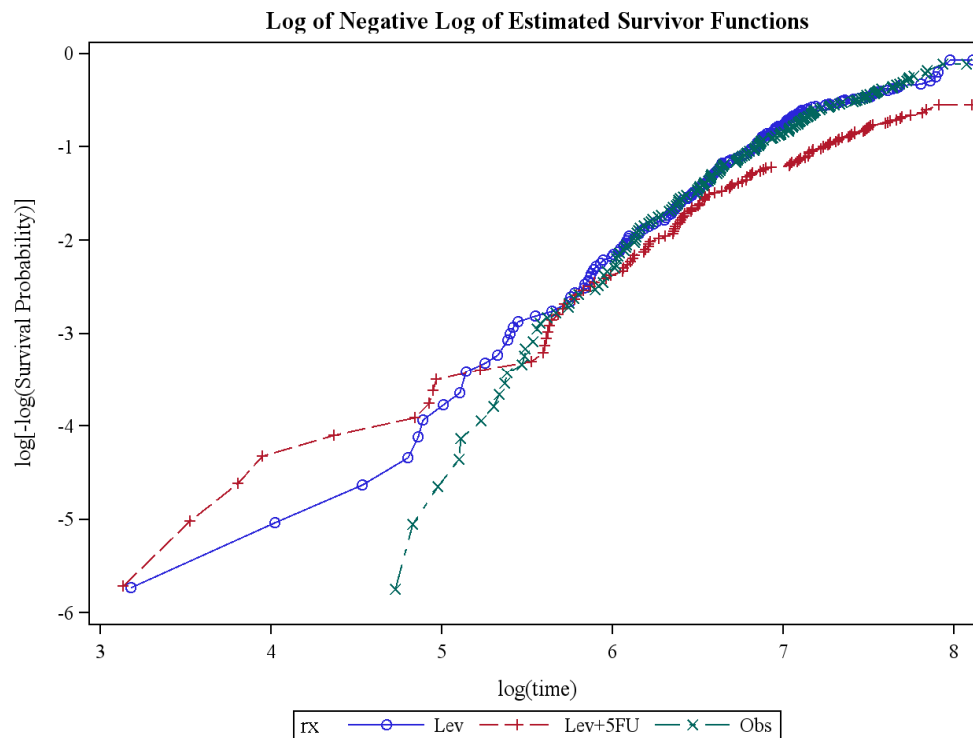
Da der Faktor  $c$  von verschiedenen Einflussgrößen abhängt, versucht man diesen Einfluss so einfach wie möglich zu erklären, in dem man ihn in eine lineare Beziehung setzt. Hierbei werden die Einflussfaktoren nicht gruppiert, sondern ihre genauen Werte verwendet. Es ist nicht sinnvoll den Faktor  $c$  in eine lineare Beziehung zu setzen, da dieser nur positiv sein kann.

Allgemeiner Ausdruck:  $\log(c) = a + b_1x_1 + b_2x_2 + \dots + b_nx_n$

Das allgemeine Verfahren der Beschreibung von  $\log(c)$  durch solche Einflussvariablen heißt Cox-Regression und ist analog zur multiplen Regression. Doch zuvor müssen Modellannahmen geprüft werden.

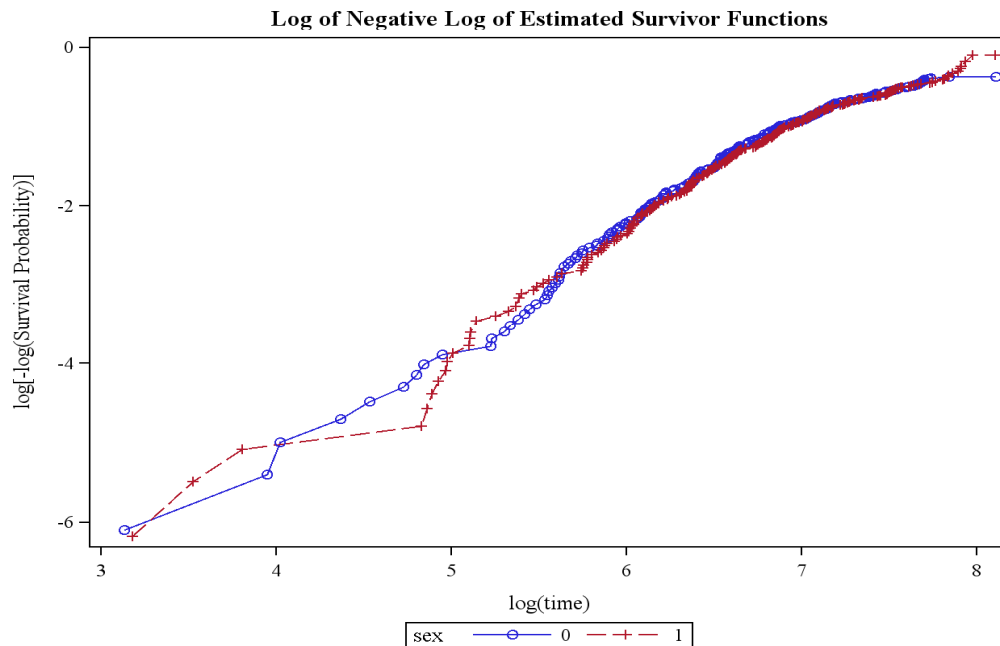
### 3.1 Prüfung der Modellannahmen

Die Cox-Regression geht von sogenannten proportional hazards aus, d.h. dass über den gesamten Beobachtungszeitraum (unabhängig von der Zeit) der Quotient zweier Ausfallraten konstant ist. Um diese Annahme zu prüfen wird ein geschichtetes Modell geschätzt und für jede Schicht die sogenannte log-minus-log-survivor-Funktion dargestellt. Die so dargestellten Kurven müssen über den gesamten Zeitverlauf „relativ“ parallel verlaufen. Diese Annahme kann mit PROC LIFETEST von SAS über die Option PLOT=(LLS) überprüft werden.



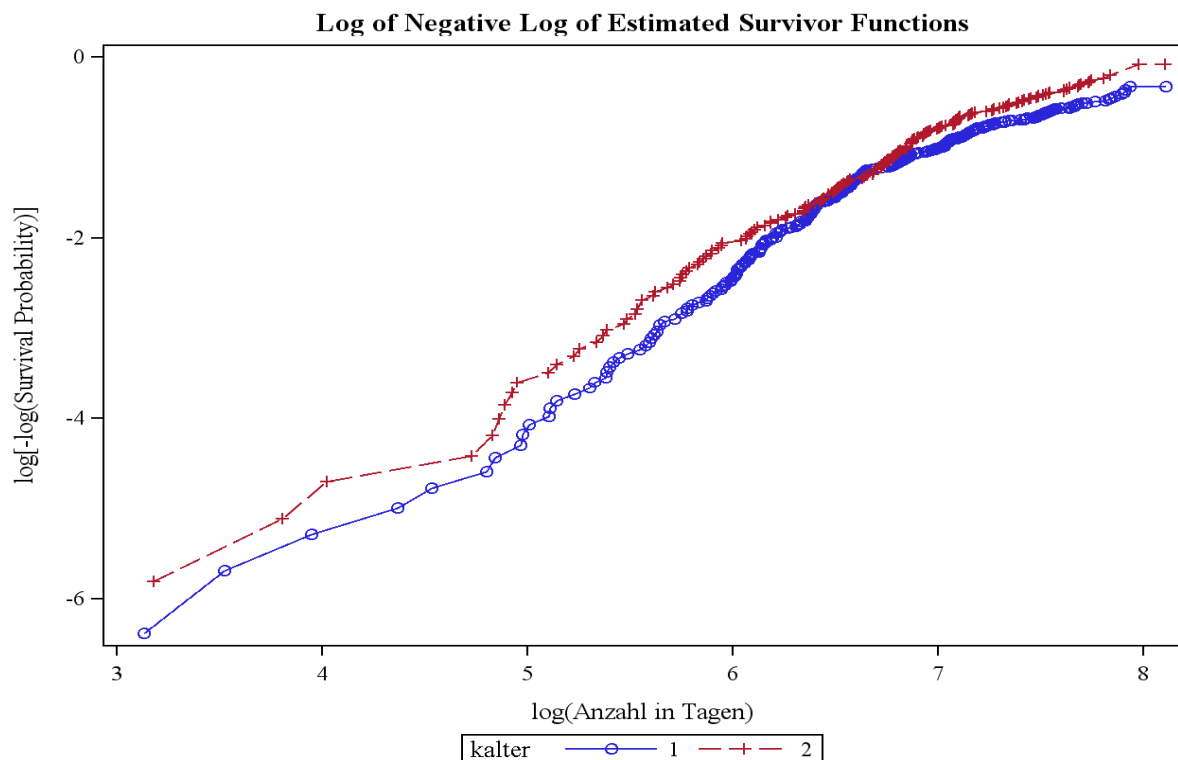
**Abbildung 13** Log-Grafik für Behandlungsart

Die Kurven für die Behandlungsart sind „relativ“ parallel. Die Modellvoraussetzung für die Cox-Regression ist somit erfüllt.



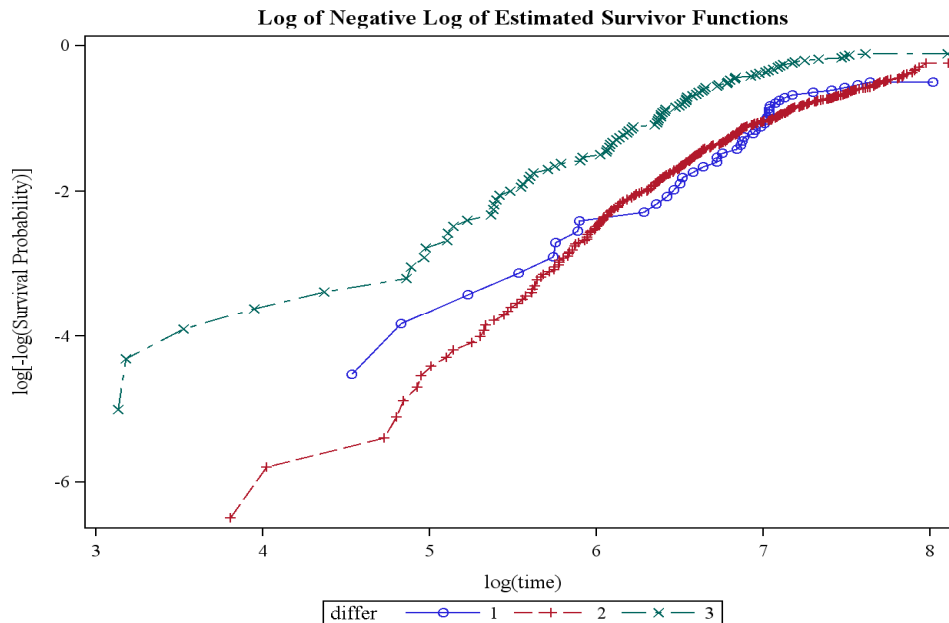
**Abbildung 14** Log-Grafik für Geschlecht

Die beiden Kurven für Geschlecht sind eher nicht parallel. Die Modellvoraussetzung für die Cox-Regression ist somit nicht erfüllt, aber nicht dramatisch unterschiedlich.



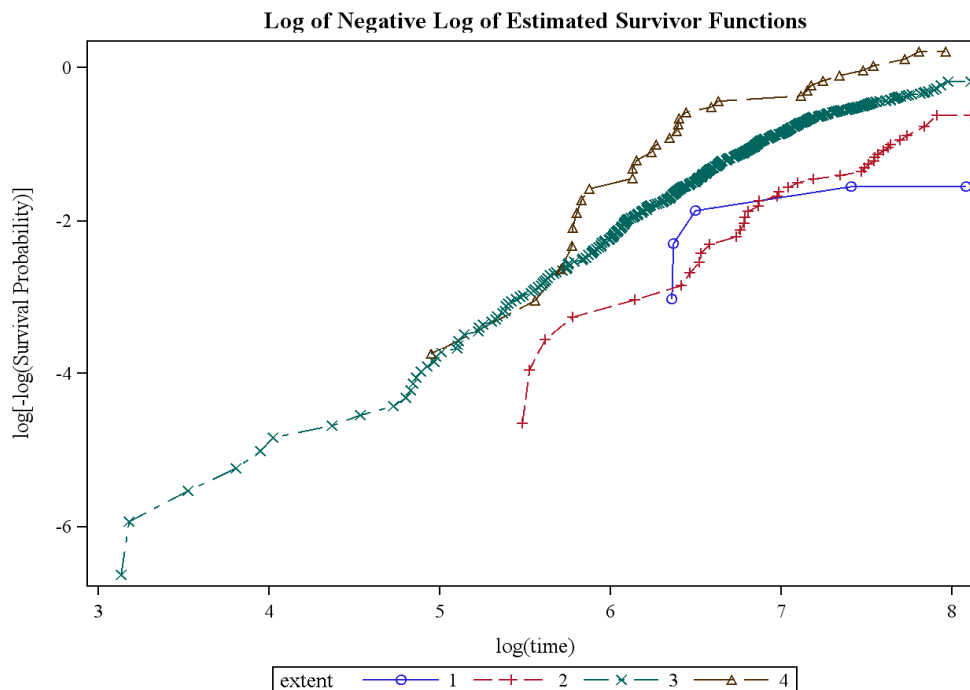
**Abbildung 15** Log-Grafik für Alter

Die beiden Kurven für das Alter sind „relativ“ parallel. Die Modellvoraussetzung für die Cox-Regression ist somit erfüllt.



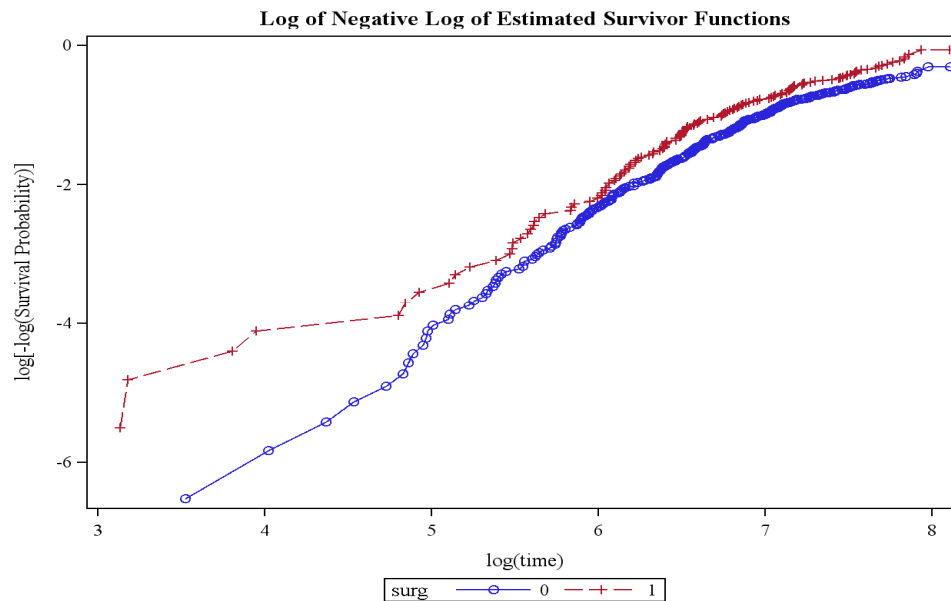
**Abbildung 16** Log-Grafik für die Differenzierbarkeit der Tumorzellen

Hier ist es schwierig zu erkennen, ob die Modellvoraussetzung der Cox-Regression erfüllt ist, da die Kurve mit guter Differenzierbarkeit der Tumorzellen sich mit der Kurve mäßige Differenzierbarkeit der Tumorzellen überschneidet, wobei die andere Kurve „relativ“ parallel verläuft.



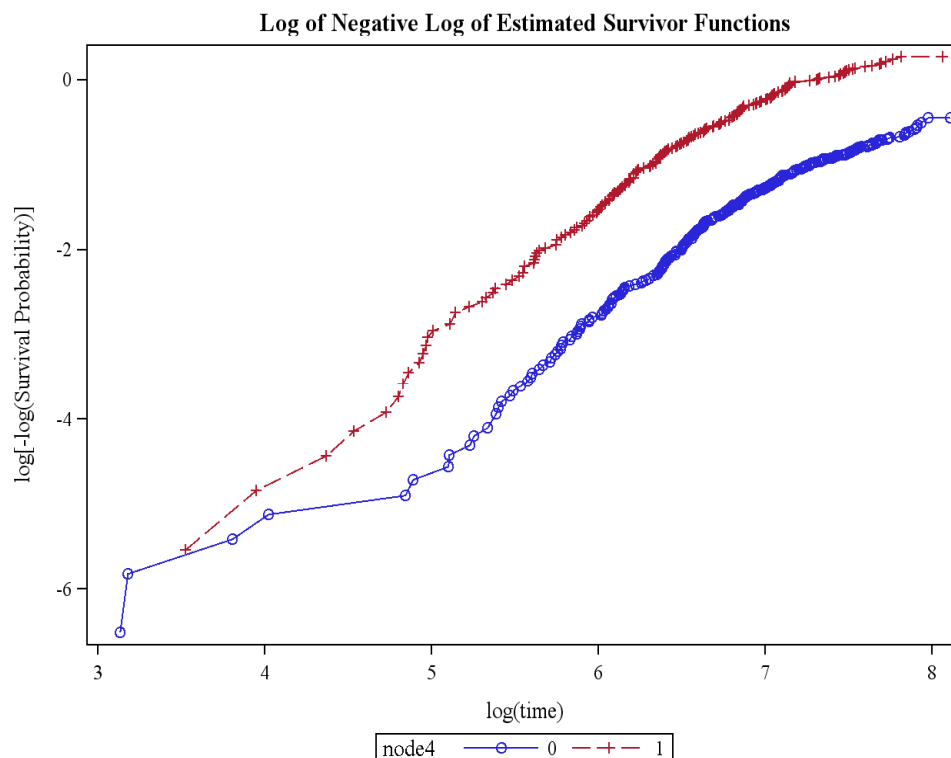
**Abbildung 17** Log-Grafik für die Tumorausdehnung

Hier ist es schwierig zu erkennen, ob die Modellvoraussetzung der Cox-Regression erfüllt ist, da die Kurve mit Darmschleimhaut sich mit der Kurve Darmmuskulatur überschneidet, wobei die anderen beiden „relativ“ parallel verlaufen.



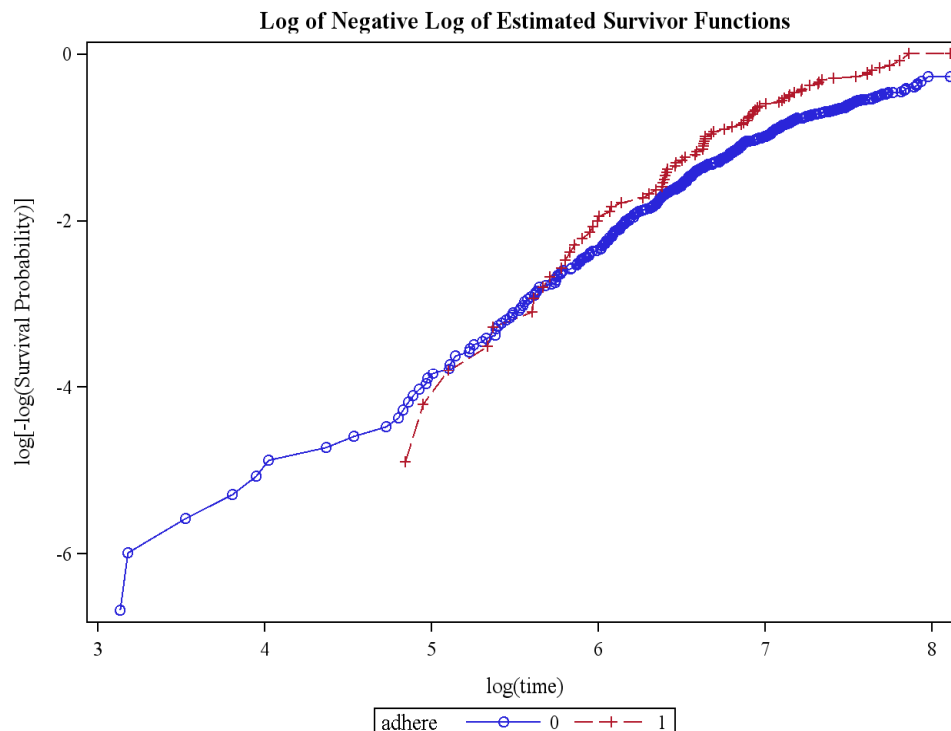
**Abbildung 18** Log-Grafik für die Behandlungsdauer

Die beiden Kurven für die Behandlungsdauer sind „relativ“ parallel. Die Modellvoraussetzung für die Cox-Regression ist somit erfüllt.



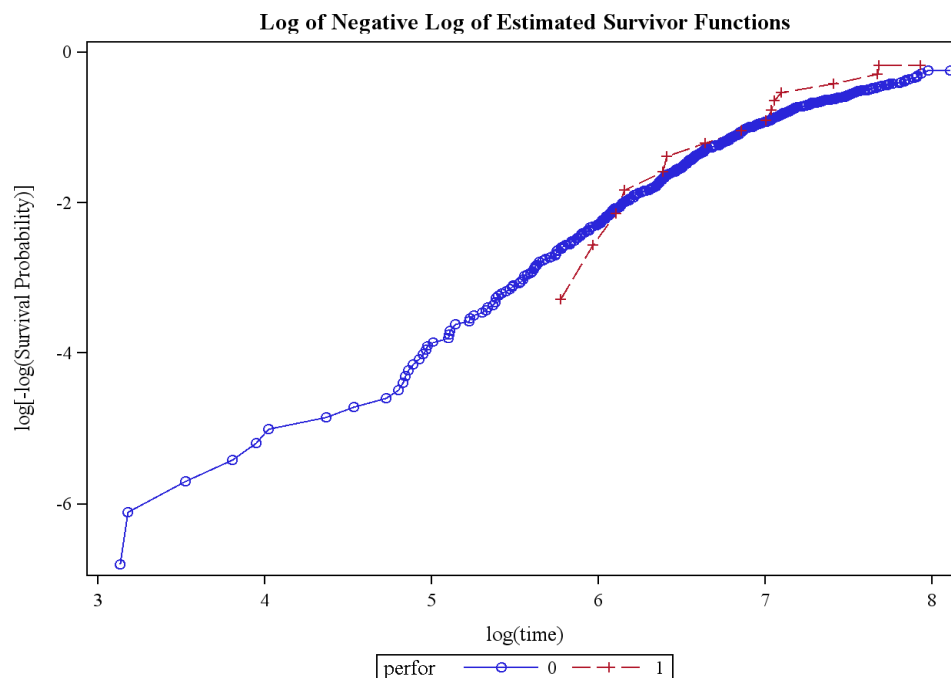
**Abbildung 19** Log-Grafik für Lymphknotenklassen

Die Kurven für die Lymphknotenklassen sind „relativ“ parallel. Die Modellvoraussetzung für die Cox-Regression ist somit erfüllt.



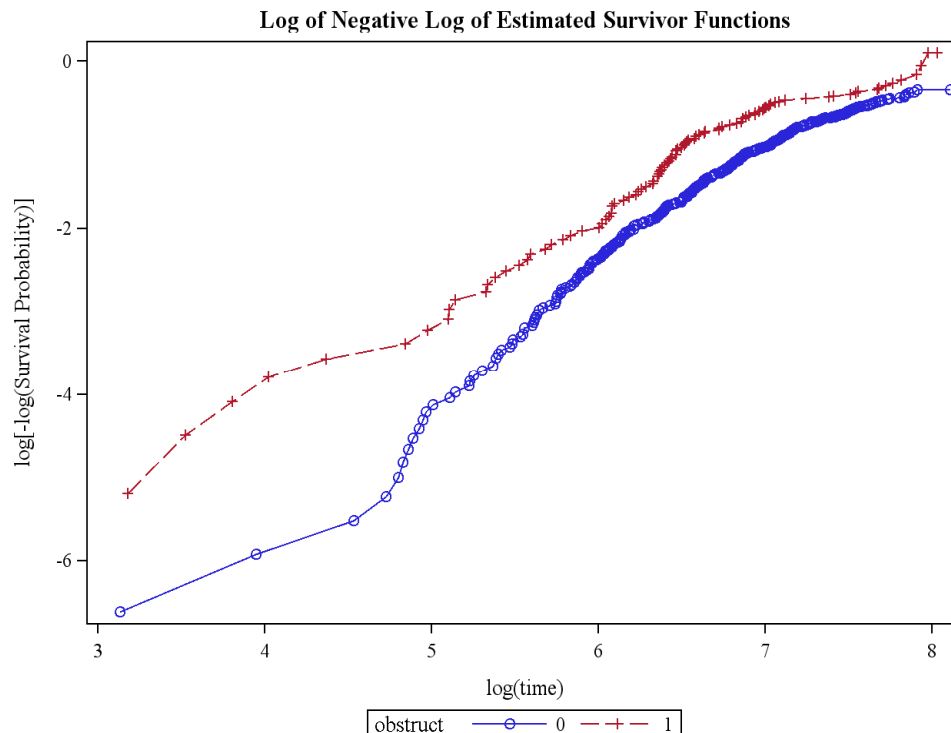
**Abbildung 20** Log-Grafik für den Tumorbefall der anliegenden Organe

Hier ist es schwierig zu erkennen, ob die Modellvoraussetzung der Cox-Regression erfüllt ist, da die Kurve mit Tumorbefall der anliegenden Organe sich mit der Kurve kein Tumorbefall der anliegenden Organe überschneidet.



**Abbildung 21** Log-Grafik für die Darmdurchbruch

Hier ist es schwierig zu erkennen, ob die Modellvoraussetzung der Cox-Regression erfüllt ist, da die Kurve mit Darmdurchbruch sich mit der Kurve kein Darmdurchbruch überschneidet.



**Abbildung 22** Log-Grafik für die Darmverstopfung

Die beiden Kurven für die Darmverstopfung sind „relativ“ parallel. Die Modellvoraussetzung für die Cox-Regression ist somit erfüllt.

Zusammengefasst sprechen die Grafiken nicht gegen die Anwendung des Cox Modells.

## 3.2 Ergebnis des Cox-Modells

Aufruf der Cox-Regression in SAS:

```
PROC PHREG data=colon.colon;
  CLASS rx sex obstruct perfor adhere differ extent surg node4;
  model time*status(0) = rx sex age obstruct perfor adhere differ
    extent surg node4;
  HAZARDATIO rx;
  HAZARDATIO sex;
  HAZARDATIO age;
  HAZARDATIO obstruct;
  HAZARDATIO perfor;
  HAZARDATIO adhere;
  HAZARDATIO differ;
  HAZARDATIO extent;
  HAZARDATIO surg;
  HAZARDATIO node4;
RUN;
```

Output besteht aus folgenden Abschnitten:

- **Modellinformation:** Es wird dort die abhängige Variable, die zensierte Variable und dessen Wert dargestellt. Es werden dann noch die Anzahl der eingelesenen Daten und die Anzahl der benutzten Daten angezeigt.
- **Class Level Information:** Dort wird die Kodierung der einzelnen Variablen dargestellt.

- **Summary of the Number of the Event and censored Values:** Dort wird die Anzahl der Zensierten und der Event Variablen angegeben.
- **Testing Global Null Hypotheses:** Dort wird die Überschreitungswahrscheinlichkeit der einzelnen Variablen angezeigt.
- **Hazard Ratio for :** Es wird die Variable angezeigt und deren Wert bei dieser Stichprobe. Gleichzeitig sieht man das 95% Konfidenzintervall in der der eigentlich wahre Wert von der Grundgesamtheit liegt.

### 3.2.1 Volles Modell

Zuerst wird das Gesamte Modell mit allen Variablen des Datensatzes auf den Einfluss der einzelnen Variablen, die einen Einfluss auf den Tod durch Dickdarmkrebs haben. Dabei wird ein Signifikanzniveau von 5% festgelegt.

Danach wird das Modell erneut gerechnet allerdings mit der Backward-Selektion. Bei dieser Selektion werden alle die Variablen verwendet, deren p-Wert die 10% nicht überschreitet. Alle anderen Variablen werden verworfen.

Danach rechnet man noch einmal das volle Modell, diesmal nur ohne die Variablen die über diesen p-Wert von 10% waren.

```

Model Information

Data Set      colon.colon
Dependent Variable  time      Überlebenszeit in Tage
Censoring Variable  status      Status am Ende der Studie
Censoring Value(s)  0
Ties Handling      BRESLOW

```

```

Number of Observations Read      929
Number of Observations Used      906

```

Fallzahl 906 23 Missings Daher geringer Selektionsbias möglich
--

#### Class Level Information

Class	Value	Design Variables
rx	Lev	1 0
	Lev+5FU	0 1
	Obs	0 0
sex	0	1
	1	0
differ	1	1 0
	2	0 1
	3	0 0
extent	1	1 0 0
	2	0 1 0
	3	0 0 1
	4	0 0 0
surg	0	1
	1	0
Node4	0	1
	1	0



adhere	0	1
	1	0
perfor	0	1
	1	0
obstruct	0	1
	1	0

Die Dummy-Codierung ist eine Null-Eins-Codierung kategorialer Variablen mit den Ausprägungen 1 und 0, die als Indikator für das Vorhandensein einer Ausprägung einer mehrstufigen Variablen dient (vorhanden = 1, nicht vorhanden = 0).

#### Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
906	441	465	51.32

#### Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

#### Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	5695.191	5555.420
AIC	5695.191	5583.420
SBC	5695.191	5640.666

#### Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	139.7710	14	<.0001
Score	153.1146	14	<.0001
Wald	142.9882	14	<.0001

#### Type 3 Tests

Effect	DF	Wald Chi-Square	Pr > ChiSq
rx	2	10.8970	0.0043
sex	1	0.1096	0.7406
age	1	3.6122	0.0574
differ	2	8.7880	0.0124
extent	3	13.3305	0.0040
surg	1	5.2409	0.0221
node4	1	87.3930	<.0001
adhere	1	1.3929	0.2379
perfor	1	0.0043	0.9477
obstruct	1	5.8044	0.0160

#### Test $\beta=0$ :

Behandlungsart: 0.0043 < 0.05  
→ signifikant

Geschlecht: 0.7406 > 0.05  
→ nicht signifikant

Alter: 0.0574 > 0.05  
→ nicht signifikant  
(Tendenz)

Differenzierbarkeit: 0.0124 < 0.05  
→ signifikant

Tumorausdehnung: 0.0040 < 0.05  
→ signifikant

Behandlungszeitraum: 0.0221 < 0.05  
→ signifikant

Lymphknotenbefall: <.0001 < 0.05  
→ signifikant

Anliegende Organe: 0.2379 > 0.05  
→ nicht signifikant

Durchbruch: 0.9477 > 0.05  
→ nicht signifikant

Verstopfung: 0.0160 < 0.05  
→ signifikant

#### Analysis of Maximum Likelihood Estimates

Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
rx	Lev	1	-0.02058	0.11340	0.0329	0.8560
rx	Lev+5FU	1	-0.36357	0.12039	9.1208	0.0025
sex	0	1	-0.03190	0.09637	0.1096	0.7406
age		1	0.00781	0.00411	3.6122	0.0574
differ	1	1	-0.28477	0.19248	2.1890	0.1390
differ	2	1	-0.36343	0.12262	8.8750	0.0030
extent	1	1	-1.11109	0.54473	4.1604	0.0414
extent	2	1	-0.85947	0.26804	10.2812	0.0013
extent	3	1	-0.36563	0.21239	2.9635	0.0852
surg	0	1	-0.23878	0.10430	5.2409	0.0221
node4	0	1	-0.93548	0.10007	87.3930	<.0001
adhere	0	1	-0.15515	0.13146	1.3929	0.2379
perfor	0	1	-0.01768	0.26974	0.0043	0.9477
obstruct	0	1	-0.28663	0.11897	5.8044	0.0160

#### Analysis of Maximum Likelihood Estimates

Parameter		Hazard Ratio	Label	
rx	Lev	0.980	rx	Lev
rx	Lev+5FU	0.695	rx	Lev+5FU
sex	0	0.969	sex	0
age		1.008	age	
differ	1	0.752	differ	1
differ	2	0.695	differ	2
extent	1	0.329	extent	1
extent	2	0.423	extent	2
extent	3	0.694	extent	3
surg	0	0.788	surg	0
node4	0	0.392	node4	0
adhere	0	0.856	adhere	0
perfor	0	0.982	perfor	0
obstruct	0	0.751	obstruct	0

#### Hazard Ratios for rx

Description	Point Estimate	95% Wald Confidence Limits	
rx Lev vs Lev+5FU	1.409	1.109	1.791
rx Lev vs Obs	0.980	0.784	1.223
Lev+5FU vs Obs	0.695	0.549	0.880

Die „1“ ist in dem Konfidenzintervall nicht enthalten, dies bedeutet, dass man einen signifikanten Einfluss der Behandlungsart erkennen kann. Patienten, die Lev+5FU erhalten haben, haben ein 1,4-fach höhere Überlebenschance an Darmkrebs nicht zu versterben als Patienten nur mit der Behandlung Levamisol und nur einer Beobachtung.

#### Hazard Ratios for sex

Description	Point Estimate	95% Wald Confidence Limits	
sex m vs w	0.969	0.802	1.170

#### Hazard Ratios for age

Description	Point Estimate	95% Wald Confidence Limits	
age Unit=1	1.008	1.000	1.016

#### Hazard Ratios for differ

Description	Point Estimate	95% Wald Confidence Limits	
differ 1 vs differ 2	1.082	0.780	1.500
differ 1 vs differ 3	0.752	0.516	1.097
differ 2 vs differ 3	0.695	0.547	0.884

Die „1“ ist in dem Konfidenzintervall nicht enthalten, dies bedeutet, dass man einen signifikanten Einfluss der Differenzierung der Tumorzellen erkennen kann. Patienten, deren Tumorzellen gering differenzierbar sind, haben ein 0,695-fach höheres Risiko an Darmkrebs zu versterben als Patienten mit differenzierbaren Tumorzellen

#### Hazard Ratios for extent

Description	Point Estimate	95% Wald Confidence Limits	
extent 1 vs extent 2	0.778	0.275	2.198
extent 1 vs extent 3	0.475	0.176	1.279
extent 1 vs extent 4	0.329	0.113	0.957
extent 2 vs extent 3	0.610	0.431	0.865
extent 2 vs extent 4	0.423	0.250	0.716
extent 3 vs extent 4	0.694	0.458	1.052

#### Hazard Ratios for surg

Description	Point Estimate	95% Wald Confidence Limits	
Surg 0 vs 1	0.788	0.642	0.966

Die „1“ ist in dem Konfidenzintervall nicht enthalten, dies bedeutet, dass man einen signifikanten Einfluss der Behandlungsdauer des Kolonkarzinoms erkennen kann. Patienten, deren Behandlungsdauer kürzer ist, haben ein 0,788-fach höheres Risiko an Darmkrebs zu versterben als Patienten mit einer längeren Behandlungsdauer.

#### Hazard Ratios for node4

Description	Point Estimate	95% Wald Confidence Limits	
node4 0 vs 1	0.392	0.323	0.477

#### Hazard Ratios for adhere

Description	Point Estimate	95% Wald Confidence Limits	
adhereode 0 vs 1	0.856	0.662	1.108

#### Hazard Ratios for perfor

Description	Point Estimate	95% Wald Confidence Limits	
perfor 0 vs 1	0.982	0.579	1.667

#### Hazard Ratios for obstruct

Description	Point Estimate	95% Wald Confidence Limits	
obstruct 0 vs 1	0.751	0.595	0.948

Die „1“ ist in dem Konfidenzintervall nicht enthalten, dies bedeutet, dass man einen signifikanten Einfluss der Darmverstopfung durch den Tumor erkennen kann. Patienten, deren Darm verstopft ist, haben ein 0,751-fach höheres Risiko an Darmkrebs zu versterben als Patienten mit keiner Darmverstopfung durch den Tumor.

### 3.2.2 Modell mit Backward-Selektion (slstay 10%)

#### Model Information

Data Set	colon.colon	
Dependent Variable	time	Überlebenszeit
Censoring Variable	status	Status am Ende der Studie
Censoring Value(s)	0	
Ties Handling	BRESLOW	
Number of Observations Read		929
Number of Observations Used		906

#### Class Level Information

Class	Value	Design Variables		
rx	Lev	1	0	
	Lev+5FU	0	1	
	Obs	0	0	
sex	0	1		
	1	0		
differ	1	1	0	
	2	0	1	
	3	0	0	
extent	1	1	0	0
	2	0	1	0
	3	0	0	1
	4	0	0	0
surg	0	1		
	1	0		
node4	0	1		
	1	0		
adhere	0	1		
	1	0		
perfor	0	1		
	1	0		
obstruct	0	1		
	1	0		

#### Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
906	441	465	51.32

Step 0. The model contains the following effects:

rx sex age differ extent surg node4 adhere perfor obstruct

#### Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

#### Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	5695.191	5555.420
AIC	5695.191	5583.420
SBC	5695.191	5640.666

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	139.7710	14	<.0001
Score	153.1146	14	<.0001
Wald	142.9882	14	<.0001

Step 1. Effect perfor is removed. The model contains the following effects:

rx sex age differ extent surg node4 adhere obstruct

Bei einem slstay-Wert von 10% fällt der Darmdurchbruch aus dem Modell heraus, weil der p-Wert von dem Darmdurchbruch über diesen 10% liegt. Bei der Backward-Selektion startet man mit dem vollen Modell und der größte p-Wert wird iterativ entfernt, so lange bis alle p-Werte der Einflussgrößen unter dem slstay-Wert liegen.

Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	5695.191	5555.424
AIC	5695.191	5581.424
SBC	5695.191	5634.581

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	139.7667	13	<.0001
Score	153.1130	13	<.0001
Wald	143.0037	13	<.0001

Step 2. Effect sex is removed. The model contains the following effects:

rx age differ extent surg node4 adhere obstruct

Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	5695.191	5555.535
AIC	5695.191	5579.535
SBC	5695.191	5628.603

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	139.6558	12	<.0001
Score	153.0207	12	<.0001
Wald	142.9832	12	<.0001

Step 3. Effect adhere is removed. The model contains the following effects:

rx age differ extent surg node4 obstruct

#### Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

#### Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	5695.191	5556.930
AIC	5695.191	5578.930
SBC	5695.191	5623.909

#### Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	138.2609	11	<.0001
Score	151.8029	11	<.0001
Wald	142.0902	11	<.0001

NOTE: No (additional) effects met the 0.1 level for removal from the model.

#### Type 3 Tests

Effect	DF	Wald Chi-Square	Pr > ChiSq
rx	2	11.1199	0.0038
age	1	3.8962	0.0484
differ	2	9.5979	0.0082
extent	3	14.4968	0.0023
surg	1	5.5684	0.0183
node4	1	87.1904	<.0001
obstruct	1	5.8811	0.0153

Test  $\beta=0$ :  
 Behandlungsart: 0.0038 > 0.05  
 → signifikant  
 Alter 0.0484 < 0.05  
 → signifikant  
 Differenzierbarkeit: 0.0082 < 0.05  
 → signifikant  
 Tumorausdehnung: 0.0023 < 0.05  
 → signifikant  
 Behandlungszeitraum: 0.0183 < 0.05  
 → signifikant  
 Lymphknotenbefall: <.0001 < 0.05  
 → signifikant  
 Darmverstopfung: 0.0153 < 0.05  
 → signifikant

Bei einem slstay-Wert von 10% fallen die Behandlungsart, das Alter, die Differenzierbarkeit der Tumorzellen, die Tumorausdehnung, der Behandlungszeitraum, der Lymphknotenbefall und die Darmverstopfung nicht aus dem Modell heraus, weil deren p-Werte jeweils unter 10% liegen. Aber auch alle unter 5% liegen.

Es fehlen bei diesem slstay-Wert von 10% folgende Variablen: der Darmdurchbruch, das Geschlecht und der Befall der anliegenden Organe durch den Tumor.

### Analysis of Maximum Likelihood Estimates

Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
rx	Lev	1	-0.01564	0.11324	0.0191	0.8901
rx	Lev+5FU	1	-0.36453	0.12031	9.1808	0.0024
age		1	0.00809	0.00410	3.8962	0.0484
differ	1	1	-0.30980	0.19112	2.6275	0.1050
differ	2	1	-0.37796	0.12201	9.5959	0.0020
extent	1	1	-1.18153	0.54148	4.7613	0.0291
extent	2	1	-0.90163	0.26456	11.6149	0.0007
extent	3	1	-0.40973	0.20809	3.8771	0.0490
surg	0	1	-0.24590	0.10420	5.5684	0.0183
node4	0	1	-0.93105	0.09971	87.1904	<.0001
obstruct	0	1	-0.28571	0.11782	5.8811	0.0153

### Analysis of Maximum Likelihood Estimates

Parameter		Hazard Ratio	Label
rx	Lev	0.984	rx Lev
rx	Lev+5FU	0.695	rx Lev+5FU
age		1.008	age
differ	1	0.734	differ 1
differ	2	0.685	differ 2
extent	1	0.307	extent 1
extent	2	0.406	extent 2
extent	3	0.664	extent 3
surg	0	0.782	surg 0
node4	0	0.394	node4 0
obstruct	0	0.751	obstruct 0

### Summary of Backward Elimination

Step	Effect Removed	DF	Number In	Wald Chi-Square	Pr > ChiSq	Effect Label
1	perfor	1	9	0.0043	0.9477	perfor
2	sex	1	8	0.1109	0.7391	sex
3	adhere	1	7	1.4394	0.2302	adhere

### Hazard Ratios for rx

Description	Point Estimate	95% Wald Confidence Limits	
rx Lev vs Lev+5FU	1.417	1.116	1.801
rx Lev vs Obs	0.984	0.789	1.229
rx Lev+5FU vs Obs	0.695	0.549	0.879

### Hazard Ratios for age

Description	Point Estimate	95% Wald Confidence Limits	
age Unit=1	1.008	1.000	1.016



#### Hazard Ratios for differ

Description	Point Estimate	95% Wald Confidence Limits	
differ 1 vs 2	1.071	0.773	1.482
differ 1 vs 3	0.734	0.504	1.067
differ 2 vs 3	0.685	0.540	0.870

#### Hazard Ratios for extent

Description	Point Estimate	95% Wald Confidence Limits	
extent 1 vs 2	0.756	0.268	2.133
extent 1 vs 3	0.462	0.172	1.244
extent 1 vs 4	0.307	0.106	0.877
extent 2 vs 3	0.611	0.432	0.866
extent 2 vs 4	0.406	0.242	0.682
extent 3 vs 4	0.664	0.442	0.998

#### Hazard Ratios for surg

Description	Point Estimate	95% Wald Confidence Limits	
surg 0 vs 1age	0.782	0.638	0.959

#### Hazard Ratios for node4

Description	Point Estimate	95% Wald Confidence Limits	
node4 0 vs 1	0.394	0.324	0.479

#### Hazard Ratios for obstruct

Description	Point Estimate	95% Wald Confidence Limits	
obstruct 0 vs 1age	0.751	0.597	0.947

### 3.2.3 Endmodell ohne Variable aus 3.2.2

Ein Problem könnte hier sein, dass sich eventuell eine andere Fallzahl ergibt durch Missings in den Variablen.

#### Model Information

Data Set	colon.colon	
Dependent Variable	time	Überlebenszeit in Tage
Censoring Variable	status	Status am Ende der Studie
Censoring Value(s)	0	
Ties Handling	BRESLOW	

Number of Observations Read	929
Number of Observations Used	906

Hier ist die gleiche Fallzahl vorhanden wie in 3.2.2. Daraus folgt, dass man das gleiche Modell wie in 3.2.2 verwenden darf.

#### Type 3 Tests

Effect	DF	Wald Chi-Square	Pr > ChiSq
rx	2	11.1199	0.0038
age	1	3.8962	0.0484
differ	2	9.5979	0.0082
extent	3	14.4986	0.0023
surg	1	5.5684	0.0183
node4	1	87.1904	<.0001
obstruct	1	5.8811	0.0153

Test  $\beta=0$ :

Einfluss Behandlungsart auf  
Überlebenszeit:  
 $0.0038 < 0.05$   
→ signifikant

Der Test fällt signifikant aus ( $p=0.0038$ ). Es kann somit ein Einfluss der Behandlungsart und Tod festgestellt werden. Daher bleibt die Behandlungsart in dem Modell.

#### Hazard Ratios for rx

Description	Point Estimate	95% Wald Confidence Limits	
rx Lev vs Lev+5FU	1.417	1.116	1.801
rx Lev vs Obs	0.984	0.789	1.229
rx Lev+5FU vs Obs	0.695	0.549	0.879

Die „1“ ist in dem Konfidenzintervall nicht enthalten, dies bedeutet, dass man einen signifikanten Einfluss der Behandlungsart erkennen kann. Patienten, die Lev+5FU erhalten haben, haben ein 1,4-fach höhere Überlebenschance an Darmkrebs nicht zu versterben als Patienten nur mit der Behandlung Levamisol und nur einer Beobachtung.

## 4 Diskussion

### Krankheitssituation:

In dieser Studie sind Daten von 929 Patienten gegeben, die an Dickdarmkrebs erkrankt sind. Davon verstarben 452 Patienten und 477 waren zum Ende der Studie noch am Leben. Die mediane Überlebenszeit liegt bei 6.9 Jahren. Die 5 –Jahresüberlebenswahrscheinlichkeit der Patienten liegt bei 56%.

### Einflussgröße Therapie:

In dieser Studie gibt es drei verschiedene Behandlungsarten. Diese bestehen aus der Gabe von Levamisol, der Behandlung zuerst mit Levamisol und später dann mit einer Chemotherapie mit 5FU oder der Patient wurde nur beobachtet. Die Patienten wurden wie folgt eingeteilt. Es bekamen 310 Patienten Levamisol, 304 Patienten Levamisol und 5FU und 315 Patienten wurden nur beobachtet. Aus der Häufigkeitstabelle mit den Behandlungsarten und dem Tod ist schon ersichtlich, dass es bedeutend weniger Todesfälle gibt mit der Behandlung von Levamisol und 5FU (27.21%) gegenüber Levamisol (35.62%) und Beobachtung (37.17%). Wenn jetzt noch die Überlebenskurven von Kaplan-Meier mit den einzelnen Behandlungsarten verglichen werden, so stellt man fest, dass die 5 Jahresüberlebenswahrscheinlichkeit von Levamisol und 5FU bei 63.4%, Levamisol bei 53.21% liegt und die Beobachtung bei 52.57% liegt. Auch beim Log-Rank Test ist der p-Wert bei 0.0029. Dies bedeutet, dass wenn man ein Signifikanzniveau von 5% festlegt der p-Wert ( $0.0029 < 0.05$ ) ist,

dann ist der Test signifikant. Dies bedeutet, dass die Therapieart einen Einfluss hat auf die Überlebenszeit des Patienten. Bei der Cox-Regression dem Vollen Modell stellt man anhand der Überschreitungswahrscheinlichkeit fest, dass der p-Wert ( $0.0043 < 0.05$ ) ist und daher der Test signifikant bleibt, auch nach Adjustierung weiterer Einflußgrößen. Dies bedeutet, dass die Behandlungsart einen Einfluss hat auf das Überleben des Patienten. Schaut man sich nun das Ods-Ratio und das 95% Konfidenz-Intervall der einzelnen Behand-lungsarten an, so stellt man fest, dass es eine 1.4 fach höheres Risiko besteht an Darmkrebs zu sterben wenn nur die Behandlung mit Levamisol durchgeführt wird gegenüber einer Behandlung mit Levamisol und anschließend 5FU. Das Konfidenzintervall geht von (1.109 bis 1.791). Da in diesem Intervall die 1 nicht enthalten ist hat die Behandlungsart einen Einfluss. Das gleiche stellt man bei dem Vergleich zwischen der Behandlung mit Levamisol und 5FU gegenüber der Beobachtung. Das Ods-Ratio beträgt 0.695 also eine Verbesserung der Überlebenschance an Darmkrebs zu sterben als das man den Patienten nur beobachtet. Das 95% Konfidenzintervall geht von (0.549 bis 0.879). Da in diesem Intervall die 1 nicht enthalten ist hat die Behandlungsart einen Einfluss. Bei der Behandlungsart Levamisol gegenüber der Beobachtung des Patienten stellt man keinen Einfluss der Behandlungsart fest, da die 1 in dem 95% Konfidenzintervall (0.789 bis 1.229) enthalten ist. Interessant wäre es noch, wenn es eine weitere Gruppe an Patienten gäbe, die nur mit 5FU behandelt worden wären.

#### Weitere Einflussgrößen:

Einige weitere wichtige Einflussgrößen, bei einem Signifikanzniveau von 5%, neben der Behandlungsart auf die Überlebenszeit ist das Alter der Patienten ( $p = 0.0484$ ), die Differenzierbarkeit der Tumorzellen ( $p = 0.0082$ ), Ausdehnung des Tumors ( $p = 0.0023$ ), die Behandlungsdauer der Patienten ( $p = 0.0183$ ), ob eine Darmverstopfung durch den Tumor vorliegt ( $p = 0.0153$ ) und den Lymphknotenbefall ( $p = <.0001$ ). Weniger Einfluss auf die Überlebenszeit haben: der Befall nebenliegender Organe des Darms ( $p = 0.9477$ ), das Geschlecht ( $p = 0.7406$ ) und der Darmdurchbruch durch den Tumor ( $p = 0.9477$ ).

Berücksichtigt man diese Einflussfaktoren im Cox Modell, dann bleiben die Effekte der Therapie allerdings stabil und hängen wenig von diesen weiteren Einflussfaktoren ab.

# SAS-Code

```
*****
*
*   Studiengang: Mathematische Biometrie BSc
*   Fach: Consulting Class
*   Fachsemester: 5
*
*   Beispiel: Colon Karzinom
*   Literatur: R-Datensatz aus der Library boot
*   Analysestrategie: Survival Analyse
*   Inhalt:
*       -Daten einlesen
*       -Deskriptive Statistik
*       -Kaplan-Meier Überlebenskurven + LogRank Test
*       -Proportional Hazard Regression Modell
*           -Vorraussetzungen prüfen
*           -Modellierung
*           -Variablenselektion
*
*   Autor: Alexander Monte
*   Datum: 31. August 2012
*****;
```

```
LIBNAME library 'T:\gemeinsam\Praktikanten\Alexander\colon';
LIBNAME colon 'T:\gemeinsam\Praktikanten\Alexander\colon';
```

```
*****
*
*   Formate für die diskreten Merkmale anlegen
*
*****;
```

```
PROC FORMAT LIBRARY=colon;
    VALUE geschl    0='w'
                   1='m';

    VALUE erkenn    1='gut erkennbar'
                   2='mäßig erkennbar'
                   3='wenig erkennbar';

    VALUE ausdehn    1='Darmschleimhaut'
                   2='Darmmuskulatur'
                   3='Äußere Darmwand'
                   4='Darmdurchbruch';

    VALUE zeit      0='kurzer Zeitraum'
                   1='langer Zeitraum';

    VALUE ereig     1='Rezidiv'
                   2='Tod';

    VALUE behind    0='Nein'
                   1='Ja';

    VALUE durchbr    0='Nein'
                   1='Ja';
```

```

VALUE befall    0='Nein'
                1='Ja';

VALUE zensiert  0='Nein'
                1='Ja';

VALUE knoten    0='Nein'
                1='Ja';

VALUE alter     1='< 65 Jahre'
                2='> 65 Jahre';

VALUE $therap   'Obs'      ='Beobachtung'
                'Lev'      ='Levamisol'
                'Lev+5-FU'='Levamisol und Chemotherapie';

RUN;

*****
*
*   Einlesen der Daten aus xls in SAS nur einmal durchführen
*
*****;

PROC IMPORT
    FILE='T:\gemeinsam\Praktikanten\Alexander\colon\daten_colon.xls'
    OUT=colon.colon
    DBMS=EXCEL2000 REPLACE;
    SHEET="colon";
    GETNAMES=YES;

RUN;

*alphabetische Auflistung der Variablen auch nur einmal durchführen;
PROC CONTENTS short;
RUN;

*****
*
*   Variablen:
*
*   id              Patientenidentifikationsnummer
*   study            Patient nimmt an der Studie Teil 1->ja
*   rx               Behandlungsart Obs->Beobachtet Lev->Gabe Levamisol
*                   Lev+5Fu->Gabe Levamisol und Chemotherapie
*   sex              Geschlecht 0->weiblich 1->männlich
*   age              Alter des Patienten in Jahren ab Tumordiagnose
*   obstruct         Darmbehinderung durch Tumor 0->Nein 1->Ja
*   perfor           Darmwanddurchbruch 0->Nein 1->ja
*   adhere           Befall der umliegenden Organe 0->Nein 1->Ja
*   nodes            Anzahl der befallenen Knoten Na->keine Angabe
*   time             Anzahl der Tage bis zu einem Ereigniss
*   status           Fand eine Zensierung des Patienten statt 0->Nein 1->Ja
*   differ           Differenzierbarkeit der Tumorzellen 1->gute
*                   Differenzierbarkeit 2->mäßige Differenzierbarkeit
*                   3->geringe Differenzierbarkeit NA->keine Angabe
*   extent           Ausdehnung des Tumors 1->Darmschleimhaut
*                   2->Darmmuskulatur 3->Äußere Darmwand
*                   4->Darmdurchbruch
*   surg             Behandlungszeitraum bis zum Ereignis
*                   0->kurzer Zeitraum 1->langer Zeitraum

```

```

*      Node4              Mehr als 4 Lymphknoten befallen 0->Nein 1->Ja
*
*****;

DATA form_colon;
    SET colon.colon;

    IF nodes = . then nodes=.;

    IF differ =. then erkenn=.;

    IF age <= 65 then kalter=1;
    ELSE kalter =2 ;

    LABEL  age='Alter in Jahren' time='Anzahl in Tagen' etype='Ereignis'
           alter='Alter klassiert' sex='Geschlecht';
    FORMAT sex geschl. differ erkenn. extent ausdehn. surg zeit. etype
           ereig. obstruct behind. perfor durchbr. adher befall. status
           zensiert. node4 knoten. age alter. rx $therap.;

RUN;

PROC FREQ DATA=colon.colon;          *Häufigkeiten aufgelistet für die
Zensierung;
TABLES status;
RUN;

PROC FREQ DATA=colon.colon;          *Häufigkeiten für die einzelnen
                                     Behandlungenarten;
TABLES rx;
RUN;

PROC FREQ DATA=colon.colon;          *Häufigkeiten aufgelistet für Geschlecht;
TABLES sex;
RUN;

PROC FREQ DATA=Form_colon;           *Häufigkeitstabelle für Alter klassiert;
TABLES kalter;
RUN;

PROC FREQ DATA=colon.colon;          *Häufigkeiten der befallenen Lymphknoten;
TABLES nodes;
RUN;

PROC FREQ DATA=colon.colon;          *Häufigkeit der differenzierbarkeit der
                                     Tumorzelle;
TABLES differ;
RUN;

PROC FREQ DATA=colon.colon;          *Häufigkeit der Ausdehnung des Tumors;
TABLES extent;
RUN;

PROC FREQ DATA=colon.colon;          *Häufigkeit für den Behandlungszeitraum;
TABLES surg;
RUN;

PROC FREQ DATA=colon.colon;          *Häufigkeit befall der Lymphknoten>4;
TABLES node4;
RUN;

```

```

PROC FREQ DATA=colon.colon;          *Häufigkeit befall der näheren Organe;
TABLES adhere;
RUN;

PROC FREQ DATA=colon.colon;          *Häufigkeit der Darmdurchbrüche;
TABLES perfor;
RUN;

PROC FREQ DATA=colon.colon;          *Häufigkeit der Darmverstopfung;
TABLES obstruct;
RUN;

*****
*
*   Statistische Auswertung der Variablen
*
*****;

*****
*
*       Deskriptive Statistik
*       Gesamtzahl, Anzahl der Missings, Mittelwert,
*       Standardabweichung, Median, unteres Quartil,
*       oberes Quartil, Minimum, Maximum von Alter
*
*****;
PROC SORT DATA=colon.colon;
BY rx status;
RUN;
PROC MEANS DATA=colon.colon MAXDEC=2 PCTLDEF=4 N NMISS MEAN STD MEDIAN Q1
Q3 MIN MAX;
    VAR age nodes;
    TITLE 'Einfache Deskriptive Statistik zur Überlebenszeit, Alter und
    befallener Lymphknoten';
RUN;

*****
*
*       Histogramm: Lymphknoten, Alter
*
*****;

ODS GRAPHICS ON /Border=off;

PROC UNIVARIATE DATA=colon.colon NOPRINT PCTLDEF=4;
    VAR status;
    HISTOGRAM status /NORMAL;
    TITLE 'Histogramm des Status';
RUN;

PROC UNIVARIATE DATA=colon.colon NOPRINT PCTLDEF=4;
    VAR nodes;
    HISTOGRAM nodes /NORMAL;
    TITLE 'Histogramm der befallenen Lymphknoten';
RUN;
PROC UNIVARIATE DATA=colon.colon NOPRINT PCTLDEF=4;
    VAR age;
    HISTOGRAM age /NORMAL;
    TITLE 'Histogramm des Alter';

```

```

RUN;

PROC CAPABILITY DATA=colon.colon NOPRINT PCTLDEF=4;
    VAR age;
    HISTOGRAM age /NORMAL;
    TITLE 'Histogramm des Alters';
RUN;

PROC CAPABILITY DATA=colon.colon NOPRINT PCTLDEF=4;
    VAR status;
    HISTOGRAM status /NORMAL;
    TITLE 'Histogramm des Status';
RUN;
ODS GRAPHICS OFF;

*****
*
*   Auswertung der diskreten Variablen
*   Status, Geschlecht, Ausdehnung, Behandlung und Lymphknoten
*
*****;

PROC FREQ DATA=colon.colon;
    TABLES status sex extent rx node4 /NOCUM;
    TITLE 'Häufigkeitstabelle zu Status, Geschlecht, Behandlung und
        Lymphknoten';
RUN;

*****
*
*   Auswertung nach Status und der Behandlungsart
*   stratifiziert
*
*****;

ODS GRAPHICS ON /Border=off;

PROC FREQ DATA=colon.colon;
    TABLES sex*rx/chisq;
    TITLE 'Häufigkeitstabelle Geschlecht und Behandlungsart';
RUN;

PROC FREQ DATA=colon.colon;
    TABLES status*rx/chisq;
    TITLE 'Häufigkeitstabelle Status und Behandlungsart';
RUN;

PROC FREQ DATA=Form_colon;
    TABLES kalter*rx/chisq;
    TITLE 'Häufigkeitstabelle Alter und Behandlungsart';
RUN;

PROC FREQ DATA=colon.colon;
    TABLES obstruct*rx/chisq;
    TITLE 'Häufigkeitstabelle Verstopfung und Behandlungsart';
RUN;

PROC FREQ DATA=colon.colon;
    TABLES perfor*rx/chisq;
    TITLE 'Häufigkeitstabelle Darmdurchbruch und Behandlungsart';

```



```

RUN;

PROC FREQ DATA=colon.colon;
    TABLES adhere*rx/chisq;
    TITLE 'Häufigkeitstabelle befallene Organe und Behandlungsart';
RUN;

PROC FREQ DATA=colon.colon;
    TABLES node4*rx/chisq;
    TITLE 'Häufigkeitstabelle befallene Knoten und Behandlungsart';
RUN;

PROC FREQ DATA=colon.colon;
    TABLES differ*rx/chisq;
    TITLE 'Häufigkeitstabelle Differenzierung des Tumorgewebes und
        Behandlungsart';
RUN;

PROC FREQ DATA=colon.colon;
    TABLES extent*rx/chisq;
    TITLE 'Häufigkeitstabelle Tumorausdehnung und Behandlungsart';
RUN;

PROC FREQ DATA=colon.colon;
    TABLES surg*rx/chisq;
    TITLE 'Häufigkeitstabelle Zeitraum und Behandlungsart';
RUN;

PROC FREQ DATA=colon.colon;
    TABLES sex*status/chisq;
    TITLE 'Häufigkeitstabelle Geschlecht und Status Patient';
RUN;

PROC FREQ DATA=colon.colon;
    TABLES rx*status/chisq;
    TITLE 'Häufigkeitstabelle Behandlungsart und Satus Patient';
RUN;

PROC FREQ DATA=colon.colon;
    TABLES age*status/chisq;
    TITLE 'Häufigkeitstabelle Alter und Status Patient';
RUN;

PROC FREQ DATA=colon.colon;
    TABLES obstruct*status/chisq;
    TITLE 'Häufigkeitstabelle Verstopfung und Status Patient';
RUN;

PROC FREQ DATA=colon.colon;
    TABLES perfor*status/chisq;
    TITLE 'Häufigkeitstabelle Darmdurchbruch und Status Patient';
RUN;

PROC FREQ DATA=colon.colon;
    TABLES adhere*status/chisq;
    TITLE 'Häufigkeitstabelle befallener Organe und Status Patient';
RUN;

PROC FREQ DATA=colon.colon;
    TABLES node4*status/chisq;
    TITLE 'Häufigkeitstabelle befallene Knoten und Status Patient';
RUN;

```

```

PROC FREQ DATA=colon.colon;
    TABLES differ*status/chisq;
    TITLE 'Häufigkeitstabelle Differenzierung Tumorzellen und Status
          Patient';
RUN;

PROC FREQ DATA=colon.colon;
    TABLES extent*status/chisq;
    TITLE 'Häufigkeitstabelle Ausdehnung Tumor und Status Patient';
RUN;

PROC FREQ DATA=colon.colon;
    TABLES surg*status/chisq;
    TITLE 'Häufigkeitstabelle Behandlungszeitraum und Status Patient';
RUN;
ODS GRAPHICS OFF;

*****
*
*      Kaplan-Meier, Gesamtüberleben
*
*****;

TITLE;
ods rtf
file="T:\gemeinsam\Praktikanten\Alexander\colon\colon.RTF";
ODS GRAPHICS ON /Border=off;

PROC LIFETEST CONFBAND=EP ALPHAQT=0.05 data=colon.colon PLOT=(S (CL),LLS);
    TIME time * status(0);
RUN;
ODS GRAPHICS OFF;

*****
*
*      Kaplan-Meier, Gruppierung: Behandlungsart
*
*****;

ODS GRAPHICS ON /Border=off;

PROC LIFETEST CONFBAND=EP ALPHAQT=0.05 data=colon.colon PLOT=(S (CL),LLS);
    TIME time * status(0);
    STRATA rx;
RUN;
ODS GRAPHICS OFF;

*****
*
*      Kaplan-Meier, Gruppierung: Geschlecht
*
*****;

ODS GRAPHICS ON /Border=off;

PROC LIFETEST CONFBAND=EP ALPHAQT=0.05 data=colon.colon PLOT=(S (CL),LLS);
    TIME time * status(0);
    STRATA sex;
RUN;
ODS GRAPHICS OFF;

```

```

*****
*
*      Kaplan-Meier, Gruppierung: Alter
*
*****;

ODS GRAPHICS ON /Border=off;

PROC LIFETEST CONFBAND=EP ALPHAQT=0.05 data=Form_colon PLOT=(S (CL),LLS);
    TIME time * status(0);
    STRATA kalter;
RUN;

ODS GRAPHICS OFF;

*****
*
*      Kaplan-Meier, Gruppierung: Lymphknotenbefall
*
*****;

ODS GRAPHICS ON /Border=off;

PROC LIFETEST CONFBAND=EP ALPHAQT=0.05 data=colon.colon PLOT=(S (CL),LLS);
    TIME time * status(0);
    STRATA node4;
RUN;
ODS GRAPHICS OFF;

*****
*
*      Kaplan-Meier, Gruppierung: Ausbreitung des Tumors
*
*****;

ODS GRAPHICS ON /Border=off;

PROC LIFETEST CONFBAND=EP ALPHAQT=0.05 data=colon.colon PLOT=(S (CL),LLS);
    TIME time * status(0);
    STRATA extent;
RUN;
ODS GRAPHICS OFF;

*****
*
*      Kaplan-Meier, Gruppierung: Verstopfung des Darms durch den Tumor
*
*****;

ODS GRAPHICS ON /Border=off;

PROC LIFETEST CONFBAND=EP ALPHAQT=0.05 data=colon.colon PLOT=(S (CL),LLS);
    TIME time * status(0);
    STRATA obstruct;
RUN;
ODS GRAPHICS OFF;

```

```

*****
*
*      Kaplan-Meier, Gruppierung: Darmdurchbruch
*
*****;

ODS GRAPHICS ON /Border=off;

PROC LIFETEST CONFBAND=EP ALPHAQT=0.05 data=colon.colon PLOT=(S (CL),LLS);
    TIME time * status(0);
    STRATA perfor;
RUN;
ODS GRAPHICS OFF;

*****
*
*      Kaplan-Meier, Gruppierung: Befall der anliegenden Organe durch Tumor
*
*****;

ODS GRAPHICS ON /Border=off;

PROC LIFETEST CONFBAND=EP ALPHAQT=0.05 data=colon.colon PLOT=(S (CL),LLS);
    TIME time * status(0);
    STRATA adhere;
RUN;
ODS GRAPHICS OFF;

*****
*
*      Kaplan-Meier, Gruppierung: Zelldifferenzierung des Tumors
*
*****;

ODS GRAPHICS ON /Border=off;

PROC LIFETEST CONFBAND=EP ALPHAQT=0.05 data=colon.colon PLOT=(S (CL),LLS);
    TIME time * status(0);
    STRATA differ;
RUN;
ODS GRAPHICS OFF;

*****
*
*      Kaplan-Meier, Gruppierung: Behandlungszeitraum
*
*****;

ODS GRAPHICS ON /Border=off;

PROC LIFETEST CONFBAND=EP ALPHAQT=0.05 data=colon.colon PLOT=(S (CL),LLS);
    TIME time * status(0);
    STRATA surg;
RUN;
ODS GRAPHICS OFF;
ods rtf close;

```

```

*****
*
*      PHREG Cox-Regression
*      Behandlungsart und Status klassiert
*
*****;

ODS GRAPHICS ON /Border=off;

PROC PHREG data=colon.colon;
    CLASS rx sex obstruct perfor adhere differ extent surg node4;
    model time*status(0) = rx sex age obstruct perfor adhere differ
extent
    surg node4;
    HAZARDATIO rx;
    HAZARDATIO sex;
    HAZARDATIO age;
    HAZARDATIO obstruct;
    HAZARDATIO perfor;
    HAZARDATIO adhere;
    HAZARDATIO differ;
    HAZARDATIO extent;
    HAZARDATIO surg;
    HAZARDATIO node4;
RUN;

ODS GRAPHICS OFF;

*****
*
*      PHREG Cox-Regression
*      Behandlungsart und Status klassiert
*      Modell mit Backward-Selektion
*
*****;

ODS GRAPHICS ON /Border=off;

PROC PHREG data=colon.colon;
    CLASS rx sex obstruct perfor adhere differ extent surg node4;
    model time*status(0) = rx sex age obstruct perfor adhere differ
extent
    surg node4 / selection=bw slstay=0.1;
    HAZARDATIO rx;
    HAZARDATIO sex;
    HAZARDATIO age;
    HAZARDATIO obstruct;
    HAZARDATIO perfor;
    HAZARDATIO adhere;
    HAZARDATIO differ;
    HAZARDATIO extent;
    HAZARDATIO surg;
    HAZARDATIO node4;
RUN;
ODS GRAPHICS OFF;

```

```

*****
*
*   PHREG Cox-Regression
*   Behandlungsart und Status klassiert
*   Endmodell mit Wechselwirkung Behandlungsart und Tod
*
*****;
ODS GRAPHICS ON /Border=off;

PROC PHREG data=colon.colon;
  CLASS rx obstruct differ extent surg node4;
  model time*status(0) = rx age obstruct differ extent
    surg node4;
  HAZARDATIO rx;
  HAZARDATIO age;
  HAZARDATIO obstruct;
  HAZARDATIO differ;
  HAZARDATIO extent;
  HAZARDATIO surg;
  HAZARDATIO node4;
RUN;
ODS GRAPHICS OFF;

```