

3 Ergebnisse

3.1 Analyse der simulierten Daten

Um die verschiedenen Modelle zu untersuchen, wurden Daten simuliert. Das Vorgehen bei der Simulation ist im Methodenteil dieser Arbeit beschrieben. Von Abbildung 7 bis Abbildung 16 sind die Überlebenskurven für die einzelnen simulierten Variablen jeweils für Trainings- und Testdaten dargestellt. Die Variablen F2 und F5 wurden als ordinal skalierte und die Variablen F1, F3 und F4 als stetig skalierte Faktoren betrachtet. Bei den stetigen Faktoren wurde der Median in der Trainingsdatei als Grenzwert für Dichotomisierung zur Darstellung in Kaplan-Meier-Überlebenskurven verwendet.

3.1.1 Kaplan-Meier-Kurven und Log-Rank-Tests bei den simulierten Daten

Zu Darstellung der Variable F1 wurde als Grenzwert der Median ($m=1,253$) bei den Trainingsdaten ausgewählt. Diese Zahl wurde auch bei den Testdaten als Grenzwert verwendet. Insgesamt zeigt sich eine signifikante längere Überlebenszeit bei den 1999 Fällen mit einem Wert, der kleiner ist als der Median.

Mit Ausnahme von Variable F4 zeigte sich bereits auch bei den anderen Variablen ein signifikanter Einfluß auf die Prognose.

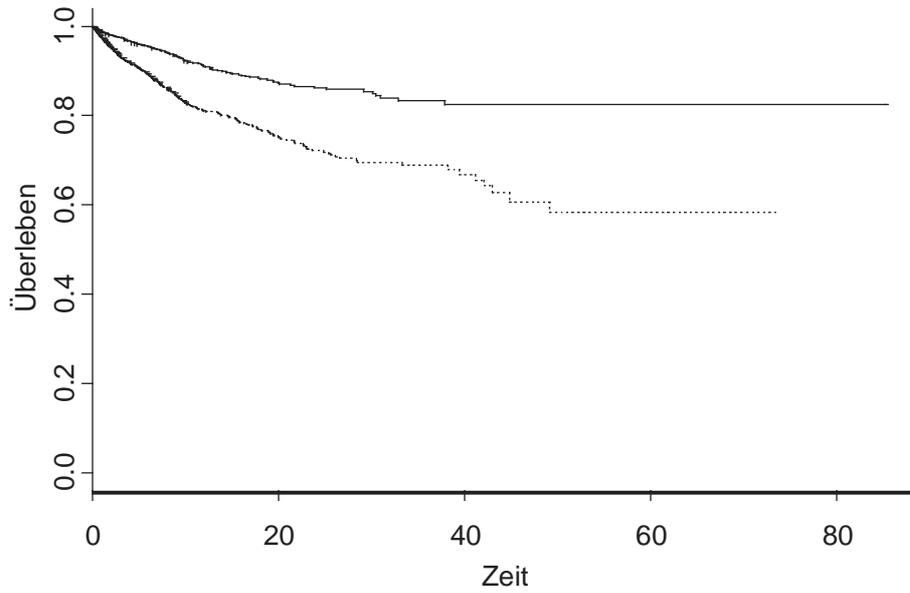


Abbildung 7: Überlebenskurven für die Variable F1 der simulierten Daten (Trainingsdaten)

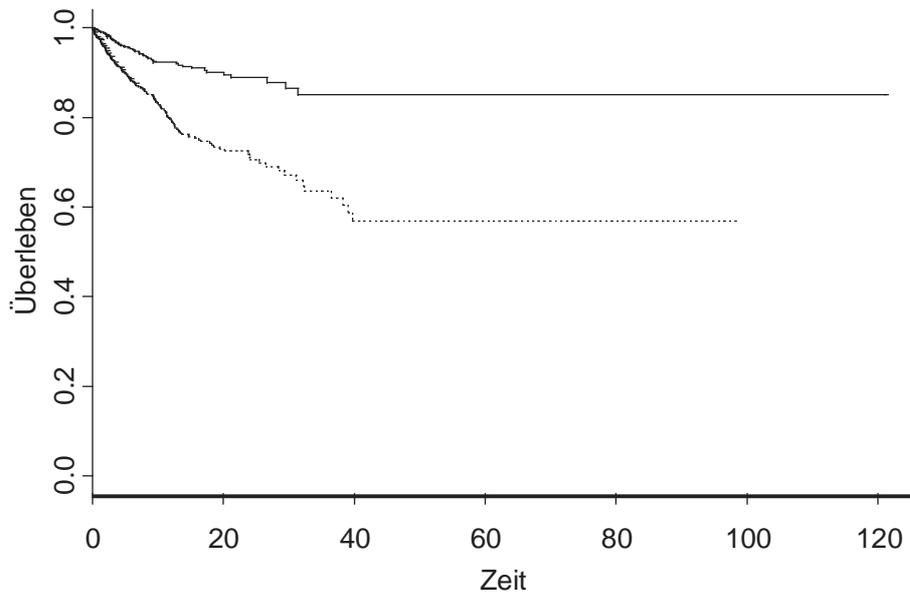


Abbildung 8: Überlebenskurven für die Variable F1 der simulierten Daten (Testdaten)

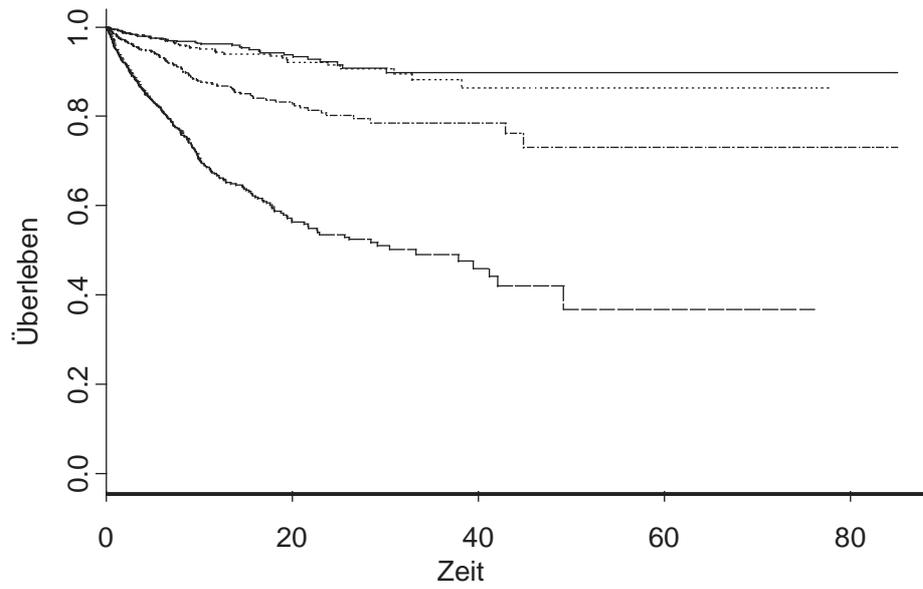


Abbildung 9: Überlebenskurven für die Variable F2 der simulierten Daten (Trainingsdaten)

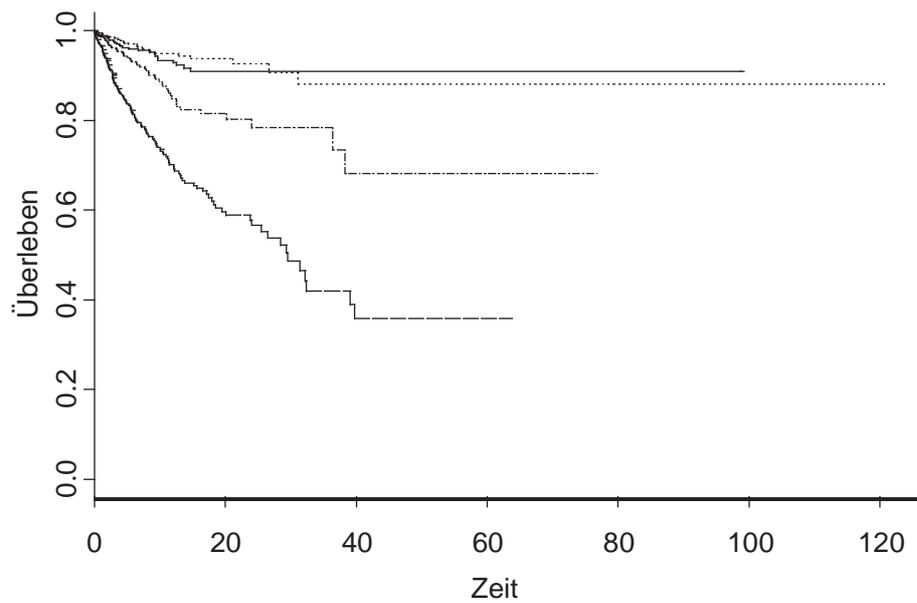


Abbildung 10: Überlebenskurven für die Variable F2 der simulierten Daten (Testdaten)

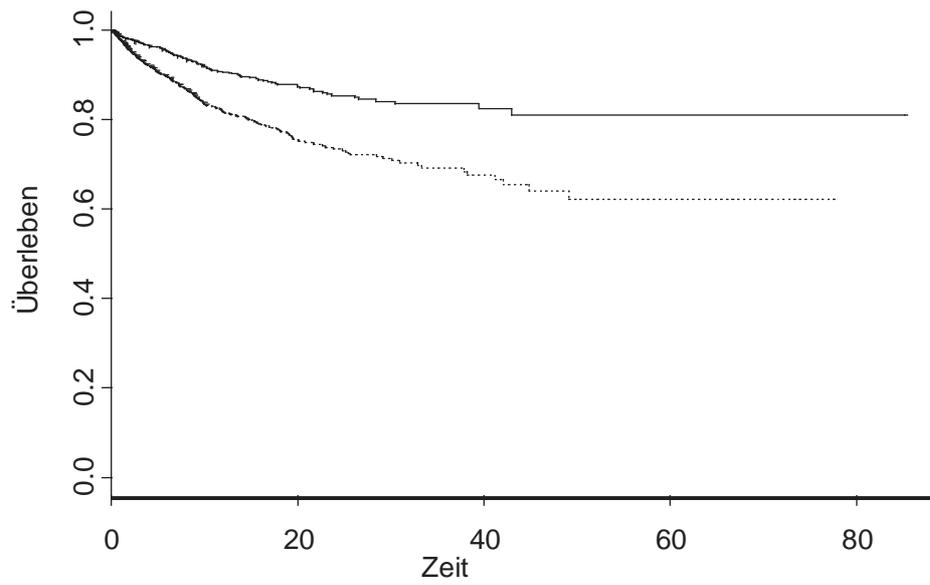


Abbildung 11: Überlebenskurven für die Variable F3 der simulierten Daten (Trainingsdaten)

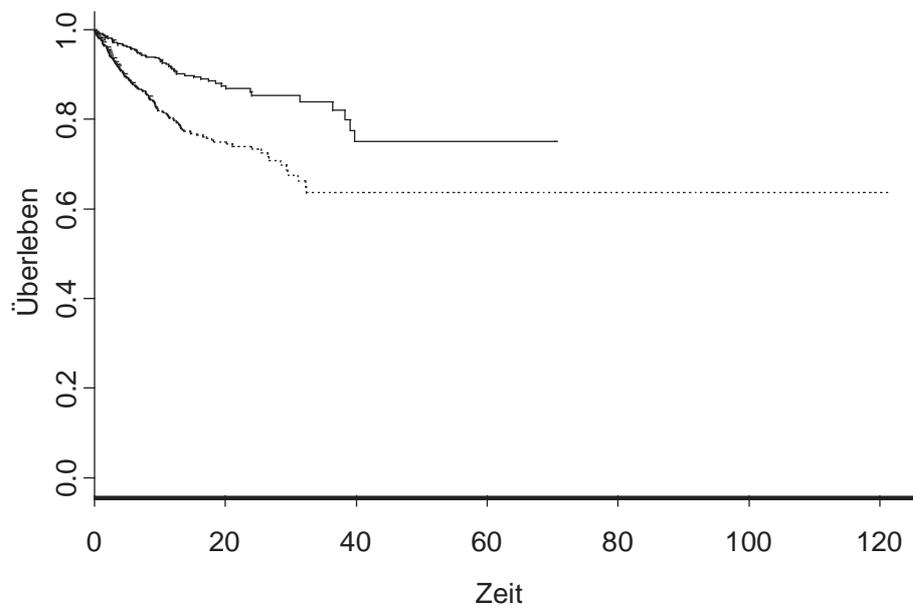


Abbildung 12: Überlebenskurven für die Variable F3 der simulierten Daten (Testdaten)

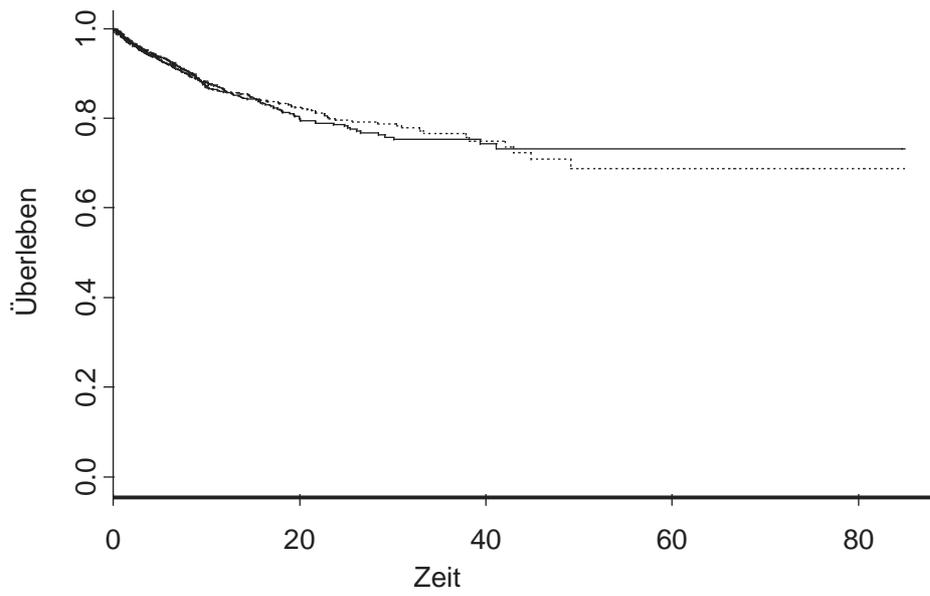


Abbildung 13: Überlebenskurven für die Variable F4 der simulierten Daten (Trainingsdaten)

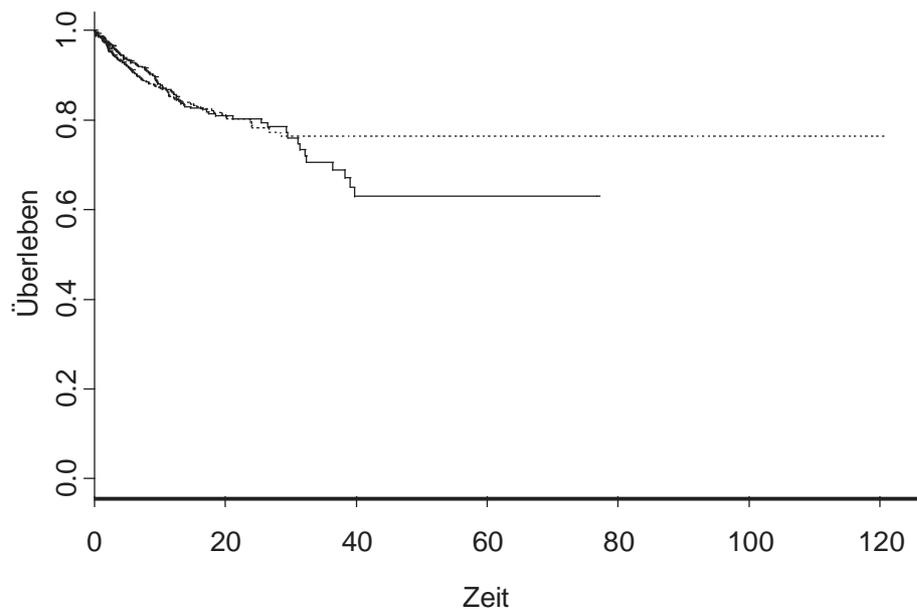


Abbildung 14: Überlebenskurven für die Variable F4 der simulierten Daten (Testdaten)

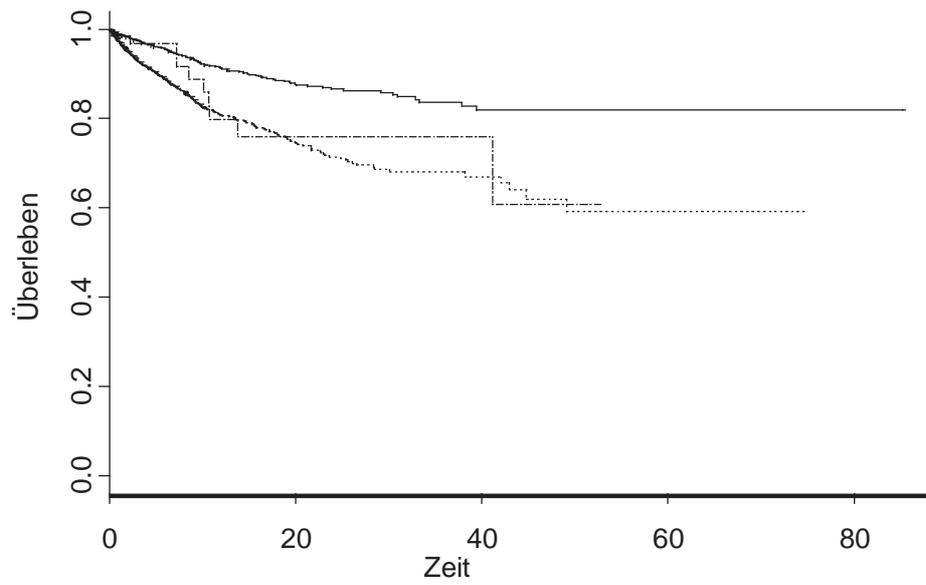


Abbildung 15: Überlebenskurven für die Variable F5 der simulierten Daten (Trainingsdaten)

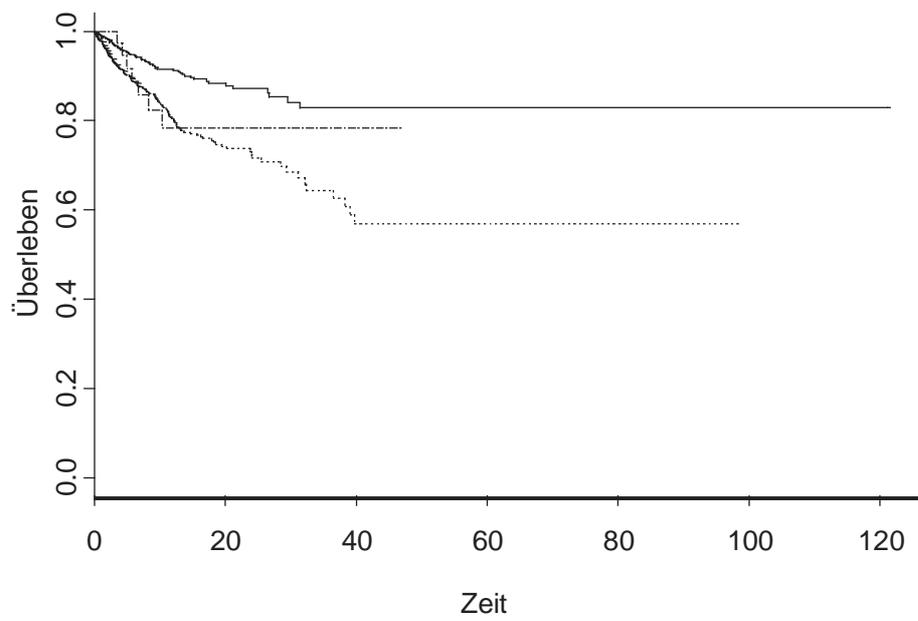


Abbildung 16: Überlebenskurven für die Variable F5 der simulierten Daten (Testdaten)

3.1.2 CART-Analyse

Bei der CART-Analyse ist die Festlegung der Baumgröße von zentraler Bedeutung. Dabei wird empfohlen, nach Möglichkeit vorhandene Hintergrundinformationen zu nutzen.

Ein nützliches Werkzeug ist die Betrachtung der relativen Fehler in Abhängigkeit von der Komplexität des Baumes. Die gestrichelte Linie liegt um einen Standardfehler oberhalb vom kleinsten Fehler. Die relativen Fehler bei dieser Art der Darstellung werden durch Kreuzvalidierungen bestimmt. Aus Abbildung 17 wird ersichtlich, daß eine Verwendung von mehr als 16 Knoten im Baum bei diesen Daten wieder zu einer Vergrößerung des relativen Fehlers führt. Um Artefakte auszuschließen, sollte der Baum eher aus wenigen Knoten bestehen.

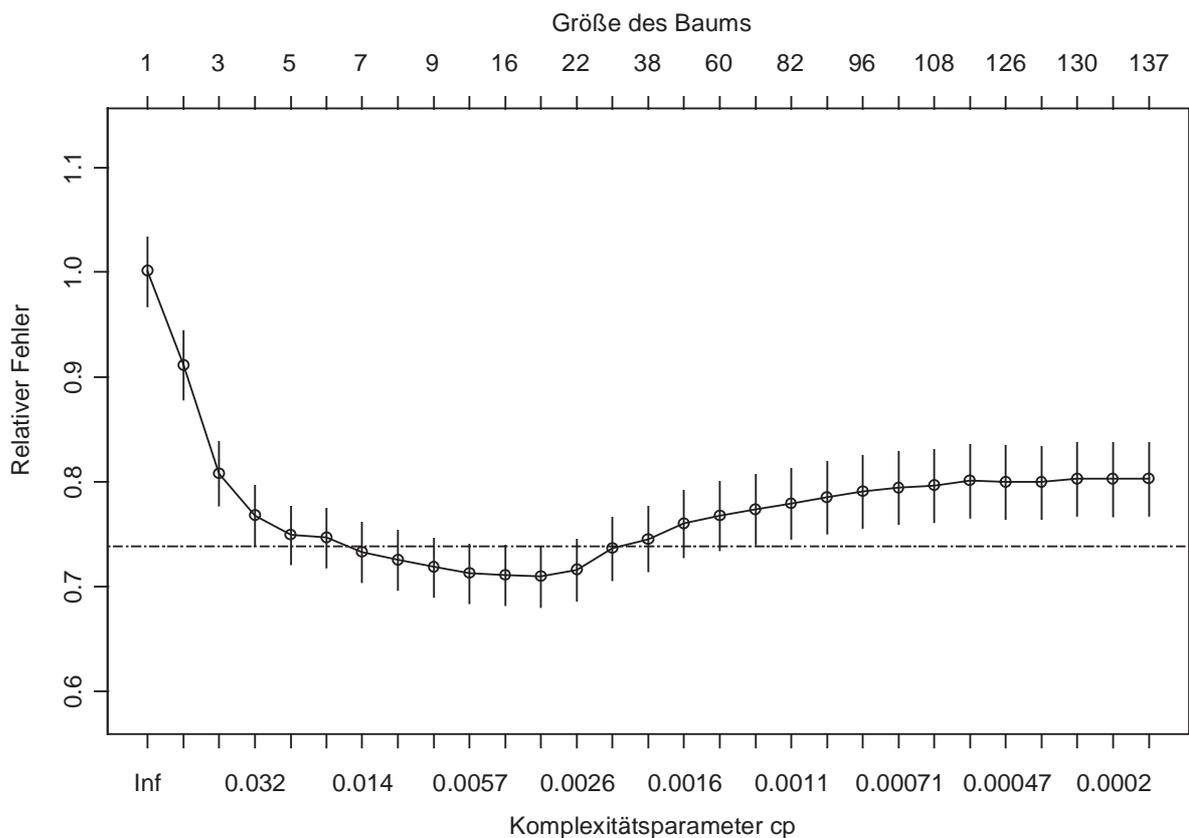


Abbildung 17: Komplexität des CART-Baumes und relativer Fehler bestimmt durch Kreuzvalidierungen

In einem interaktiven Verfahren kann das Aussehen des Baumes beeinflusst werden. Der hier dargestellte Baum wurde weitgehend automatisch erstellt. Lediglich bei der Festlegung der Anzahl der Knoten wurde das Aussehen des endgültigen Baumes manuell mitbestimmt.

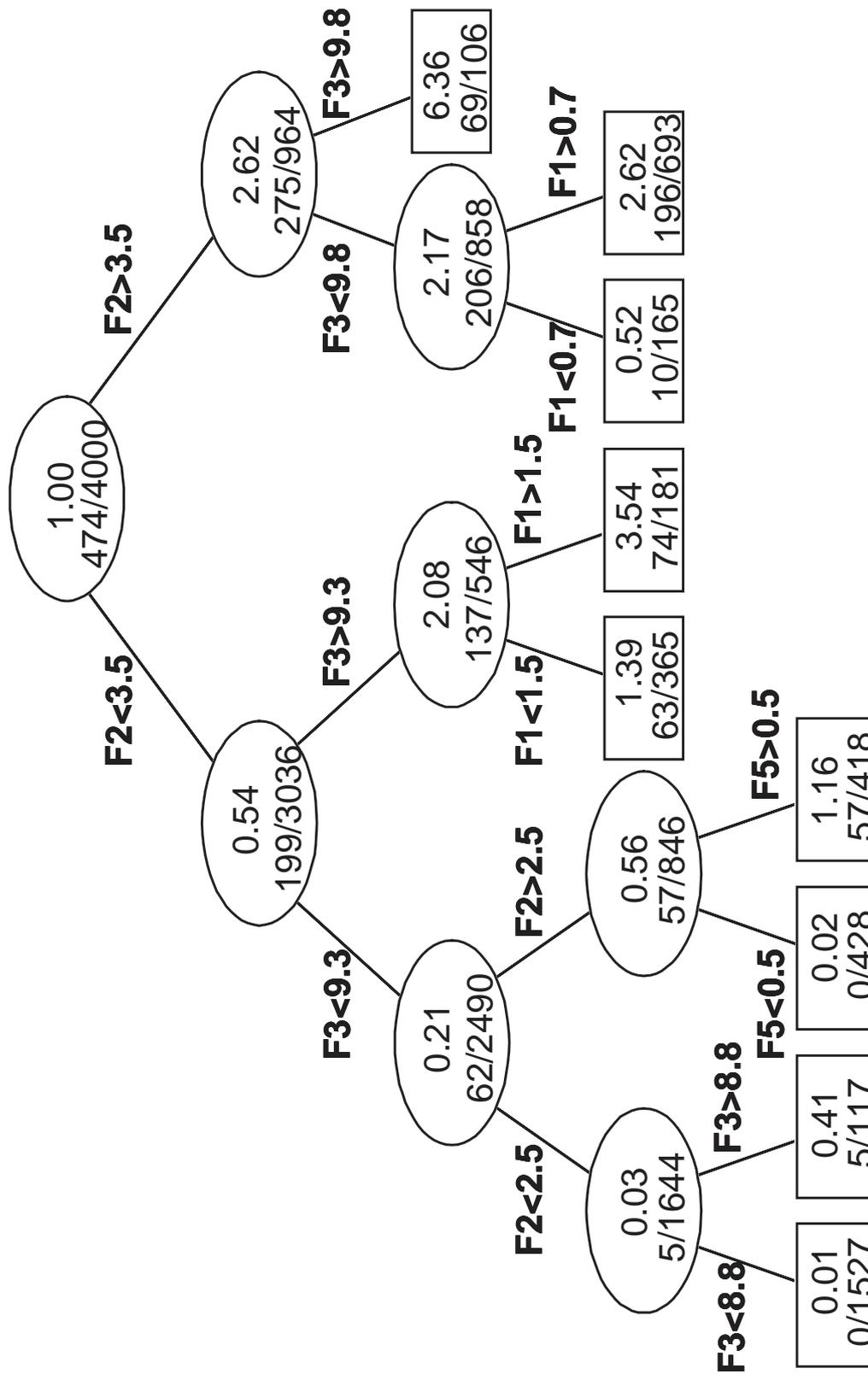


Abbildung 18: CART-Baum für die simulierten Daten unter Berücksichtigung von zensierten Beobachtungen

Als wichtigstes prognostisches Merkmal ergab sich bei der CART-Analyse der Faktor F2. In der nächsten Ebene des Baumes wurde in beiden Ästen F3 als nächste wichtige Variable in den Baum aufgenommen. Die Variable F4 wurde bei Einstellung des gewählten Komplexitätsparameters nie in den Baum aufgenommen.

Der linke Teil des Baumes beinhaltet die Fälle mit einer günstigen Prognose. Bei Betrachtung der drei linken Endpunkte wird ersichtlich, daß für eine große Zahl der untersuchten Fälle nur ein sehr geringes Risiko für das Eintreten des Zielereignisses besteht. Das Aussehen des Baumes läßt auf eine Interaktion zwischen den Variablen F2 und F3 schließen.

3.1.3 Cox-PH-Analyse

Bei Analyse der simulierten Daten wurden alle potentiellen Faktoren einzeln als lineare Variablen in das Modell hineingenommen. Interaktionen oder nichtlineare Anteile wurden im ersten Analyseschritt nicht berücksichtigt. Da die Zeitvariable bei der Simulation nicht zur Definition von Ereignissen verwendet wurde, muß die Proportionalitätsannahme erfüllt sein. In diesem speziellen Fall ist das Risiko für ein Ereignis über den gesamten Zeitraum konstant. Beim Testen auf Verletzung der Proportionalitätsannahme beträgt der p-Wert für F2 $p=0,997$ und für F3 $p=0,506$. Für die Variablen F2 und F3 ist die grafische Überprüfung der Proportionalitätsannahme in Abbildung 19 und Abbildung 20 dargestellt. Die Proportionalitätsannahme gilt als verletzt, wenn sich durch das Konfidenzband (± 2 Standardabweichungen, Punkte) keine horizontal verlaufende Gerade legen läßt. In diesen beiden Abbildungen ist kein Anhalt für eine Verletzung der Proportionalitätsannahme zu finden.

In Abbildung 21 sind die relativen Hazard-Raten für die Variablen F1, F2, F3 und F4 dargestellt. Bis auf F4 ist bei jedem der potentiellen Faktoren ein Einfluß auf das Zielereignis zu erkennen.

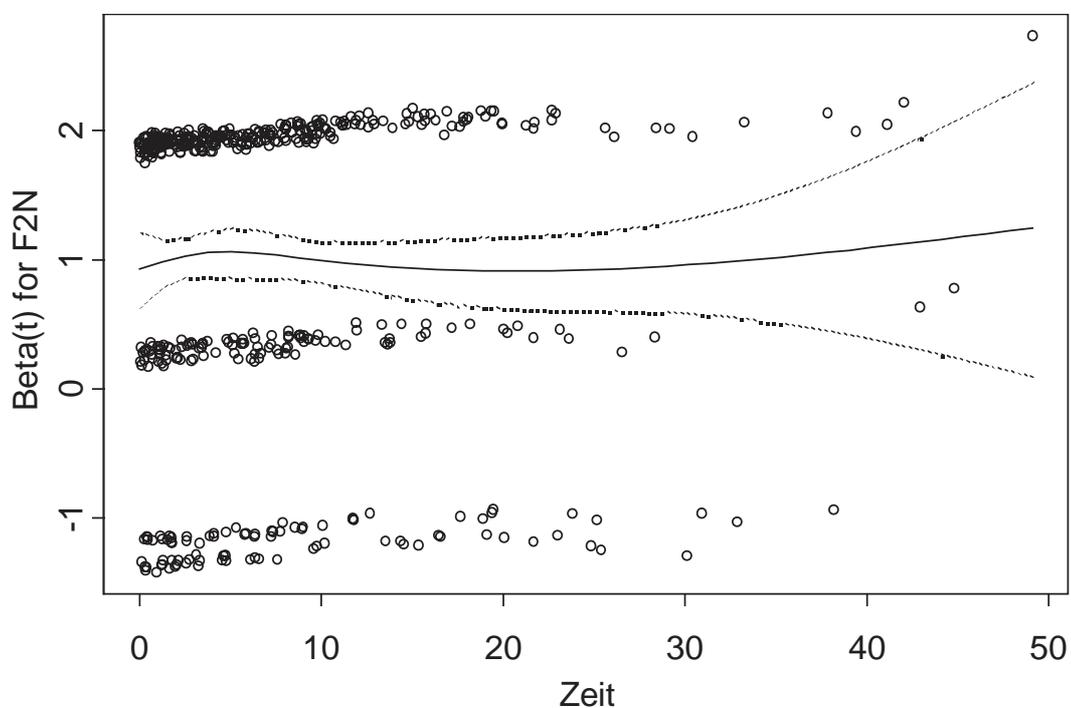


Abbildung 19: Überprüfung der Proportionalitätsannahme für die Variable F2.

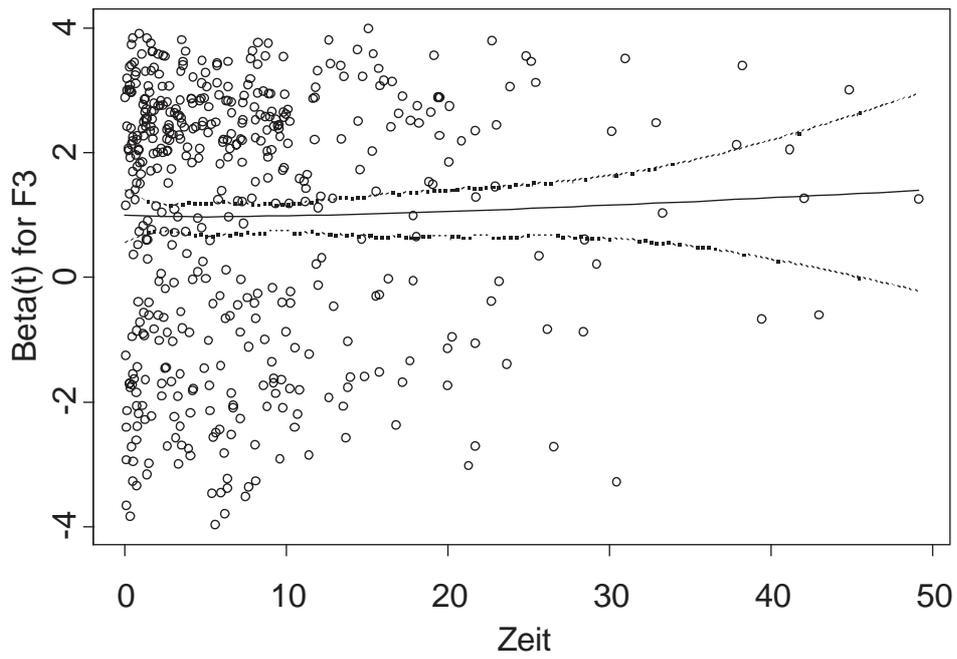


Abbildung 20: Überprüfung der Proportionalitätsannahme für die Variable F3

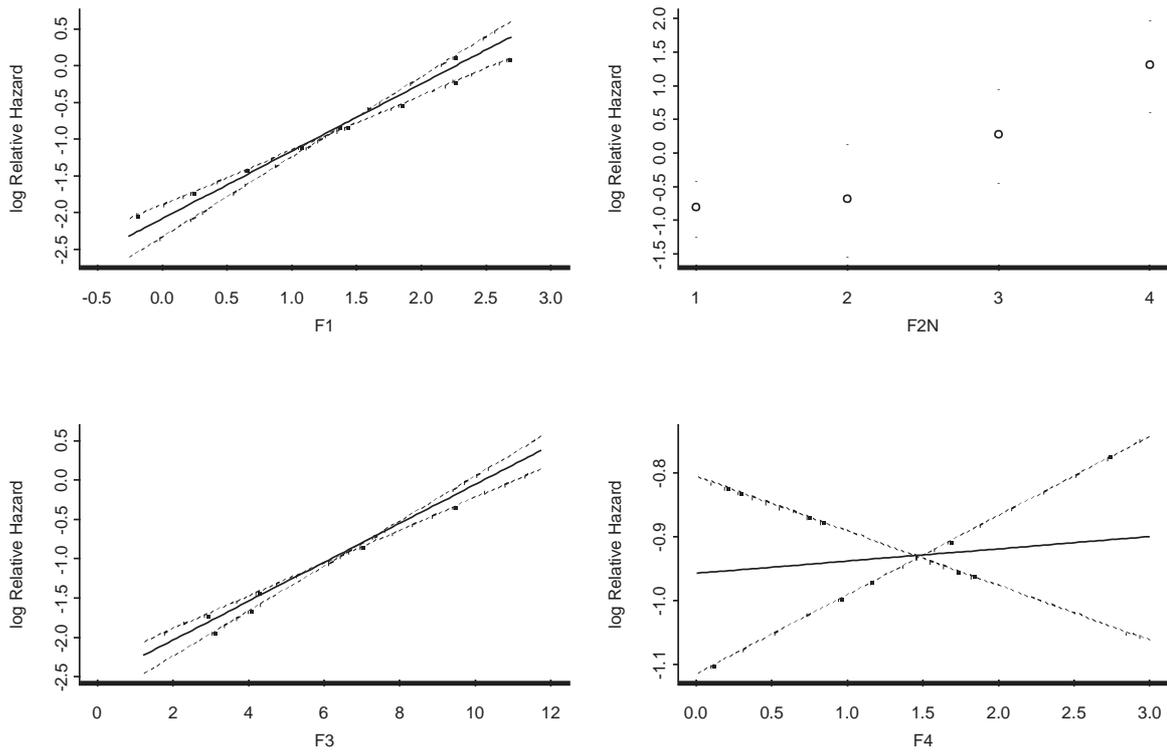


Abbildung 21: Log-Relative Hazards für die Variablen F1-F4, adjustiert für: F1=1.253, F2=2; F3=6.462, F4=1.468

3.1.4 Künstliche neuronale Netze

Die Simulation eines künstlichen neuronalen Netzes erfolgte mit einem vollständig verbundenen Modell mit sechs Neuronen in der Eingabeschicht ($i_1 - i_6$), einer Fehler-Einheit (b), zwei Neuronen in der versteckten Schicht (h_1 und h_2) und einem Neuron (o) in der Ausgabeschicht (Abbildung 22). Daraus ergaben sich 23 Gewichte, die in der Lernphase trainiert werden mußten. Das Modell konvergierte nach weniger als 100 Iterationen. Die reine Rechenzeit auf einem Pentium-II-Prozessor mit 266 MHz betrug weniger als eine Minute.

Die aus der Lernphase resultierenden Gewichte des Netzes sind in Tabelle 6 dargestellt.

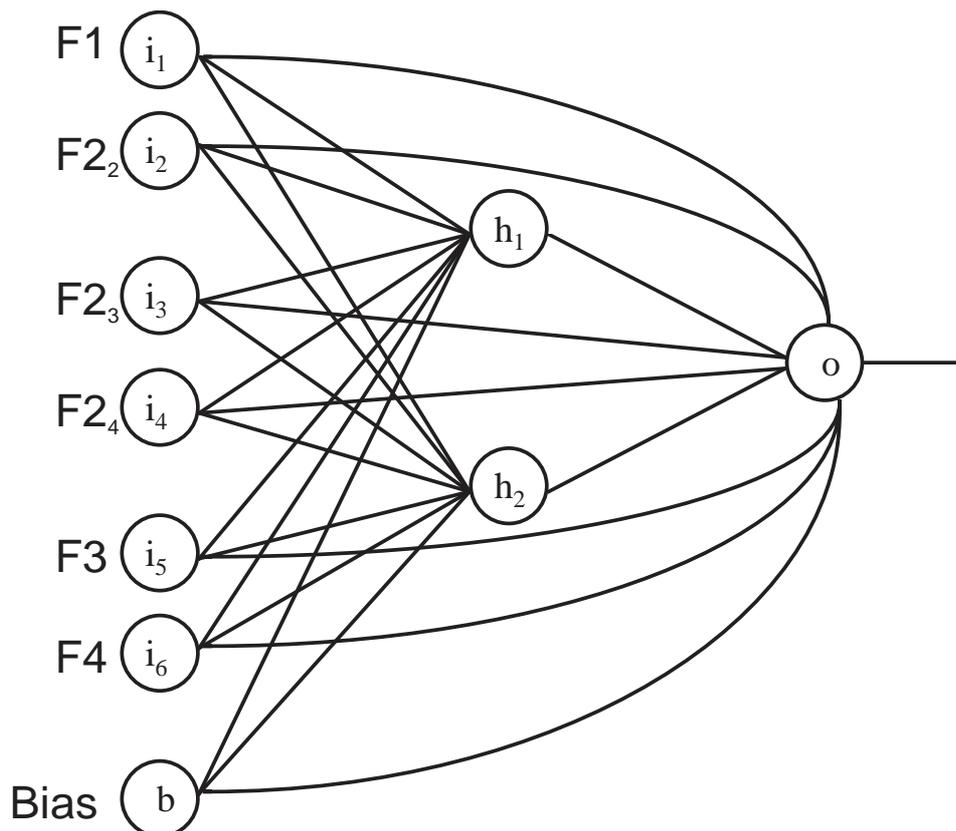


Abbildung 22: Struktur des verwendeten Netzwerks für die Analyse der simulierten Daten

$i1 \rightarrow h1$	$i2 \rightarrow h1$	$i3 \rightarrow h1$	$i4 \rightarrow h1$	$i5 \rightarrow h1$	$i6 \rightarrow h1$	$b \rightarrow h1$
-0.25	2.54	-1.45	0.78	-0.41	-0.03	-8.23
$i1 \rightarrow h2$	$i2 \rightarrow h2$	$i3 \rightarrow h2$	$i4 \rightarrow h2$	$i5 \rightarrow h2$	$i6 \rightarrow h2$	$b \rightarrow h2$
-1.05	-1.07	-4.05	2.63	0.91	0.51	3.38
$i1 \rightarrow o$	$i2 \rightarrow o$	$i3 \rightarrow o$	$i4 \rightarrow o$	$i5 \rightarrow o$	$i6 \rightarrow o$	$b \rightarrow o$
1.06	-0.02	0.53	0.20	0.99	0.12	0.0
$h1 \rightarrow o$	$h2 \rightarrow o$					
9.91	-6.85					

Tabelle 6: Gewichte des neuronalen Netzes

Wie bei künstlichen neuronalen Netzen üblich, fällt die Bewertung der Gewichte schwer. Zu erkennen ist ein starker Einfluß der Variablen F2 und F3 durch die betragsmäßig größten Gewichte. Dabei wirkt der Einfluß von F3 vor allem auf die Ausgabeinheit. Allerdings besteht bei der Bewertung der einzelnen Gewichte die Gefahr, Zusammenhänge zu übersehen. Deshalb soll hier auf die Gewichte nicht weiter eingegangen werden.

3.1.5 Vergleich der Ergebnisse bei den simulierten Daten

In Abbildung 23 bis Abbildung 26 sind die ROC-Kurven für vier Zeitpunkte (bis 3 Monate, bis 7,5 Monate, bis 15 Monate und die gesamte Beobachtungszeit) und jede Methode dargestellt. Dabei beziehen sich die Diagramme A, B, C und D auf die Trainingsdaten und die Diagramme E, F, G und H auf die Testdaten. Zusätzlich ist für jedes Diagramm die Wahrscheinlichkeit für das Eintreten eines Ereignisses zum jeweiligen Zeitpunkt $p(T)$ angegeben. Außerdem können der Flächeninhalt unter der ROC-Kurve c und der damit verbundene Wert Somer's D abgelesen werden. Diese Werte sind als Maß für die Qualität der Prognoseschätzungen zu werten. Auffällig bei der Prognoseschätzung mit einem Parameter (Variable F3) ist für alle Zeitpunkte, daß im Bereich mit niedriger Sensitivität die Spezifität bereits relativ hoch ist. (Abbildung 23)

Ein anderes Aussehen haben die ROC-Kurven, die auf einer Schätzung mit einem CART-Baum beruhen (Abbildung 24). Hier liegt Somer's D mit Werten oberhalb 0,7 erheblich höher. Allerdings zeigt der Vergleich zwischen Trainings- und Testdaten eine Überanpassung (overfitting). Die Werte für Somer's D bei den Testdaten liegen um bis zu 0,06 unter den Werten der Trainingsdaten.

Bei den Kurven für das Cox-PH-Modell ist die Anpassung an die Trainingsdaten etwas schlechter verlaufen als beim CART-Modell (Abbildung 25).

Die Prognosequalität des künstlichen neuronalen Netzes zeigt in diesem Vergleich von allen geprüften Modellen die höchsten Werte für Somer's D bei den Testdaten (Abbildung 26). Offenbar ist hier die Anpassung an die Struktur der Daten am besten gelungen.

Bei den hier verwendeten Daten hat der zeitliche Verlauf nur einen geringen Einfluß auf die Prognosequalität.

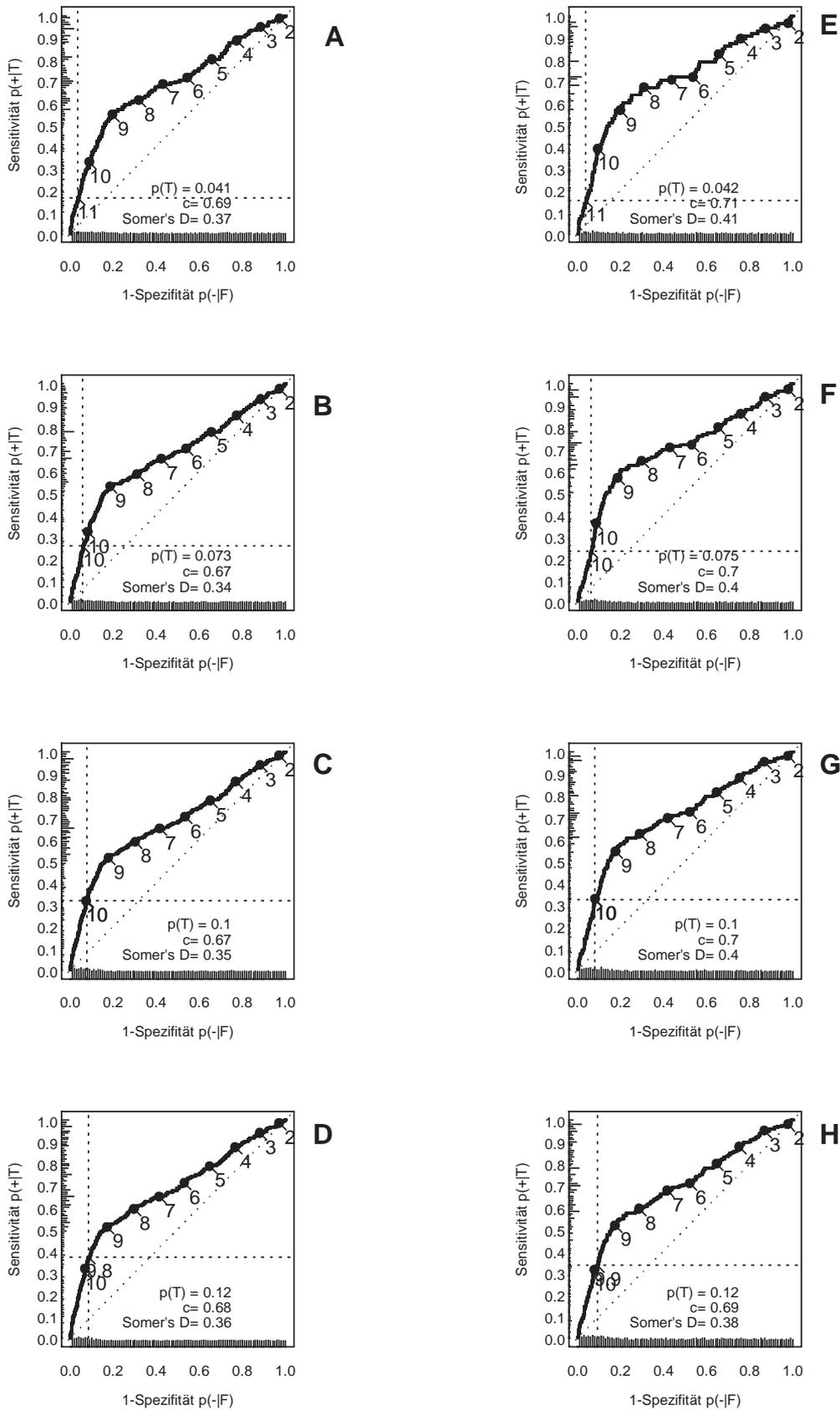


Abbildung 23: ROC-Kurven für Prognoseschätzung bei den simulierten Daten mit einem Parameter (Variable F3)

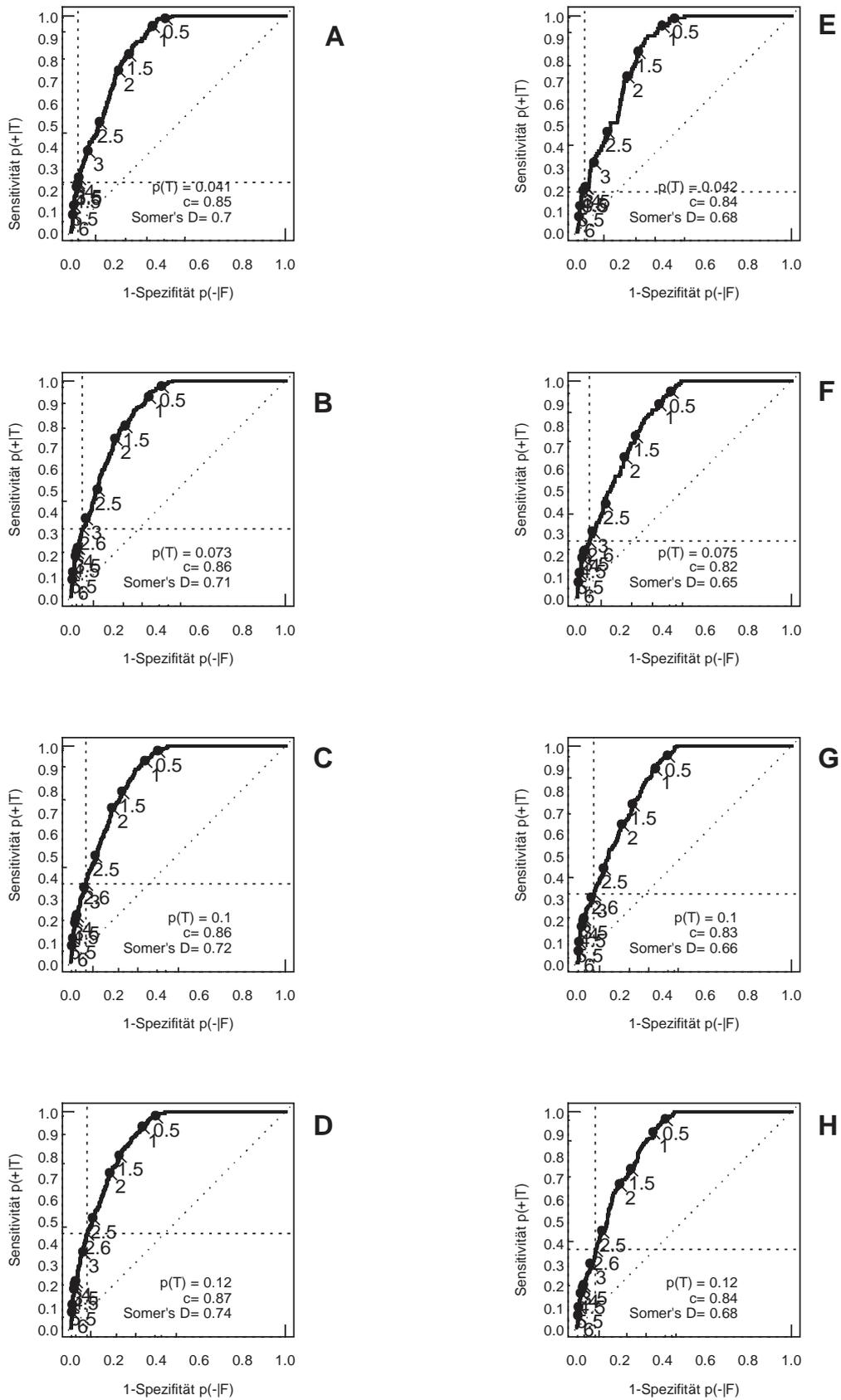


Abbildung 24: ROC-Kurven für die Prognoseschätzung bei den simulierten Daten mit einem CART-Baum

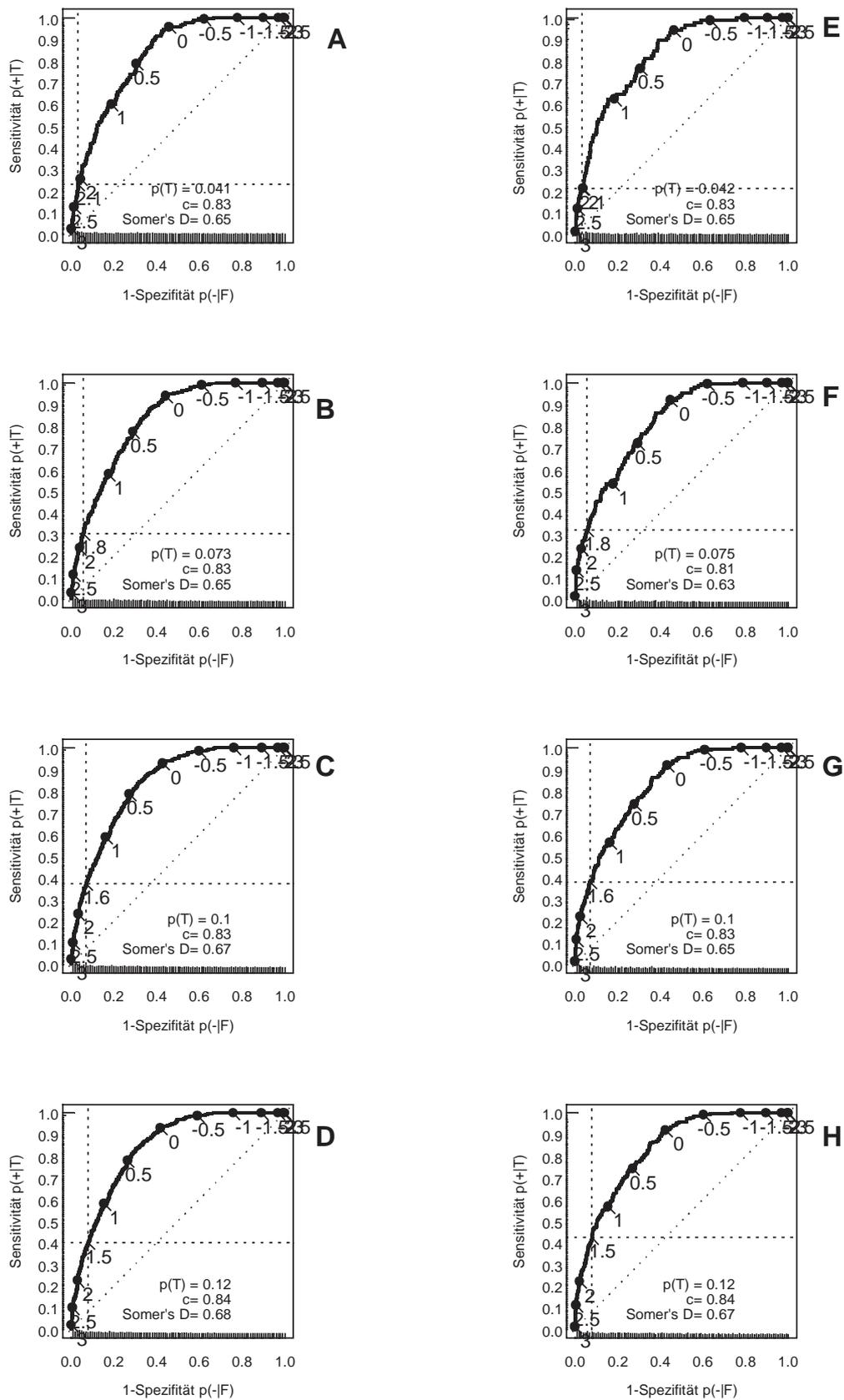


Abbildung 25: ROC-Kurven für die Prognoseschätzung bei den simulierten Daten mit einem Cox-PH-Modell

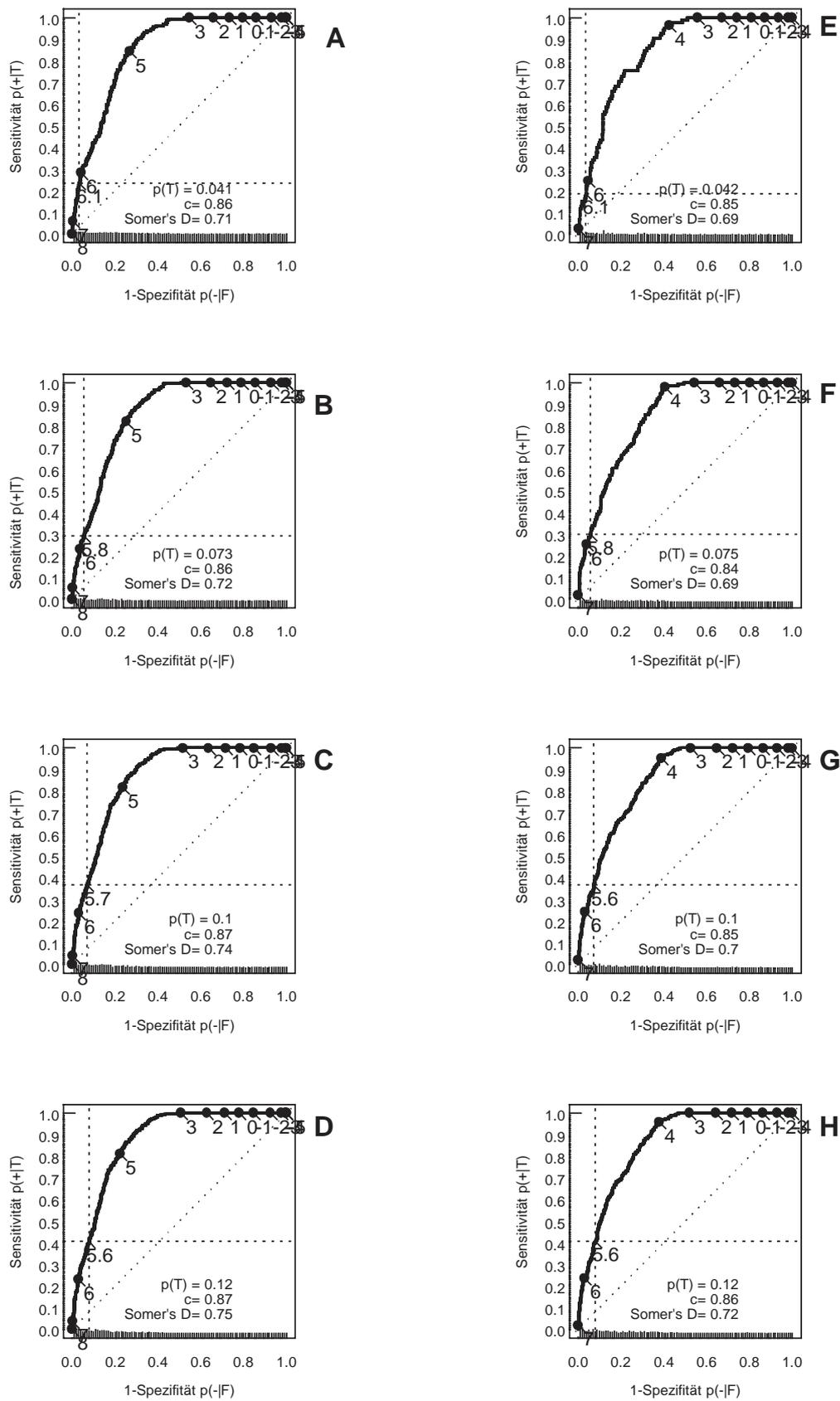


Abbildung 26: ROC-Kurven für die Prognoseschätzung bei den simulierten Daten mit einem künstlichen neuronalen Netz

Beim Vergleich der Rang-Korrelationen wurde Somer's D berechnet. Dazu wurde eine spezielle Funktion aus Harrell's Hmisc-Library verwendet, die mit rechtszensierten Daten umgehen kann. Die Ergebnisse für die Trainingsdaten sind in Tabelle 7 dargestellt. Bei diesen Daten wurde mit dem künstlichen neuronalen Netz die höchste Prognosequalität erreicht. Nahezu gleichgute Prognoserrgebnisse lieferte auch das CART-Modell.

Tabelle 7: Rang-Korrelationen für die Trainingsdaten und die verschiedenen Modelle bei den simulierten Daten (4000 Fälle, davon 474 nicht zensiert und 2251972 relevante Paare)

Gruppe	C	Index	S.D.	Konkordante Paare
F3	0.66	0.32	0.031	1493828
CART	0.84	0.68	0.013	1889216
COX-PH	0.82	0.63	0.016	1836776
KNN	0.84	0.68	0.014	1893820

Bei Überprüfung der Modelle erreichten das künstliche neuronale Netz das CART-Modell die besten Schätzungen (Tabelle 8). Die anderen drei Modelle liegen von der Wertigkeit dicht beieinander. Wie bei der ROC-Analyse für die Klassifikationsgüte zu bestimmten Zeitpunkten liefert das künstliche neuronale Netz die besten Prognoseschätzungen.

Tabelle 8: Rang-Korrelationen für die Testdaten und die verschiedenen Modelle bei den simulierten Daten (2000 Fälle, davon 240 nicht zensiert und 562994 relevante Paare)

Gruppe	C	Index	S.D.	Konkordante Paare
F3	0.69	0.38	0.041	388854
CART	0.82	0.64	0.021	462759
COX-PH	0.80	0.61	0.026	452764
KNN	0.83	0.66	0.021	467164

Analyse der Daten des Zentralregisters Malignes Melanom der Deutschen Dermatologischen Gesellschaft

Zu Melanomdaten gibt es bereits eine Vielzahl prognostischer Auswertungen. Als die wichtigsten prognostischen Parameter werden derzeit die Tumordicke nach Breslow und das Vorhandensein von Tumorulzeration im histologischen Präparat angesehen. Bei der prognostischen Auswertung in dieser Arbeit werden die Faktoren

- Tumordicke nach Breslow
- Ulzeration
- Lokalisation des Primärtumors
- Invasionslevel nach Clark
- Alter
- Geschlecht
- histologischer Subtyp

betrachtet. Die wichtigsten Variablen sind in der folgenden Zusammenstellung enthalten. Dabei wurden die Daten insgesamt betrachtet. Die für den Methodenvergleich vorgenommene Unterteilung in Trainings- und Testdaten ist ebenfalls dargestellt.

3.1.6 Beschreibung der simulierten Daten

In die Analyse gingen die Daten von 8908 Patienten ein. Für den Methodenvergleich wurden mittels Zufalls-generator drei gleich große Patientengruppen gebildet. Die Verteilung der einzelnen Patienten auf die Gruppen ist Tabelle 6 zu entnehmen. In Anlehnung an die Terminologie aus dem Bereich der neuronalen Netze werden im folgenden die Daten aus Gruppe 1 als Trainingsdaten, aus Gruppe 2 als Validierungsdaten und in Gruppe 3 als Testdaten bezeichnet. Für statistische Analysen ist eine Unterteilung in Trainings- und Validierungsdaten nicht üblich. Daher werden in der folgenden Darstellung der Verteilungen der einzelnen Variablen die Gruppen 1 und 2 zusammengefaßt.

Alle relevanten Variablen wurden auf Unterschiede, bezogen auf die drei Gruppen, geprüft. Dabei wurde bei keinem Faktor ein signifikanter Unterschied gefunden. In Tabelle 9 ist die Auswahl der Patienten in Abhängigkeit vom meldenden Zentrum dargestellt. Auch auf die Klinik bezogen, gab es keine signifikanten Unterschiede in den Verteilungen. In dieser Arbeit wurden für die meisten Auswertungen Gruppe 1 und Gruppe 2 zu den Trainingsdaten zusammengefaßt. Die Fälle aus Gruppe 3 wurden als Testdaten verwendet.

Tabelle 9: Verteilung der Patienten aus den ausgewählten acht Kliniken auf die drei Analysegruppen

Klinik	Gruppe			Total
	1	2	3	
Berlin (UKBF)	298	292	265	855
Greifswald	98	105	115	318
Lübeck	145	122	151	418
Magdeburg	376	376	371	1123
Halle	470	436	436	1342
Münster	146	137	124	407
Tübingen	1238	1281	1258	3777
Dresden-Friedrichst.	199	219	250	668
Total	2970	2968	2970	8908

Tabelle 10: Verteilung der untersuchten Faktoren bei 8908 Patienten

Faktor	Verteilung	
Geschlecht (n=8908)		
männlich	3722	(41,8 %)
weiblich	5186	(58,2 %)
Alter (n=8890)		
Mittelwert	52,9 Jahre	
25%-Quantil	41 Jahre	
Median	53 Jahre	
75%-Quantil	65 Jahre	
Tumordicke (n=8204)		
Mittelwert	1,64 mm	
25%-Quantil	0,5 mm	
Median	0,95 mm	
75%-Quantil	2,0 mm	
≤ 0,75 mm	3370	
0,76 - 1,5 mm	2109	
1,51 - 4 mm	2058	
> 4 mm	667	
Invasionslevel nach Clark (n=8327)		
II	1809	(21,7 %)
III	3543	(42,5 %)
IV	2632	(31,6 %)
V	343	(4,1 %)
Lokalisation (n=8866)		
Gesicht	819	(9,2 %)
sonstiger Kopf	295	(3,3 %)
Hals	169	(1,9 %)
Brust/Bauch	697	(7,9 %)
Rücken	2262	(5,69 %)
Unterbauch	255	(2,9 %)
Gesäß	99	(1,1 %)
Anogenitalregion	45	(0,5 %)
Oberarm	844	(9,5 %)
Unterarm	452	(5,1 %)

Hand	114	(1,3 %)
Oberschenkel	729	(8,2 %)
Unterschenkel	1557	(17,67 %)
Fuß	513	(5,8 %)
Schleimhaut	9	(0,1 %)
Histologischer Typ (n=8689)		
superfiziell spreitendes Melanom (SSM)	5219	(60,1 %)
noduläres Melanom (NM)	1722	(19,8 %)
Lentigo-Maligna Melanom (LMM)	824	(9,5 %)
Akrolentiginöses Melanom (ALM)	296	(3,4 %)
Sonstiges	628	(7,2 %)
histologisch Ulzeration (n=6795)		
Ulzeration	876	(12,8 %)
keine Ulzeration	5919	(87,1 %)
histologisch Regression (n=6578)		
Regression	717	(10,8 %)
keine Regression	5861	(89,2 %)
Art der operativen Versorgung (n=8836)		
einzeitig	4629	(52,0 %)
zweizeitig	4207	(47,2 %)
Nachbeobachtungszeit (n=8908)		
Mittelwert	60,1 Monate	
25%-Quantil	23 Monate	
Median	51 Monate	
75%-Quantil	93 Monate	
Status (n=8908)		
lebt/zensiert	7960	(89,4 %)
Tod (Todesursache Melanom)	948	(10,6 %)

Tabelle 11: Verteilung der untersuchten Faktoren bei 7780 Patienten (Patienten mit fehlenden Angaben ausgeschlossen)

Faktor	Verteilung	
Geschlecht		
männlich	3266	(42 %)
weiblich	4515	(58 %)
Alter		
Mittelwert	52,7 Jahre	
25%-Quartil	41 Jahre	
Median	53 Jahre	
75%-Quartil	65 Jahre	
Tumordicke (n=8204)		
Mittelwert	1,59 mm	
25%-Quantil	0,5 mm	
Median	0,9 mm	
75%-Quantil	2,0 mm	
≤ 0,75 mm	3271	(42,0 %)
0,76 - 1,5 mm	2035	(26,2 %)
1,51 - 4 mm	1885	(24,2 %)
> 4 mm	589	(7,6 %)
Invasionslevel nach Clark		
II	1691	(21,7 %)
III	3369	(43,3 %)
IV	2430	(31,2 %)
V	290	(3,7 %)
Lokalisation		
Gesicht/sonstiger Kopf	1083	(13,9 %)
Rumpf	3025	(38,9 %)
obere Extremität	1245	(16,0 %)
untere Extremität	2427	(31,2 %)

Histologischer Typ

superfiziell spreitendes Melanom (SSM)	4952	(63,7 %)
noduläres Melanom (NM)	1397	(18,0 %)
Lenigo-Maligna Melanom (LMM)	733	(9,4 %)
Akrolentiginöses Melanom (ALM)	241	(3,1 %)
Sonstiges	457	(5,9 %)

Nachbeobachtungszeit

Mittelwert	59,5 Monate
25%-Quantil	24 Monate
Median	51 Monate
75%-Quantil	92 Monate

Status

lebt/zensiert	7046	(90,6 %)
Tod (Todesursache Melanom)	734	(9,4 %)

Gruppe

1	2593	(32,9 %)
2	2594	(33,7 %)
3	2593	(33,4 %)

3.2 Analyse der Melanomdaten ohne Fälle mit fehlenden Angaben

3.2.1 Kaplan-Meier-Kurven und Log-Rank-Tests

In Tabelle 12 und ab Abbildung 27 sind die Kaplan-Meier-Überlebenskurven und die Verteilungen für die untersuchten Variablen bei den Trainingsdaten dargestellt. Im Log-Rank-Test haben alle Variablen einen deutlichen Einfluß auf die Prognose. Am stärksten trennt der Faktor Tumordicke zwischen guter und schlechter Prognose, am schlechtesten die Faktoren Lokalisation und Alter. Faktoren, wie Invasionslevel und Ulzeration sind stark mit der Tumordicke korreliert. Hier muß in der multifaktoriellen Analyse der Daten überprüft werden, welchen unabhängigen Einfluß die einzelnen Faktoren auf das Überleben besitzen.

Tabelle 12: Untersuchte Variablen, Fallzahlen und Ereignisse

Faktor	Anzahl	beobachtete Ereignisse	Erwartungswert	p-Wert (Log-Rank)
Tumordicke				< 0,0001
pT1 ($\leq 0,75$ mm)	1793	20	132,8	
pT2 (0,76-1,5 mm)	1033	56	81,6	
pT3 (1,51-4,0 mm)	907	146	73,1	
pT4 (> 4 mm)	288	84	18,5	
Ulzeration				< 0,0001
beschrieben	492	113	33,6	
nicht beschrieben	3529	193	272,4	
Geschlecht				< 0,0001
Männer	1723	183	106	
Frauen	2295	123	113	
Invasionslevel				< 0,0001
II	873	7	65,39	
III	1797	93	143,76	
IV	1211	176	87,7	
V	140	30	9,15	

Fortsetzung Tabelle 12: Untersuchte Variablen, Fallzahlen und Ereignisse

Faktor	Anzahl	beobachtete Ereignisse	Erwartungs-wert	p-Wert (Log-Rank)
Alter				0,0022
Gruppe 1 (≤ 40 Jahre)	1309	78	104,6	
Gruppe 2 (40-70 Jahre)	1342	118	112,8	
Gruppe 3 (> 70 Jahre)	1370	110	88,6	
Lokalisation				0,0044
nonTANS	2015	130	155	
TANS	2006	176	151	
Histologischer Typ				$< 0,0001$
SSM	2671	157	207,33	
NM	642	106	48,21	
LMM	334	16	25,49	
ALM	116	13	8,21	
anderes	258	14	16,75	

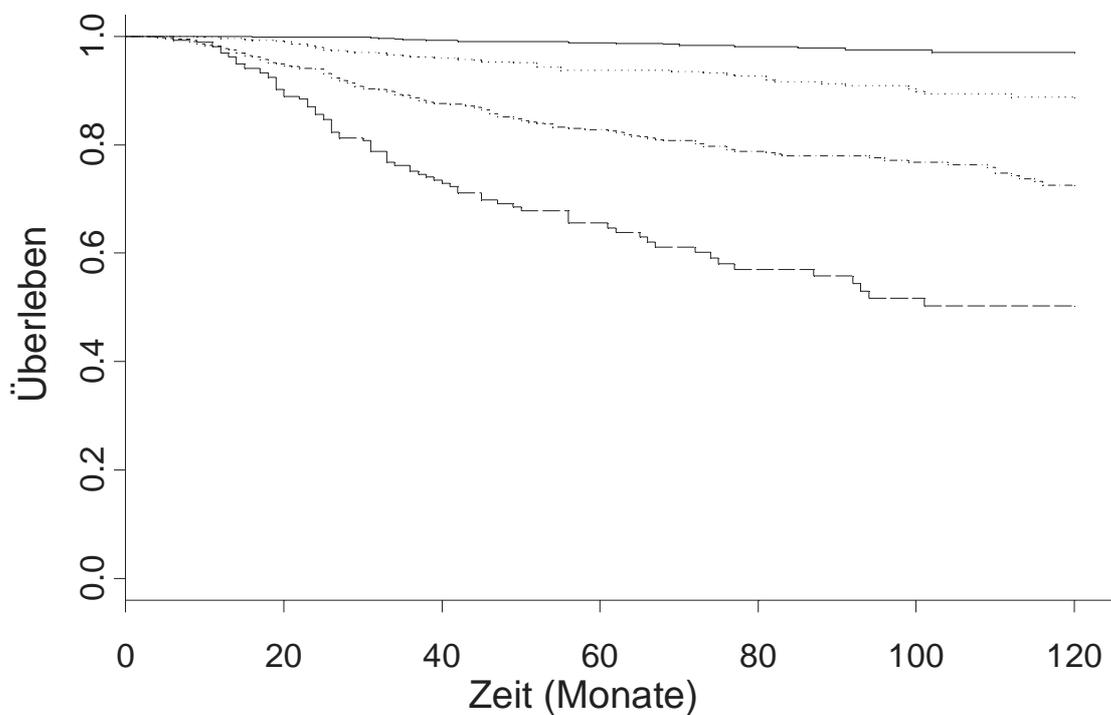


Abbildung 27: Überlebenskurven für die Variable Tumordicke

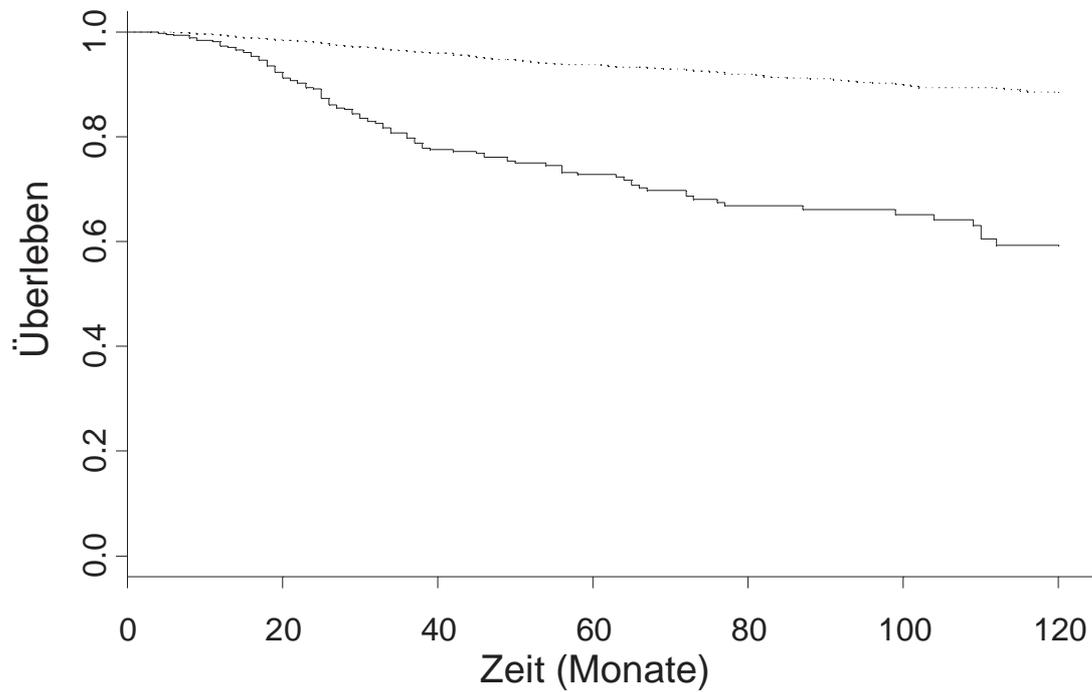


Abbildung 28: Überlebenskurven für die Variable Ulzeration

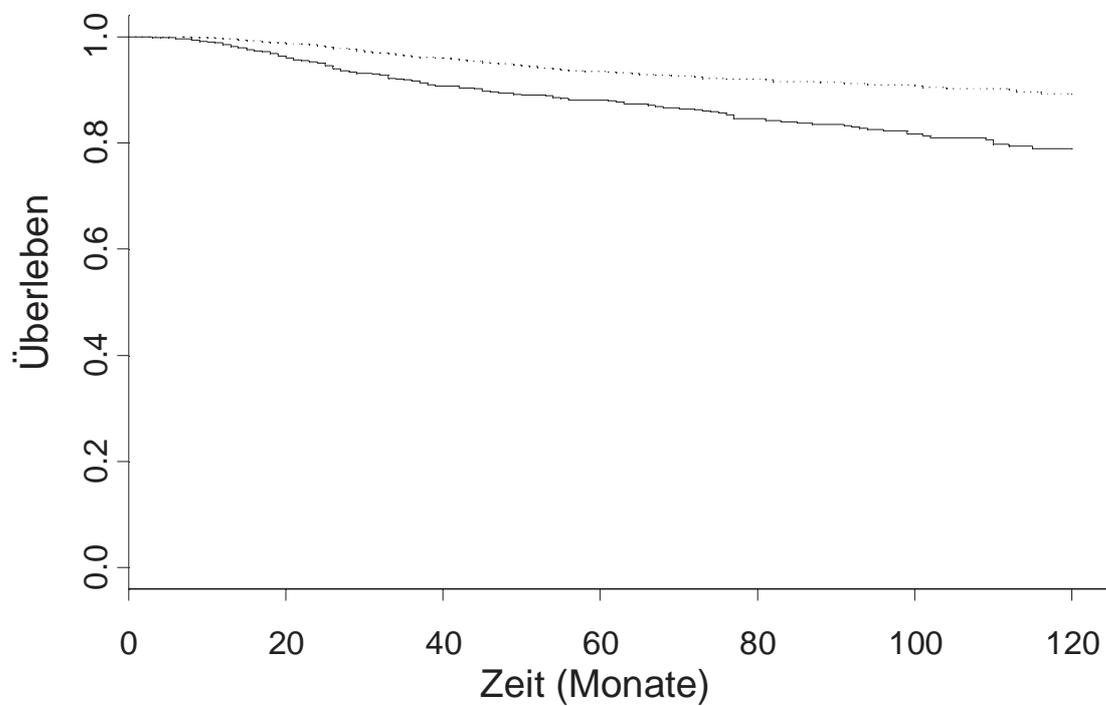


Abbildung 29: Überlebenskurven für die Variable Geschlecht

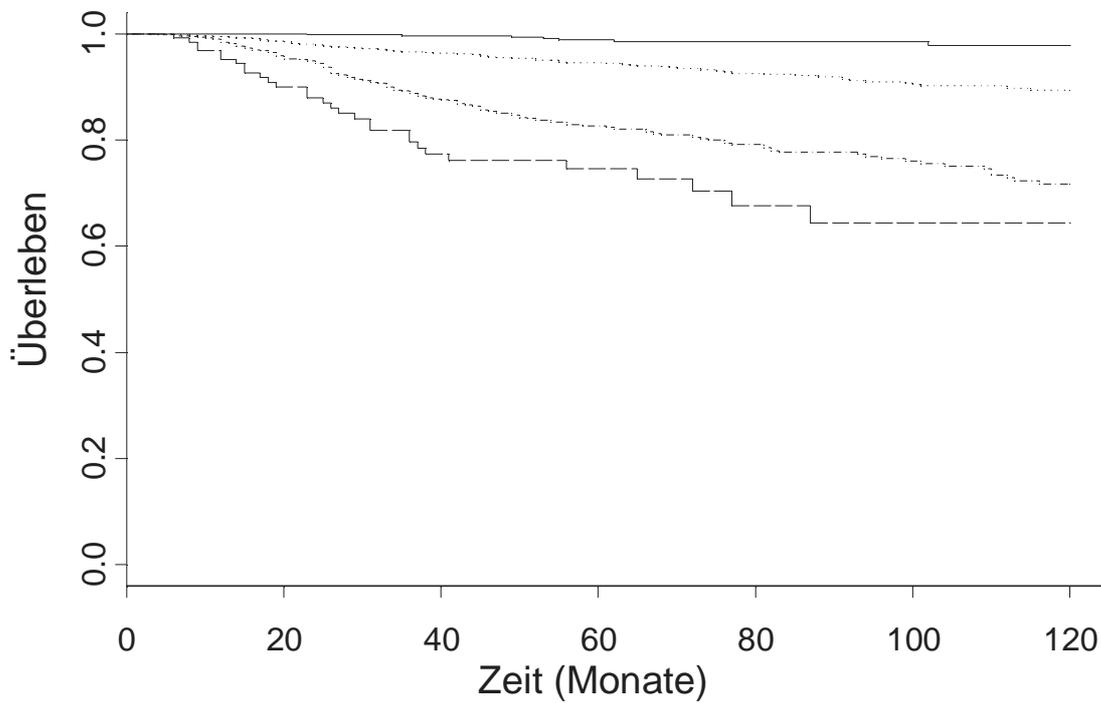


Abbildung 30: Überlebenskurven für die Variable Invasionslevel

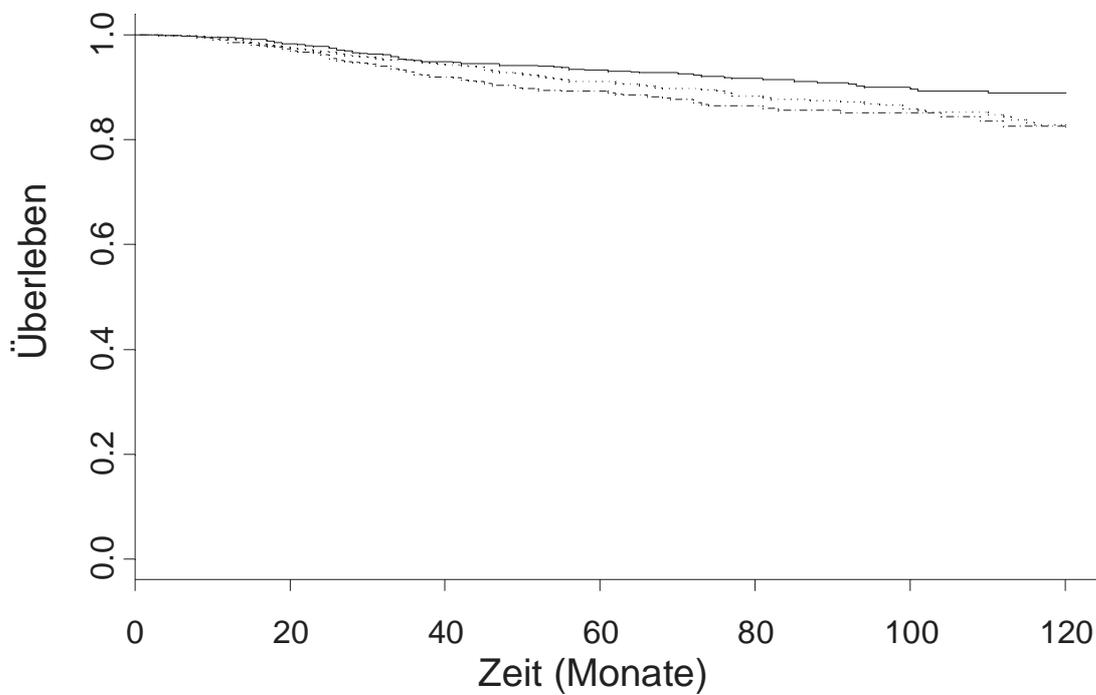


Abbildung 31: Überlebenskurven für die Variable Alter

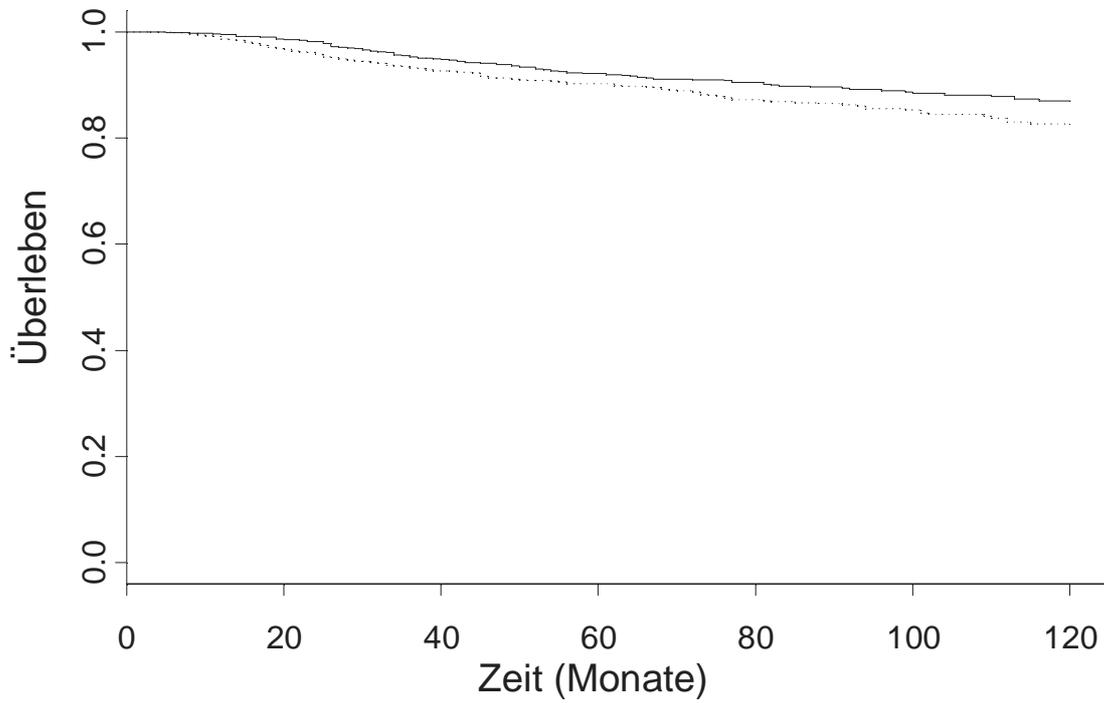


Abbildung 32: Überlebenskurven für die Variable Lokalisation

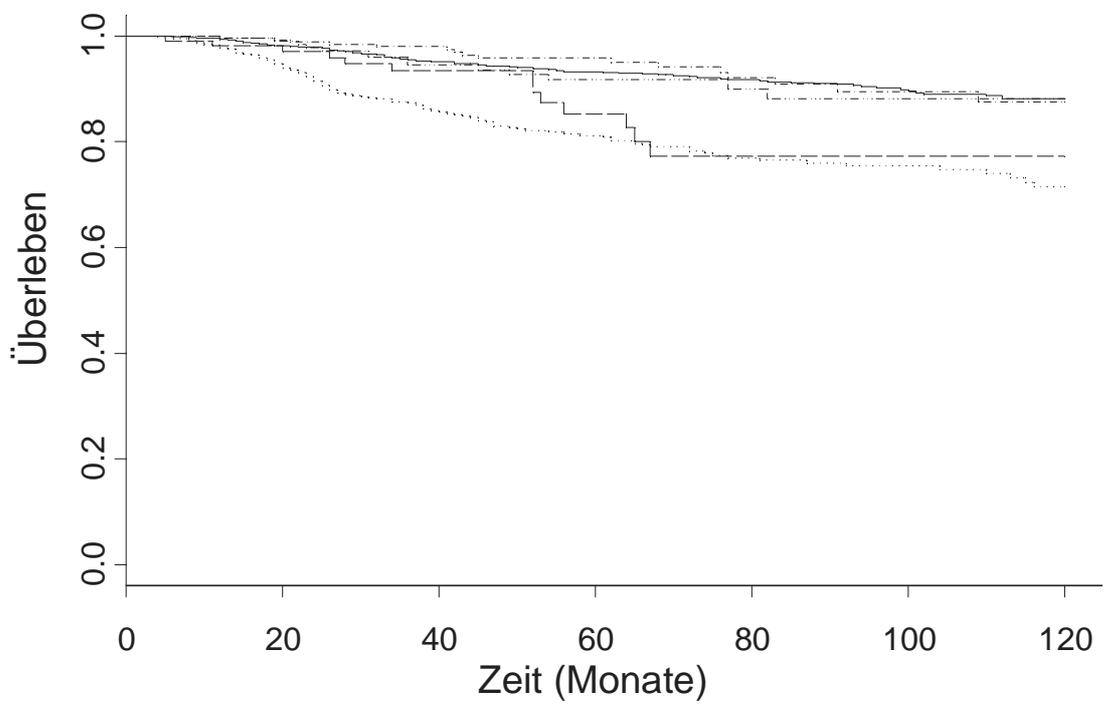


Abbildung 33: Überlebenskurven für die Variable histologischer Typ

3.2.2 CART-Analyse

Bei der CART-Analyse erweist sich ein Baum mit relativ wenigen Knoten bereits als optimal. Mit mehr als acht Knoten läßt sich keine weitere Reduktion des Fehlers erreichen. Bereits mit vier Knoten läßt sich nach den Ergebnissen in Abbildung 35 eine gute Prognosequalität erzielen.

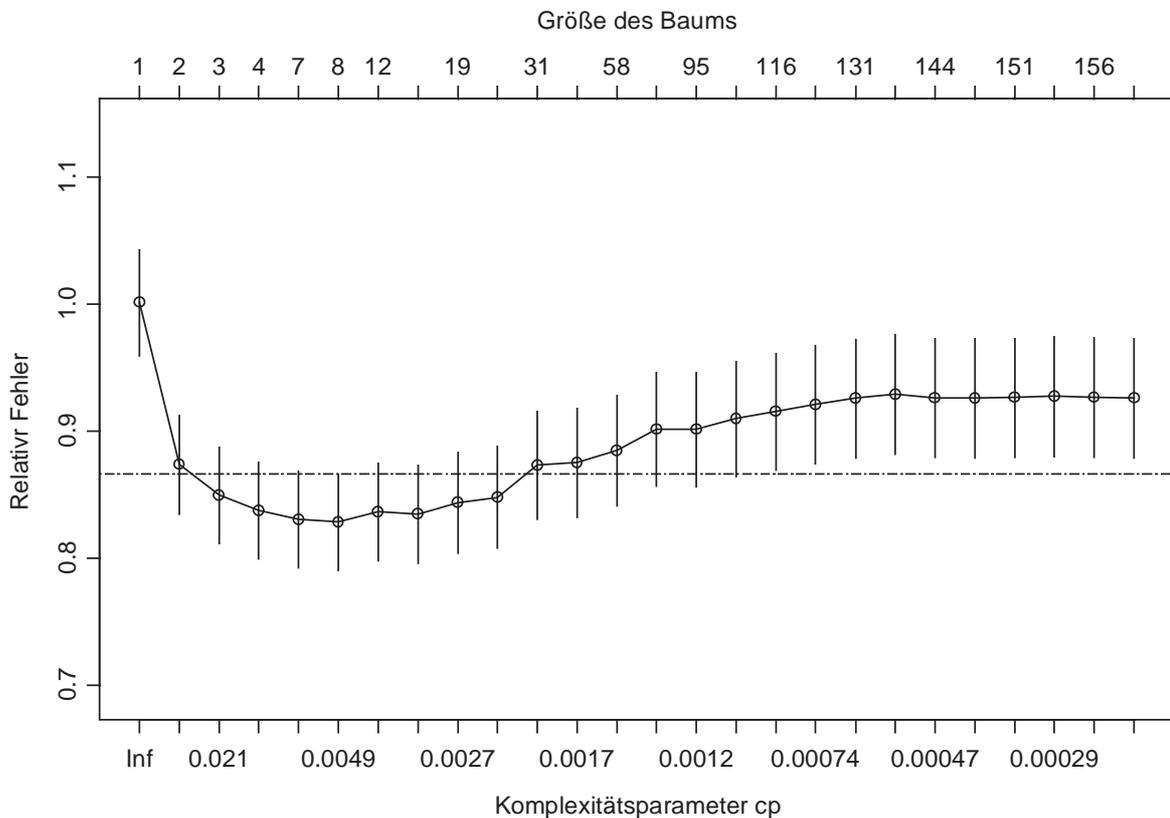


Abbildung 34: Darstellung der optimalen Baumgröße basierend auf einem Parameter für die Komplexität

Als wichtigster prognostischer Faktor erweist sich auch mit dieser Methode die Tumordicke nach Breslow. In den obersten sechs Knoten steht nur die Tumordicke. Diese Ergebnisse sind zustande gekommen, ohne daß auf die Struktur des Baumes Einfluß genommen wurde. Lediglich die Beschränkung auf sieben Knoten erfolgte manuell. Als achter Knoten wäre das Alter als weitere Variable in das Modell aufgenommen worden. Da das Alter als von der Tumordicke unabhängiger Faktor in diesem Zusammenhang nur schwer erklärbar ist, wurde auf den achten Knoten verzichtet. Der aus der Analyse resultierende CART-Baum ist in Abbildung 35 dargestellt. Die verwendeten Grenzen bei der Tumordicke sind bedingt durch die Auswahl der Patienten. Bei diesen Grenzwerten handelt es sich nicht um biologische Schwellen, sondern um das Ergebnis einer Optimierungsprozedur.

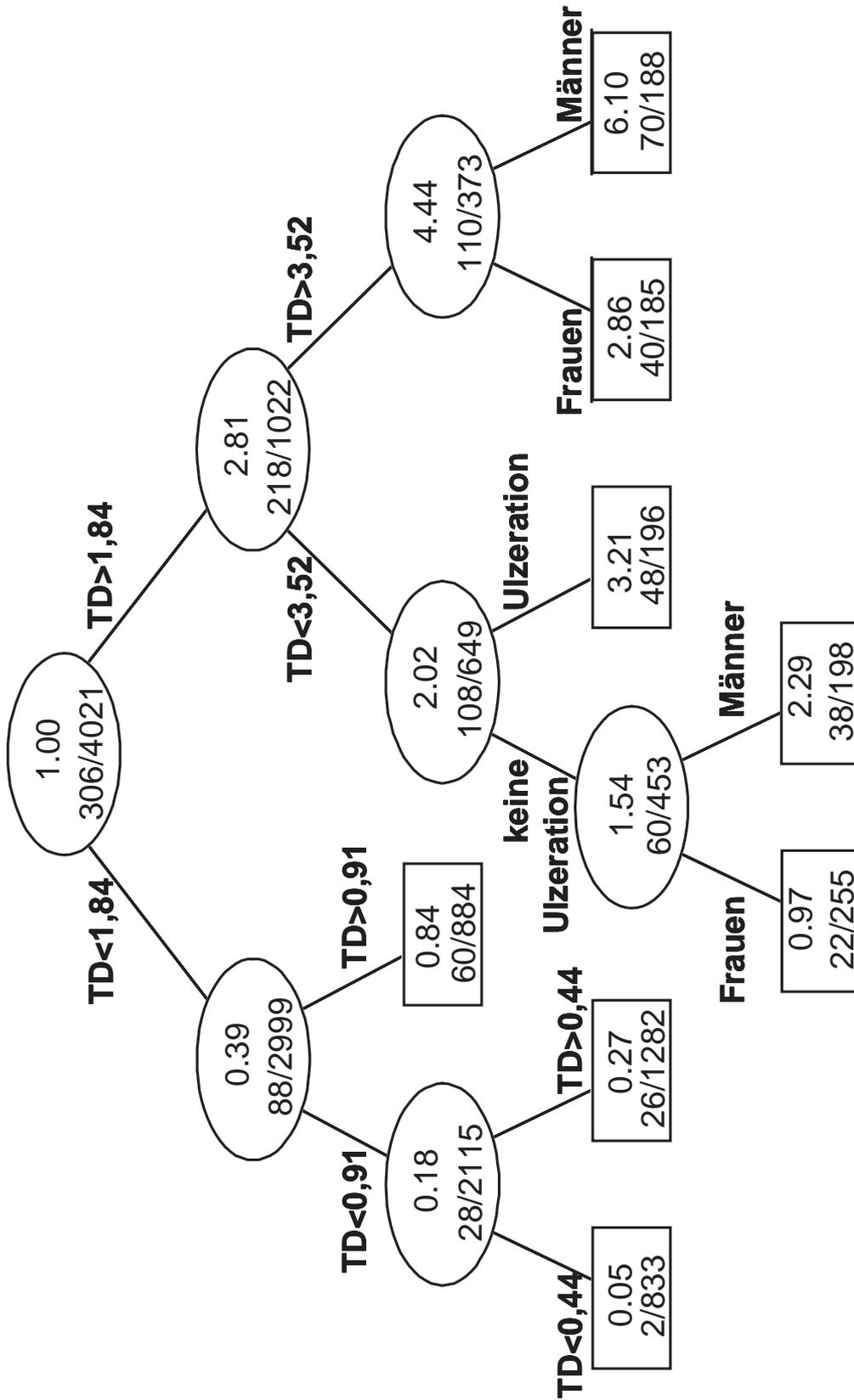


Abbildung 35: CART-Baum mit sieben Knoten für die Melanomdaten (TD steht für die Tumordicke nach Breslow)

3.2.3 Cox-PH-Analyse

Auch bei der Cox-PH-Analyse erwies sich die Tumordicke als wichtigster prognostischer Faktor. Die χ^2 -Werte der Wald-Statistik mit den zugehörigen p-Werten sind in Tabelle 13 dargestellt.

Tabelle 13: Wald-Statistik für das Cox-PH-Modell

Faktor	χ^2 -Wert	Freiheitsgrade	p-Wert
Geschlecht	21,07	1	<0,0001
Tumordicke	130,31	3	<0,0001
nichtlinear	7,48	2	0,0237
Ulzeration	9,89	1	0,0017
Histologischer Typ	6,81	4	0,1464
Invasionslevel	4,61	3	0,2025
nichtlinear	3,82	2	0,1481
Lokalisation	6,78	1	0,0092
Alter	0,35	1	0,5553
Total nichtlinear	18,06	4	0,0012
Total	342,47	14	<0,0001

Nach Anwendung einer schrittweisen Variablenreduktion verbleiben im endgültigen Modell die Faktoren Tumordicke, Ulzeration, Geschlecht und Lokalisation.

Aus den in der Einleitung beschriebenen Gründen wird für die weiteren Analysen das vollständige Modell verwendet.

Ein Problem bei Anwendung der PH-Modelle stellt die Forderung nach Proportionalität der Hazards dar. Die Überprüfung dieser Bedingung kann grafisch erfolgen oder unter Betrachtung der Schoenfeld-Residuen. Bei grafischer Betrachtung müssen sich die Kurven der Hazards bei den verschiedenen Faktoren proportional verhalten. Die in Abbildung 36 dargestellten Schoenfeld-Residuen erlauben eine einfache Überprüfung: wenn sich durch das Konfidenzband eine horizontale Linie legen lässt, ist die Proportionalitätsannahme erfüllt. Bei den hier analysierten Daten ist keine oder nur eine geringe Abweichung von der Proportionalität festzustellen. Deshalb wurde diese Bedingung als erfüllt angesehen.

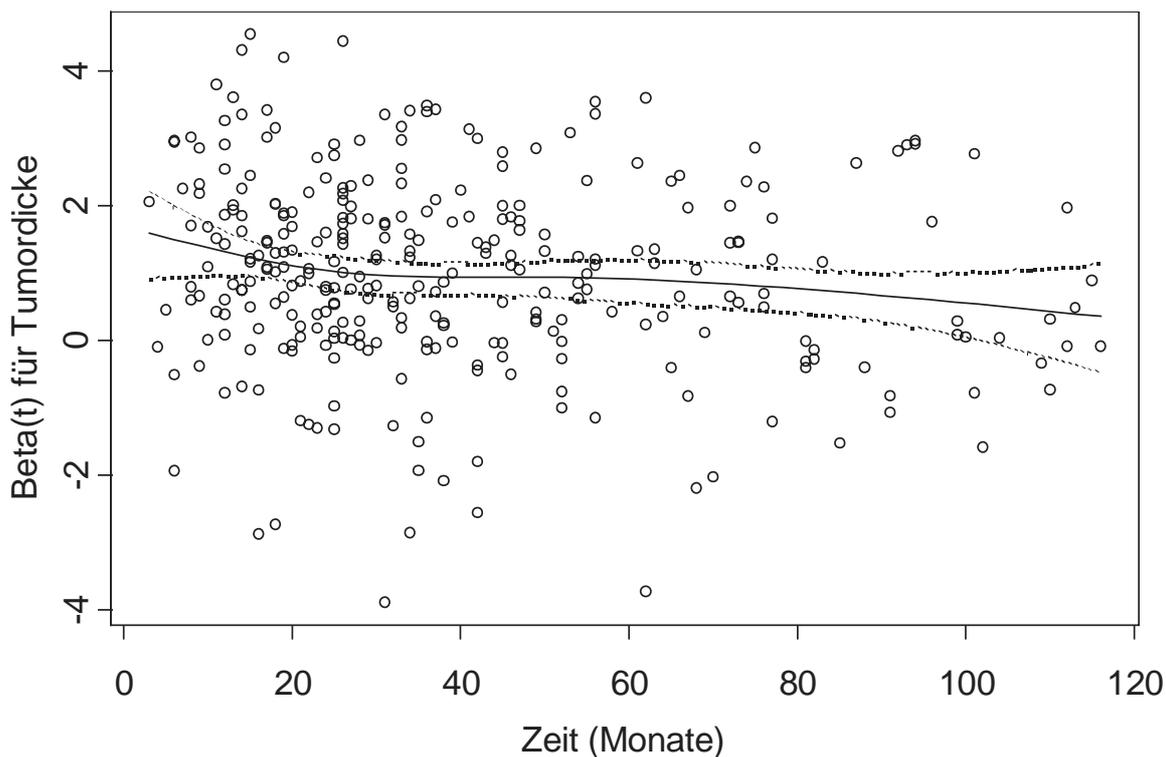


Abbildung 36: Schoenfeld-Residuen für die Tumordicke

In Abbildung 37 sind die logarithmierten relativen Hazard-Raten für die verschiedenen Faktoren dargestellt. Bei nicht stetigen Merkmalen, wie z.B. beim histologischen Typ, ist für jede Merkmalsausprägung ein Wert mit 95%-Konfidenzintervall angegeben. Auch hier lässt sich noch einmal die herausragende Bedeutung der Tumordicke ablesen.

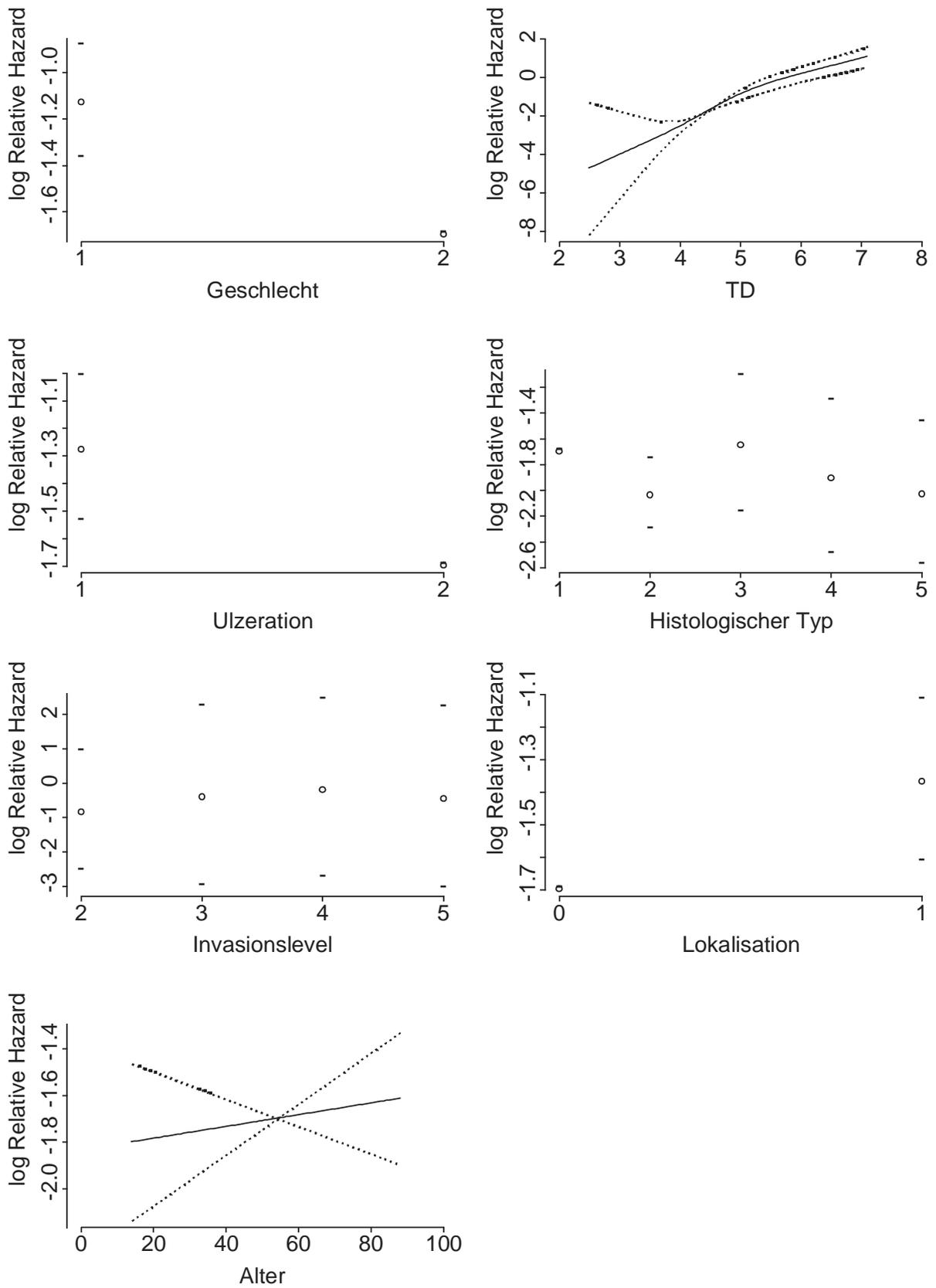


Abbildung 37: Log Relative Hazards für die untersuchten Faktoren

Um den zufälligen Effekt, der durch die Variablenselektion zustande kommen kann, einzuschätzen, wurde eine interne Validierung der Ergebnisse vorgenommen. Diese Validierung beruht auf einem Bootstrap-

Verfahren. Dabei wurden in diesem Fall 200 Stichproben mit den Daten von jeweils 200 Patienten verwendet. Die Ergebnisse dieser Validierung sind in Tabelle 14 dargestellt. In der Tabelle steht Dxy für Somer's D, R^2 für Nagelkerkes R^2 , Slope für die Abnahme der Steigung (slope shrinkage), D für Diskriminierungsindex 'discrimination index D' [(Modell L.R. χ^2-1)/L], U für den 'unreliability index' [(Differenz zwischen dem unkalibrierten X beta und X beta mit einer kalibrierten Gesamtsteigung) / L] und Q für den Gesamt-Qualitätsindex [Q = D - U. L ist -2 log likelihood mit beta=0].

Bei allen Parametern zeigte sich, daß keine wesentliche Verzerrung durch die Modellanpassung zustande gekommen ist. So kann z.B. der Faktor Slope als Korrekturfaktor angesehen werden, um für ein 'overfitting' zu korrigieren.

Tabelle 14: Ergebnisse der Validierung des COX-PH-Modells

	Original-stichprobe	Trainings-stichprobe	Test- stich-probe	Optimismus	Index (korrigiert)	Anzahl
Dxy	-0,668	-0,672	-0,661	-0,011	-0,657	200
R^2	0,153	0,157	0,149	0,008	0,145	200
Slope	1,000	1,000	0,954	0,046	0,954	200
D	0,095	0,097	0,092	0,005	0,090	200
U	0,000	0,000	0,001	0,001	0,001	200
Q	0,096	0,098	0,091	0,007	0,089	200

Tabelle 15: Ergebnisse der Kalibrierung des COX-PH-Modells

	x	Anzahl	Ereignisse	KM	Standardfehler
Punkt 1	0.7541	1005	221	0.7507	0.07663
Punkt 2	0.9321	1005	62	0.9325	0.15045
Punkt 3	0.9780	1006	20	0.9800	0.29394
Punkt 4	0.9940	1005	3	0.9968	0.70751

	Originalstichprobe	Trainingsstichprobe	Teststichprobe	Optimismus	Index (korrigiert)	Anzahl
Punkt 1	-0.003355	-0.000451	-0.001927	0.001476	-0.00483	10
Punkt 2	0.000419	-0.000678	-0.003981	0.003303	-0.00288	10
Punkt 3	0.001999	0.001373	-0.002262	0.003635	-0.00164	10
Punkt 4	0.002776	0.000993	0.000797	0.000197	0.00258	10

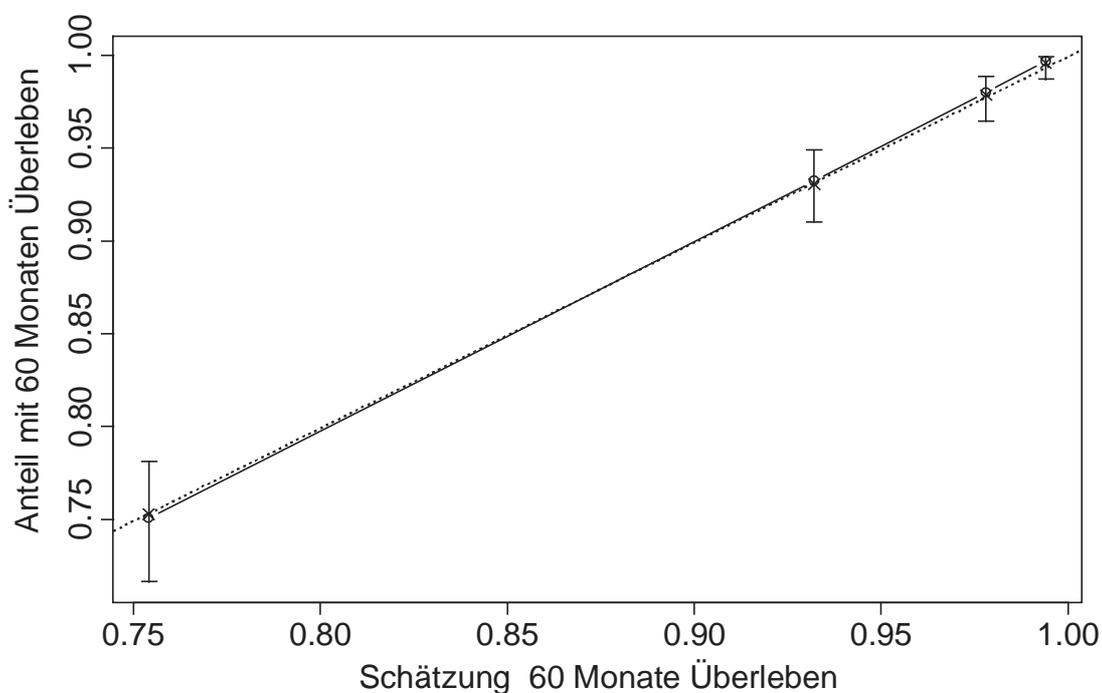


Abbildung 38: Kalibrierung des Cox-PH-Modells mit jeweils 1000 Patientendaten für jede Gruppe

3.2.4 Künstliche neuronale Netze

Bei diesen Daten wurden zwei künstliche neuronale Netze verwendet. Das erste Netz bestand aus 12 Eingabeinheiten, einer verdeckten Schicht mit zwei Einheiten und einer Ausgabereinheit. Das Netz verfügte somit über 41 zu optimierende Gewichte.

Tabelle 16: Netzgewichte des einfachen künstlichen neuronalen Netzes

	b	i1	i2	i3	i4	i5	i6
h1	-4.43	0.00	0.89	-0.66	0.03	1.58	1.14
h2	8.17	-7.00	-2.36	5.79	-2.72	6.35	3.33
o	0.00	-0.59	0.24	0.23	-0.54	-1.26	-0.94

	i7	i8	i9	i10	i11	i12	o
h1	0.48	-0.21	0.07	4.74	-0.13	0.03	6.00
h2	1.34	3.55	4.30	0.40	1.76	-0.59	-6.55
o	-0.75	0.84	-0.46	-2.10	0.46	-0.04	

Beim zweiten Netz wurde eine etwas komplexere Netzwerkstruktur gewählt. Hier bestand das Netz aus 12 Eingabeinheiten, einer verdeckten Schicht mit 3 Einheiten und einer Ausgabereinheit. Dieses Netz verfügte über 55 Gewichte.

Tabelle 17: Netzgewichte des komplexen künstlichen neuronalen Netzes

	b	i1	i2	i3	i4	i5	i6
h1	-1.84	5.93	0.61	-7.08	3.13	-4.40	-7.45
h2	-17.19	-1.82	3.86	0.85	-4.56	-5.19	1.23
h3	7.21	0.34	-0.90	0.19	0.72	0.04	-1.60
o	0.00	0.03	-0.48	-0.17	0.92	0.32	-1.45

	i7	i8	i9	i10	i11	i12	o
h1	1.41	-1.82	-4.17	-0.57	-1.10	0.48	7.51
h2	-7.07	-0.05	-1.20	3.17	-0.70	-0.16	4.91
h3	0.97	-1.33	1.20	-2.36	0.15	0.00	-9.49
o	1.53	-1.96	2.11	1.34	0.67	0.00	

3.2.5 Vergleich der Ergebnisse bei den Registerdaten

Bei diesen Daten liegen die Ergebnisse für die verschiedenen Modelle dicht beieinander. Wenn die Tumordicke allein als prognostischer Parameter verwendet wird, ergibt $c = 0,82$ (Fläche unter der ROC-Kurve). Der vergleichbare Wert bei der CART-Analyse betrug 0,81, beim Cox-PH-Modell 0,83, beim einfachen künstlichen neuronalen Netz 0,78 und beim komplexeren künstlichen neuronalen Netz 0,81. Damit unterschieden sich die Modelle nur geringfügig. Bei den Abbildungen A und E wurde jeweils die Klassifikationsgüte für einen sehr frühen Zeitpunkt (12 Monate) bestimmt. Wegen des hohen Anteils an Zensierungen insgesamt, traten hier nur unbedeutende Unterschiede zu den Daten nach 10 Jahren auf.

Insgesamt dominierte die Tumordicke als wichtigster prognostischer Faktor in allen Modellen. Die Hinzunahme weiterer potentieller Faktoren brachte keine spürbare Verbesserung der Prognoseschätzung.

Der Vergleich zwischen den beiden untersuchten künstlichen neuronalen Netzen (Abbildung 42 und Abbildung 43) brachte minimale Vorteile zugunsten des etwas komplexeren Modells. Andererseits können diese Ergebnisse auch als Bestätigung der Methode interpretiert werden, da eine völlig anderes Netz, das sogar über eine andere Netzwerkstruktur verfügt, sehr ähnlich Ergebnisse lieferte.

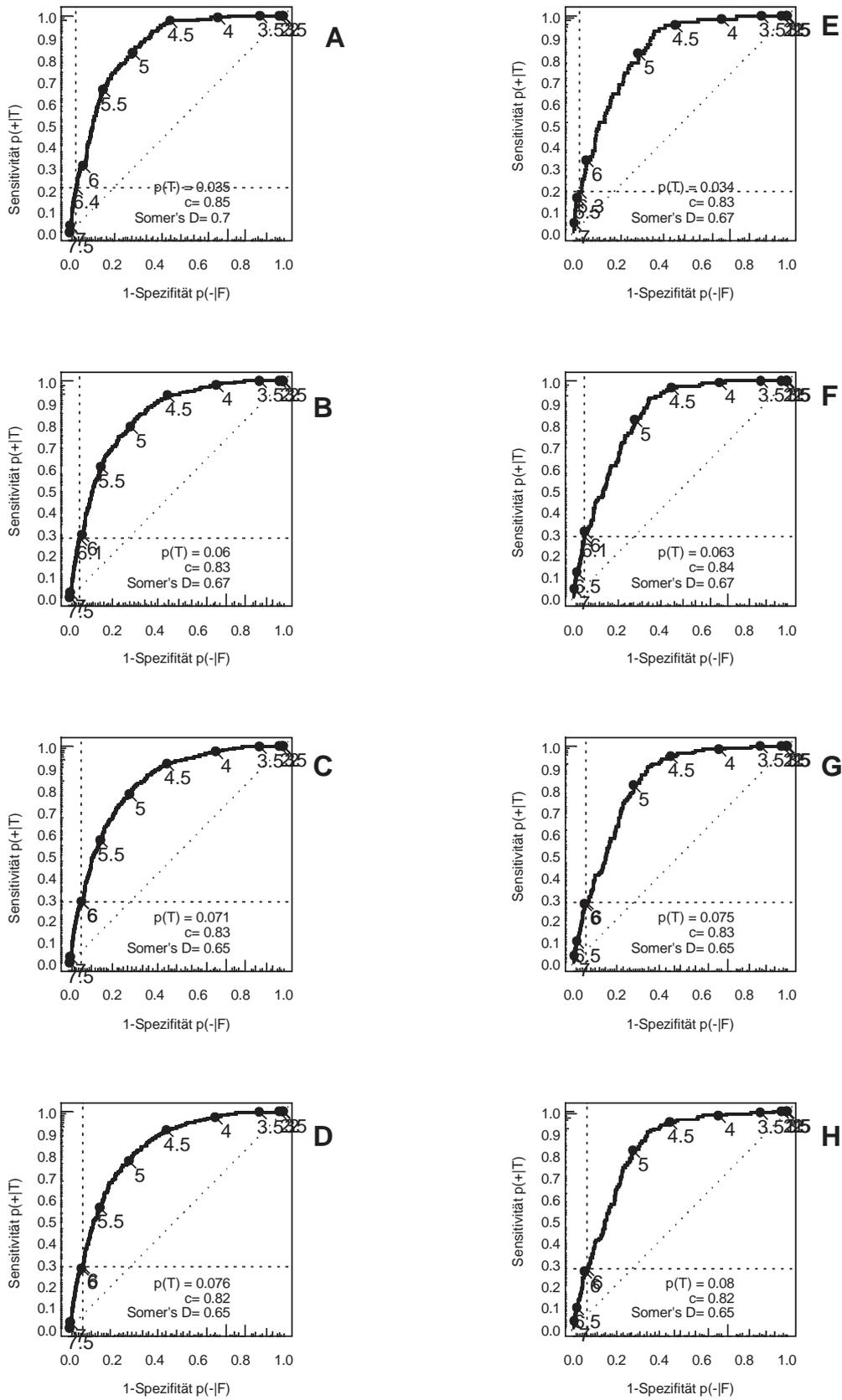


Abbildung 39: ROC-Kurven für Prognoseschätzung bei Registerdaten mit einem Parameter (Tumordicke)

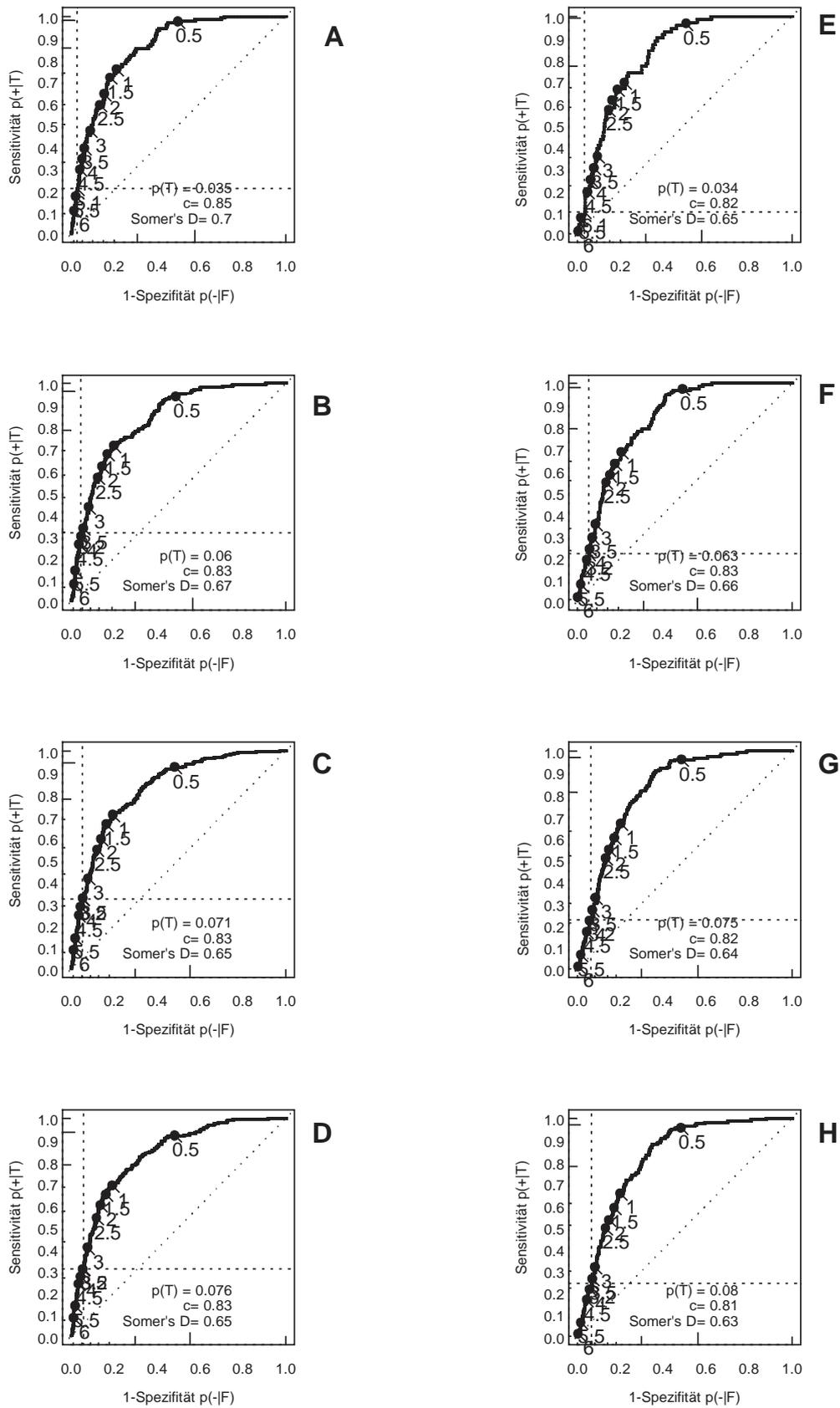


Abbildung 40: ROC-Kurven für Prognoseschätzung bei Registerdaten mit einem CART-Baum

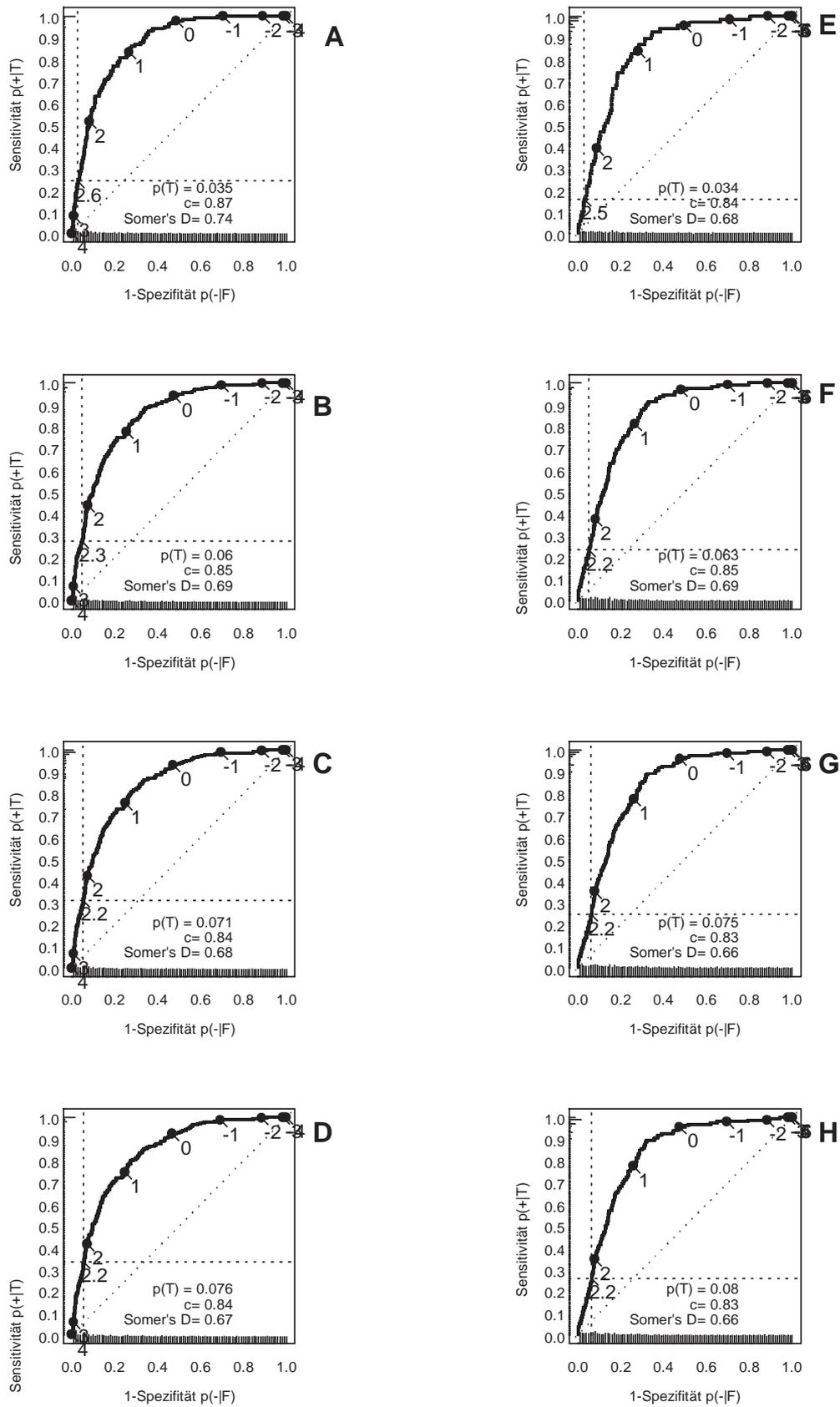


Abbildung 41: ROC-Kurven für Prognoseschätzung bei Registerdaten mit einem Cox-PH-Modell

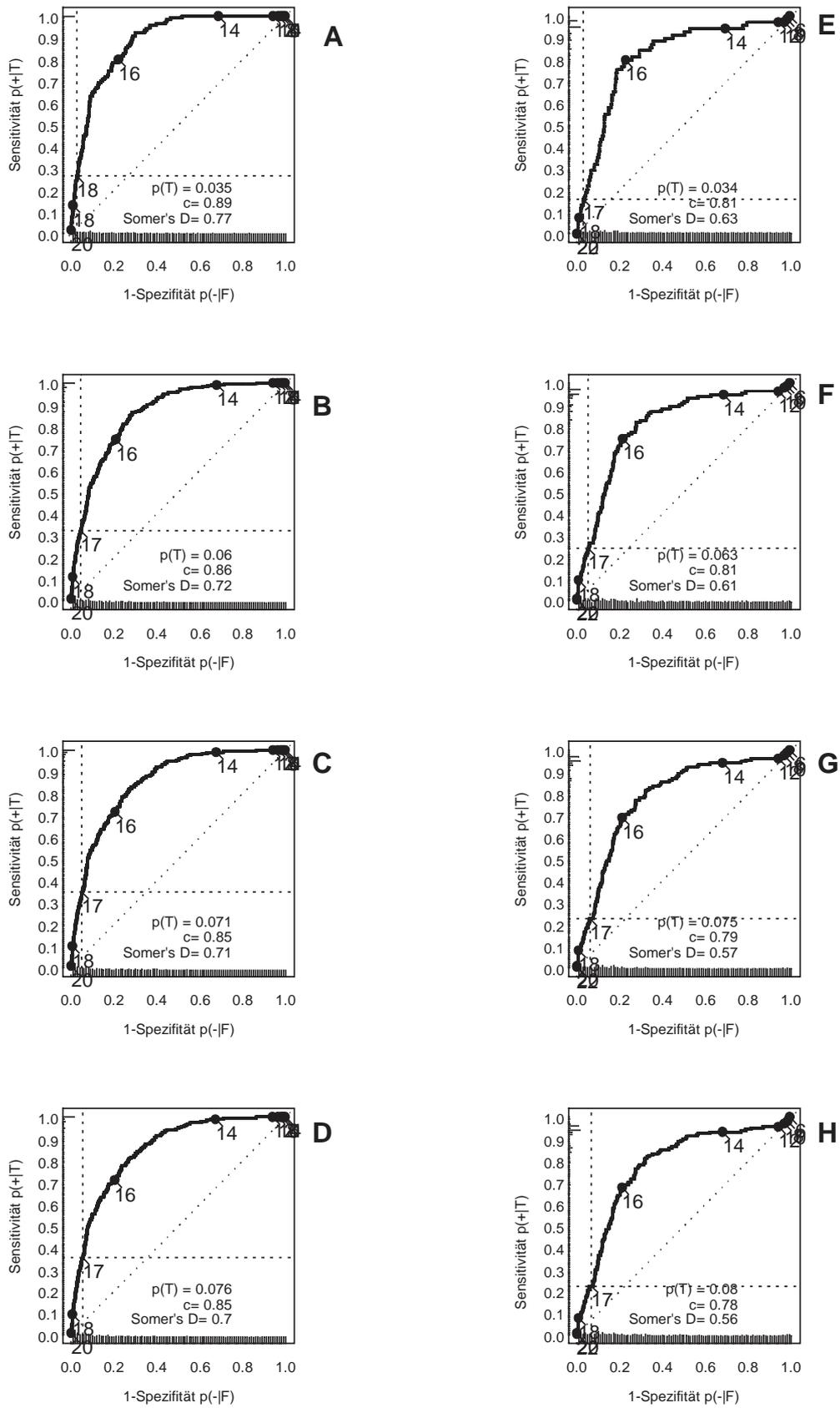


Abbildung 42: ROC-Kurven für die Prognoseschätzung bei Registerdaten mit einem künstlichen neuronalen Netz

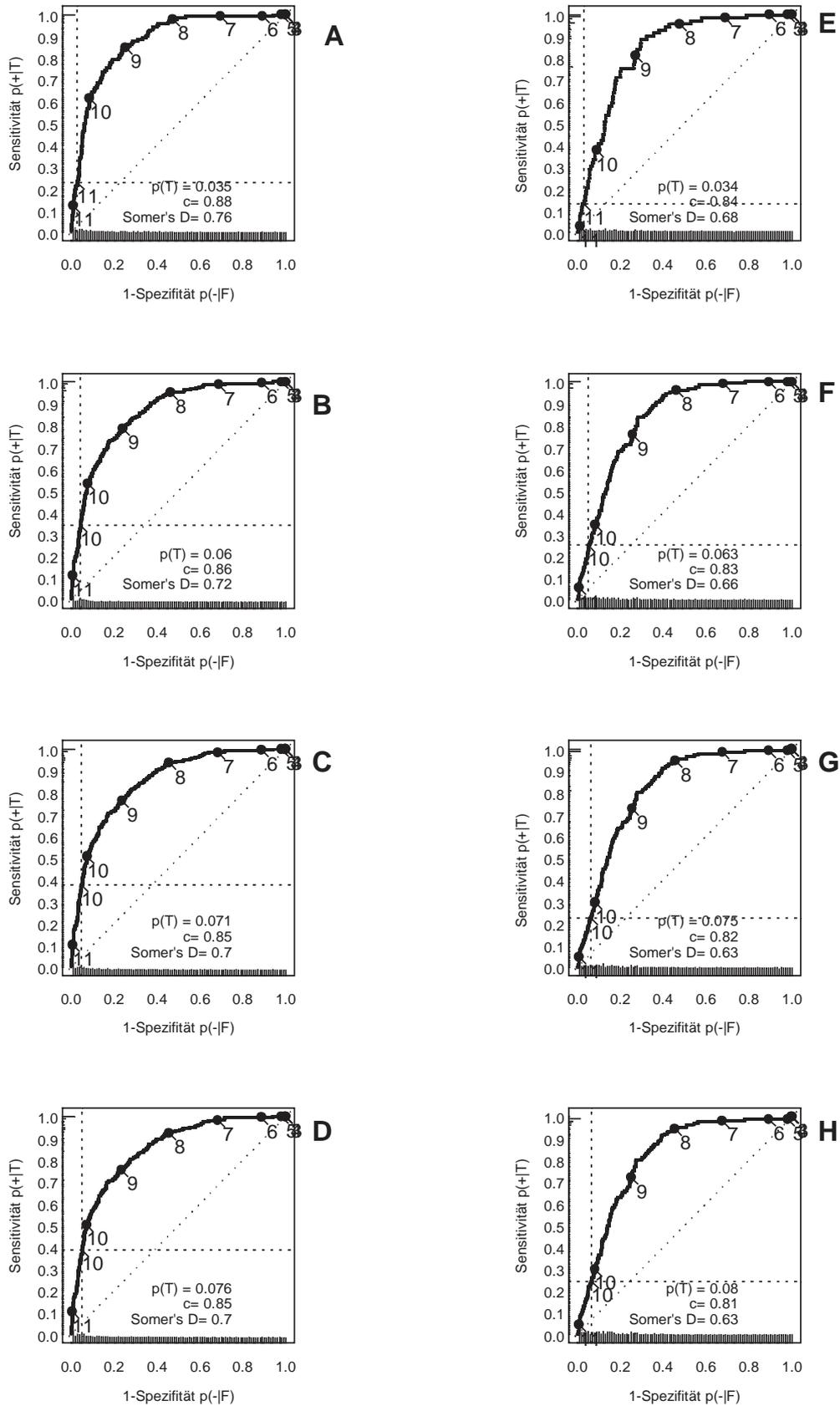


Abbildung 43: ROC-Kurven für die Prognoseschätzung bei Registerdaten mit einem komplexeren künstlichen neuronalen Netz

3.2.6 Vergleich der Rang-Korrelationen bei Registerdaten

Zum Vergleich der Methoden wurden wiederum die Rang-Korrelationen berechnet. Die Ergebnisse für die Trainingsdaten sind in Tabelle 18 dargestellt. Bei diesen Daten wurde mit dem einfacheren künstlichen neuronalen Netz höchste Prognosequalität erreicht, wenn die Überlebenszeit berücksichtigt wird.

Tabelle 18: Rang-Korrelationen für die Trainingsdaten und die verschiedenen Modelle bei 4021 Patienten, davon 306 nicht zensiert und 1460610 relevante Paare

Gruppe	C	Index	S.D.	Konkordante Paare
TD	0.82	0.63	0.021	1192891
CART	0.82	0.64	0.021	1198104
COX-PH	0.83	0.67	0.020	1218078
KNN	0.85	0.70	0.018	1243206
KNN2	0.85	0.70	0.020	1238292

Bei Überprüfung der Modelle erreicht das COX-Modell die besten Schätzungen (Tabelle 19). Die Tumordicke allein als prognostischer Faktor belegt mit nahezu identischen Werten im Vergleich zu den Trainingsdaten, daß die Zufallsauswahl keine Verzerrung enthält. Diese sehr ähnlichen Parameter entsprechen auch den Erwartungen, da die Tumordicke als einzelner Faktor nicht an die Daten angepaßt wurde.

Mit ähnlichen Werten bei den Trainings- und bei den Testdaten besitzt das CART-Modell eine hohe Konstanz. Beim CART-Modell spielt die Tumordicke die wichtigste Rolle. Bei diesem Modell hat aber durch Klassenbildung eine Anpassung an die Daten stattgefunden. Beim Cox-PH-Modell zeigte sich kein Unterschied zu Trainingsdaten. Von den bei diesen Daten geprüften Modellen, besitzt das Cox-Modell hinsichtlich der Schätzung der Ränge und damit hinsichtlich der Überlebenszeit die besten Eigenschaften.

Tabelle 19: Rang-Korrelationen für die Testdaten und die verschiedenen Modelle bei 2022 Patienten, davon 161 nicht zensiert und 361364 relevante Paare

Gruppe	C	Index	S.D.	Konkordante Paare
TD	0.82	0.64	0.027	296193
CART	0.81	0.63	0.027	293997
COX-PH	0.83	0.66	0.027	300148
KNN	0.79	0.58	0.038	285738
KNN2	0.82	0.64	0.027	295698

3.3 Analyse der Daten mit ersetzten Angaben

Fehlende Angaben lassen sich in der Praxis nie ganz vermeiden. Das Löschen von Datensätzen mit fehlenden Angaben führt zu falschen Ergebnissen, wenn das Fehlen der Werte nicht zufällig zustande gekommen ist. Bei den hier verwendeten Daten läßt sich aus Tabelle 20 entnehmen, daß die fehlenden Angaben nicht nur zufällig zustande gekommen sein können. Den Empfehlungen der Literatur folgend (siehe Abschnitt 2.7, S. 34), wurde auf das Generieren von Angaben beim wichtigsten prognostischen Faktor, der Tumordicke, verzichtet. Da das Generieren von Ersatzwerten für fehlende Angaben unabhängig von den Ergebnisvariablen läuft, kann durch die erzeugten Werte bei fehlenden Angaben unter Umständen eine Verzerrung in die Daten aufgenommen werden, was eher zu einer Verschlechterung als zu einer Verbesserung der Datenqualität führt.

3.3.1 Ersetzen fehlender Angaben

In Tabelle 20 ist die Verteilung der fehlenden Angaben dargestellt. Aus dieser Tabelle ist zu erkennen, daß die fehlenden Angaben nicht zufällig zustande gekommen sind. Zum Beispiel fehlten sehr häufig die Angaben zur Regression, wenn auch keine Angabe zur Ulzeration vorlag. Ein Verzicht auf die Fälle mit fehlenden Angaben kann zu Verzerrungen führen, die letztendlich die Prognosequalität beeinflussen. Wichtig beim Generieren von Ersatzwerten für die fehlenden Angaben ist die Unabhängigkeit von den Zielgrößen. Außerdem wird in der Literatur empfohlen, von Schätzungen bei den wichtigsten Parametern abzusehen. In den vorderen Abschnitten wurde gezeigt, daß die Tumordicke eine überragende Bedeutung bei diesen Daten besitzt. Aus diesem Grund wurde auf die Zuschreibung neuer Werte bei fehlender Tumordicke verzichtet. Fälle mit fehlenden Angaben bei der Tumordicke wurden bei dieser Auswertung nicht berücksichtigt.

Tabelle 20: Fehlende Angaben bei den Faktoren

Faktor	Level	Lokal.	Histo	Ulz.	Regr.	Alter	TD
fehlende Angaben	386	23	146	1413	1549	11	478
3841	*	*	*	*	*	*	*
122	fehlt	*	*	*	*	*	*
11	*	fehlt	*	*	*	*	*
56	*	*	fehlt	*	*	*	*
1	fehlt	*	fehlt	*	*	*	*
55	*	*	*	fehlt	*	*	*
4	fehlt	*	*	fehlt	*	*	*
1	*	*	fehlt	fehlt	*	*	*
1	fehlt	*	fehlt	fehlt	*	*	*
180	*	*	*	*	fehlt	*	*
8	fehlt	*	*	*	fehlt	*	*
5	*	*	fehlt	*	fehlt	*	*
1096	*	*	*	fehlt	fehlt	*	*
34	fehlt	*	*	fehlt	fehlt	*	*
1	*	fehlt	*	fehlt	fehlt	*	*
32	*	*	fehlt	fehlt	fehlt	*	*
2	fehlt	*	fehlt	fehlt	fehlt	*	*
7	*	*	*	*	*	fehlt	*
1	*	*	*	fehlt	*	fehlt	*
2	*	*	*	fehlt	fehlt	fehlt	*
154	*	*	*	*	*	*	fehlt
109	fehlt	*	*	*	*	*	fehlt
3	*	fehlt	*	*	*	*	fehlt
7	*	*	fehlt	*	*	*	fehlt
7	fehlt	*	fehlt	*	*	*	fehlt
4	*	*	*	fehlt	*	*	fehlt
2	fehlt	*	*	fehlt	*	*	fehlt
1	*	fehlt	*	fehlt	*	*	fehlt
2	fehlt	*	fehlt	fehlt	*	*	fehlt
7	*	*	*	*	fehlt	*	fehlt
7	fehlt	*	*	*	fehlt	*	fehlt
82	*	*	*	fehlt	fehlt	*	fehlt
54	fehlt	*	*	fehlt	fehlt	*	fehlt
6	fehlt	fehlt	*	fehlt	fehlt	*	fehlt
6	*	*	fehlt	fehlt	fehlt	*	fehlt
25	fehlt	*	fehlt	fehlt	fehlt	*	fehlt
1	fehlt	fehlt	fehlt	fehlt	fehlt	*	fehlt
1	fehlt	*	*	fehlt	fehlt	fehlt	fehlt

Tabelle 22: Verteilungen der Variablen in den Trainingsdaten. Bei den vervollständigten Werten fehlen die Fälle ohne Angabe zur Tumordicke

Variable	Verteilung aller Werte	Verteilung ohne fehlende Angaben	Verteilung mit vervollständigten Werten
Geschlecht			
männlich	2475 (41,7%)	1723 (42,9%)	2310 (42,3%)
weiblich	3463 (58,3%)	2298 (57,1%)	3150 (57,7%)
Alter			
Mittelwert	52,9 Jahre	52,8 Jahre	52,7 Jahre
Median	54 Jahre	54 Jahre	53 Jahre
≤ 45 Jahre	1934 (32,6%)	1309 (32,6%)	1816 (33,3%)
46 - 60 Jahre	1959 (33,1%)	1342 (33,4%)	1793 (32,8%)
> 60 Jahre	2034 (34,3%)	1370 (34,1%)	1851 (33,9%)
fehlende Angaben	11		
Tumordicke nach Breslow			
Mittelwert	1,64 mm	1,51 mm	1,64 mm
Median	0,95 mm	0,87 mm	0,95 mm
≤ 0,75 mm	2260 (41,4%)	1793 (44,6%)	2260 (41,4%)
0,76 - 1,5 mm	1388 (25,4%)	1033 (25,7%)	1388 (25,4%)
1,51 - 4,0 mm	1365 (25,0%)	907 (22,6%)	1365 (25,0%)
> 4 mm	447 (8,2%)	288 (7,2%)	447 (8,2%)
fehlende Angaben	478		478
Invasionslevel nach Clark			
II	1197 (21,6%)	873 (21,7%)	1182 (21,6%)
III	2387 (43,0%)	1797 (44,7%)	2396 (43,9%)
IV	1746 (31,4%)	1211 (30,1%)	1690 (31,0%)
V	222 (4,0%)	140 (3,5%)	192 (3,5%)
fehlende Angaben	386		0
Lokalisation			
nonTANS	3066 (51,6%)	2015 (50,1%)	2773 (50,8%)
TANS	2849 (48,0%)	2006 (49,9%)	2687 (49,2%)
fehlende Angaben	23		0
Histologischer Typ			
SSM	3494 (60,3%)	2671 (66,4%)	3380 (61,9%)
NM	1145 (19,8%)	642 (16,0%)	1022 (18,7%)
LMM	533 (9,2%)	334 (8,3%)	514 (9,4%)
ALM	202 (3,5%)	116 (2,9%)	189 (3,5%)
sonstige Einordnung	418 (7,2%)	258 (6,4%)	355 (6,5%)
fehlende Angaben	146		0

Fortsetzung Tabelle 22: Verteilungen der Variablen in den Trainingsdaten. Bei den vervollständigten Werten fehlen die Fälle ohne Angabe zur Tumordicke

Variable	Verteilung aller Werte	Verteilung ohne fehlende Angaben	Verteilung mit vervollständigten Werten
Ulzeration			
beschrieben	579 (12,8%)	425 (10,6%)	899 (16,5%)
nicht beschrieben	3946 (87,2%)	3529 (87,8%)	4561 (83,5%)
fehlende Angaben	1413		0
Regression			
beschrieben	479 (10,9%)	425 (11,1%)	682 (12,5%)
nicht beschrieben	3910 (89,1%)	3416 (88,9%)	4778 (87,5%)
fehlende Angaben	1549		0
Nachbeobachtungszeit			
Mittelwert	58,1 Monate	52,8 Monate	57,6 Monate
Median	52 Monate	46 Monate	51 Monate
Status			
lebt	5330 (89,8%)	3715 (92,3%)	4954 (90,7%)
Tod	608 (10,2%)	306 (7,6%)	506 (9,3%)

Durch das Ersetzen der fehlenden Angaben konnten bei den Trainingsdaten insgesamt über 1200 Fälle zusätzlich ausgewertet werden. Bei diesen Fällen traten 506 Ereignisse ein, also 200 mehr als bei der Analyse mit den Registerdaten, bei denen Fälle mit fehlenden Angaben gelöscht wurden.

3.3.2 Kaplan-Meier-Kurven und Log-Rank-Test

Auf eine Wiederholung der Darstellung der Kaplan-Meier-Überlebenskurven wird verzichtet. Die Ergebnisse unterschieden sich nur unwesentlich von der Analyse, bei der die Fälle mit fehlenden Angaben ausgeschlossen waren.

Als einziger neuer Faktor für die Analyse ist die histologische Regression hinzugekommen. Dieser potentielle Faktor zeigte keinen signifikanten Einfluß auf das Überleben ($p = 0,14$).

3.3.3 CART-Analyse

Die Entwicklung des CART-Baumes erfolgte wieder weitgehend automatisch. Der Komplexitätsparameter wurde manuell auf 0,0034 festgelegt. Damit ergab sich der in Abbildung 45 dargestellte CART-Baum. Wie bereits bei der vorangegangenen Analyse erwies sich im CART-Baum wiederum die Tumordicke als dominierender Faktor. Eine untergeordnete prognostische Relevanz konnte dem Geschlecht bei einer Tumordicke $> 1,3$ mm zugeordnet werden. Dabei hatte Frauen eine günstigere Prognose als Männer. Der prognostische Einfluß der Ulzeration hat abgenommen und erscheint nicht mehr im CART-Baum.

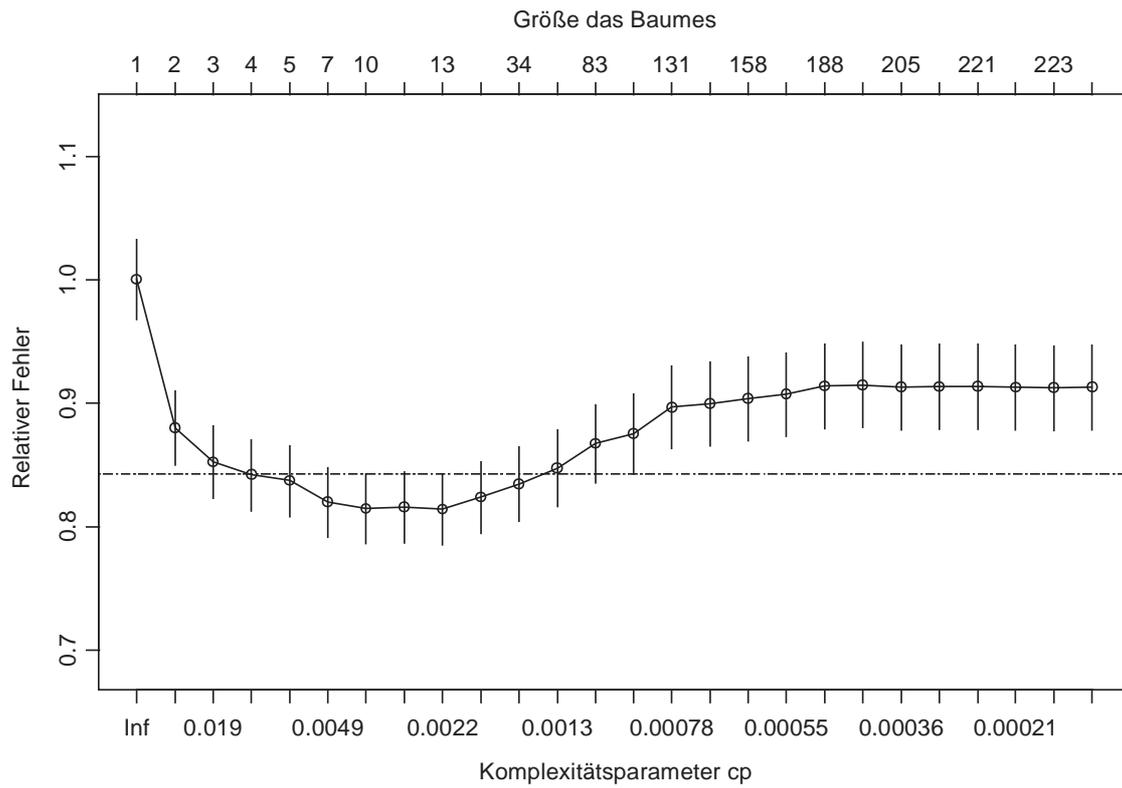


Abbildung 44: Darstellung von Fehlerrate und Komplexität des CART-Baumes für die vervollständigten Registerdaten

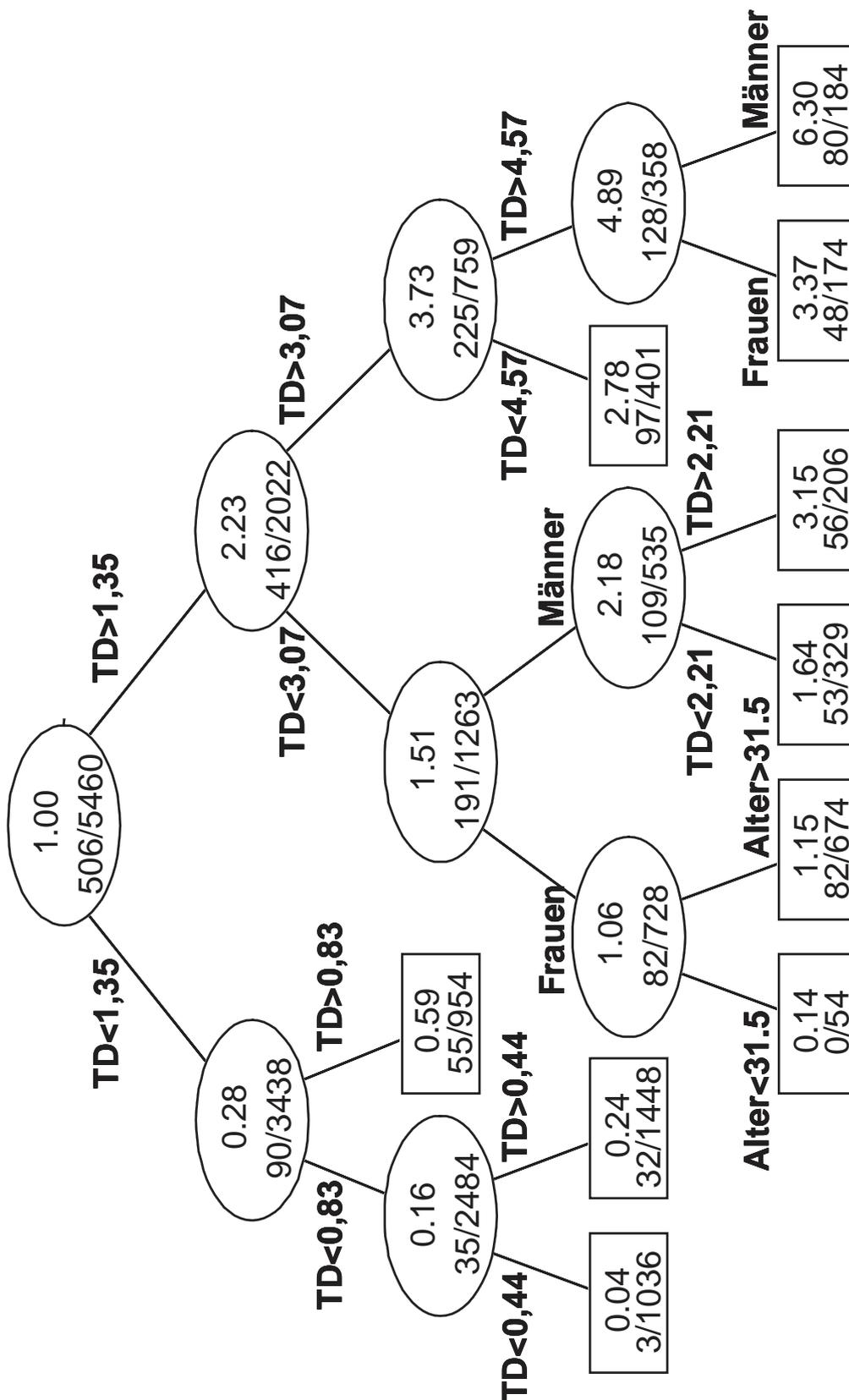


Abbildung 45: CART-Baum für die Registerdaten mit vervollständigten Angaben

3.3.4 Cox-PH-Analyse

Bei der Cox-PH-Analyse mit den vervollständigten Daten zeigten sich nur geringe Unterschiede im Vergleich zu den Daten, bei denen die fehlenden Angaben gelöscht wurden. Die Tumordicke ist mit Abstand der wichtigste prognostische Faktor. Gegenüber dem Modell ohne Fälle mit fehlenden Angaben hat der Einfluß der Tumordicke sogar noch zugenommen. Dafür ist χ^2 -Wert der Ulzeration gefallen.

Ob dieser Effekt als Artefakt durch die Ersetzung der fehlenden Angaben zu werten ist, kann mit den vorhandenen Daten nicht entschieden werden. Insgesamt ist aber der χ^2 -Wert der Wald-Statistik deutlich höher, als beim Modell ohne die Fälle mit fehlenden Angaben.

Tabelle 23: Wald-Statistik für das Cox-PH-Modell bei der Analyse von Registerdaten mit vervollständigten Angaben

Faktor	χ^2 -Wert	Freiheitsgrade	p-Wert
Geschlecht	28.42	1	<.0001
Tumordicke	213.03	3	<.0001
Tumordicke nichtlinear	12.36	2	0.0021
Ulzeration	5.55	1	0.0184
Regression	0.29	1	0.5919
Histologischer Typ	7.96	4	0.0930
Level	4.12	3	0.2487
Nonlinear	3.48	2	0.1754
Lokalisation	14.17	1	0.0002
Alter	2.16	1	0.1418
Total nichtlinear	23.27	4	0.0001
Total	520.25	15	<.0001

3.3.5 Künstliche neuronale Netze

Auch bei dieser Analyse wurden wieder zwei künstliche neuronale Netze verwendet. Das erste mit zwei Neuronen in der verdeckten Schicht, das zweite mit zehn Neuronen in der verdeckten Schicht. Auf eine Darstellung der Netzparameter wird verzichtet, da sich im Vergleich zur vorhergehenden Analyse keine neuen Resultate ergaben.

3.3.6 Vergleich der Ergebnisse bei den vervollständigten Registerdaten

Die Hinzunahme der Fälle mit fehlenden Angaben und die Ersetzung der fehlende Werte hat nur zu unwesentlichen Veränderungen geführt.

Die Tumordicke blieb der mit Abstand wichtigste prognostische Faktor. Die Hinzunahme eines anderen Faktors brachte, unabhängig von der Wahl des Modells, keine deutlich sichtbaren Unterschiede.

Die Verwendung des komplexen künstlichen neuronalen Netzes brachte eine sehr gute Anpassung an die Trainingsdaten. Bei der Überprüfung anhand der Testdaten fiel das Modell in der Qualität der Prognose stark ab, so daß von einer erheblichen Überanpassung des Modells ausgegangen werden muß. Die anderen Modelle wiesen alle sehr ähnliche Ergebnisse beim Vergleich von Trainings- und Testdaten auf.

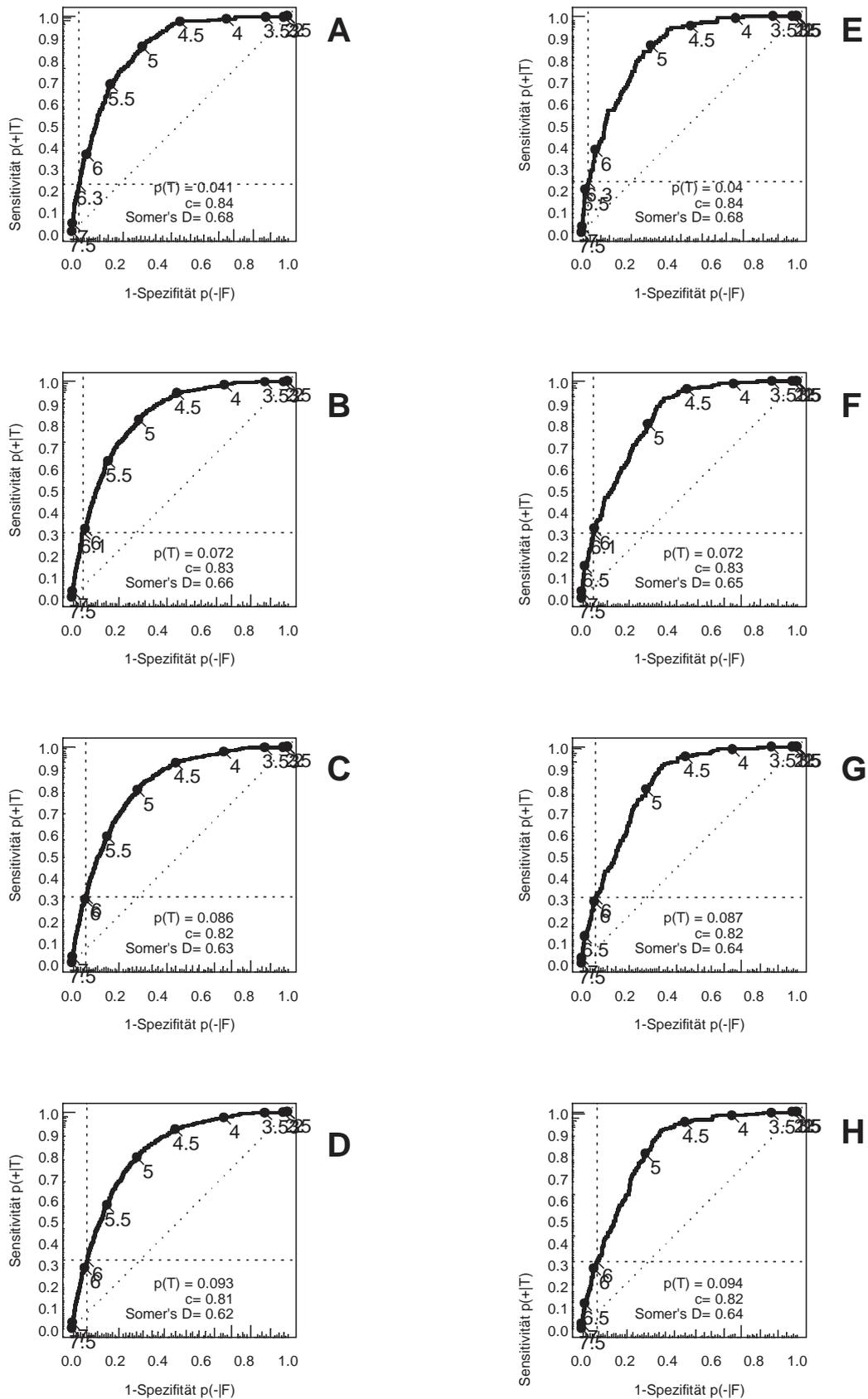


Abbildung 46: ROC-Kurven für Prognoseschätzung bei vervollständigten Registerdaten mit der Tumordicke

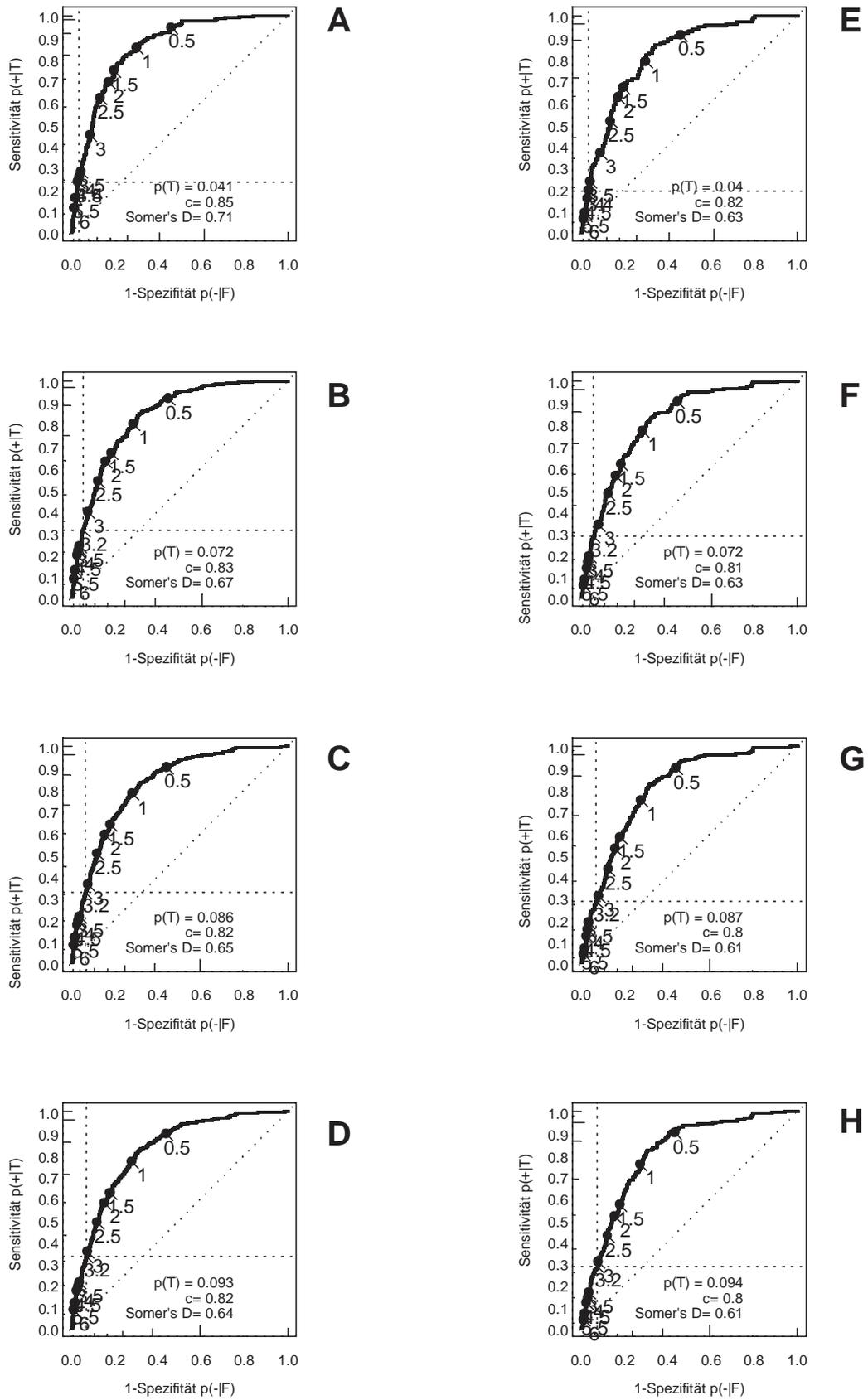


Abbildung 47: ROC-Kurven für Prognoseschätzung bei vervollständigten Registerdaten mit einem CART-Baum

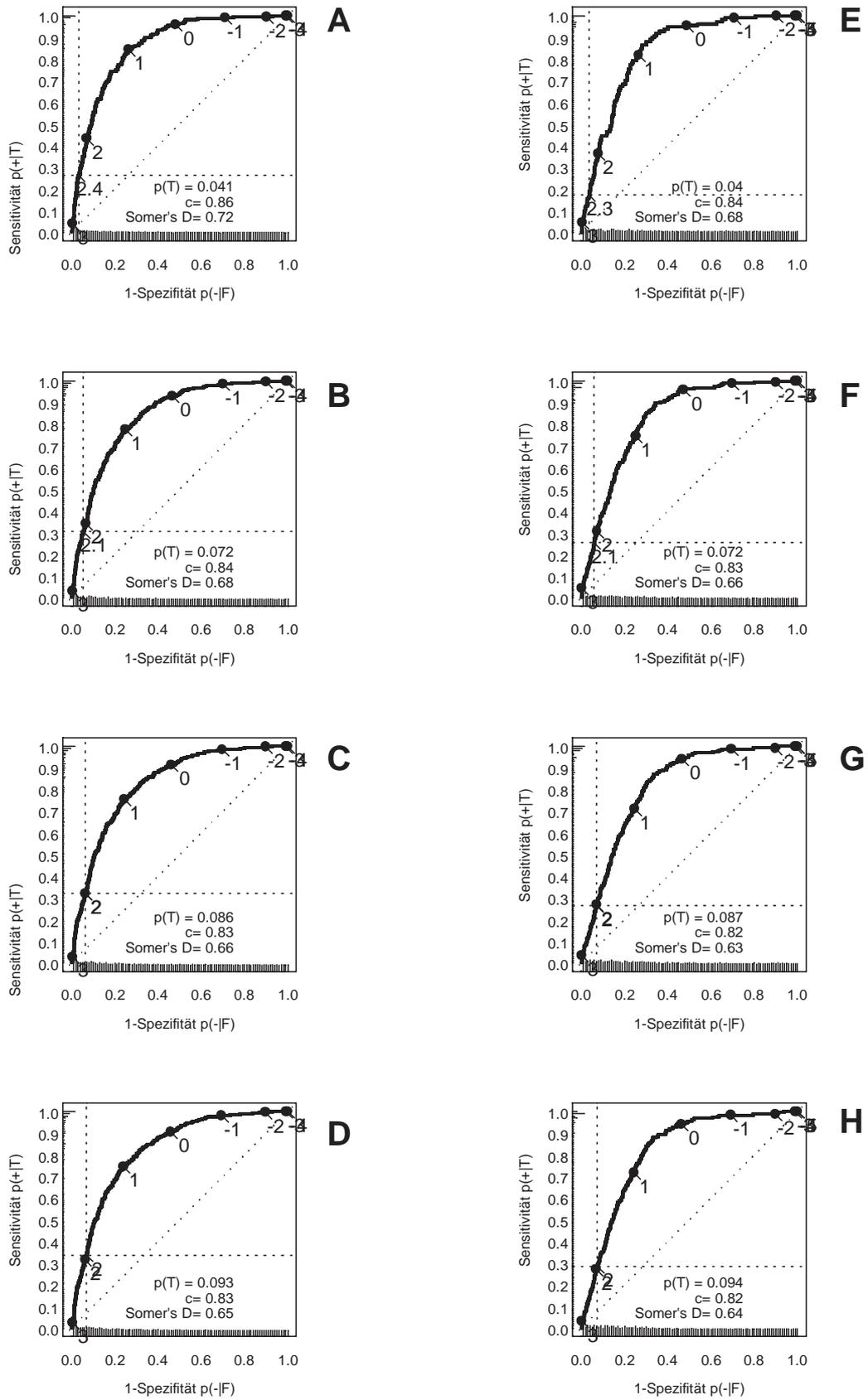


Abbildung 48: ROC-Kurven für Prognoseschätzung bei vervollständigten Registerdaten mit einem Cox-Modell

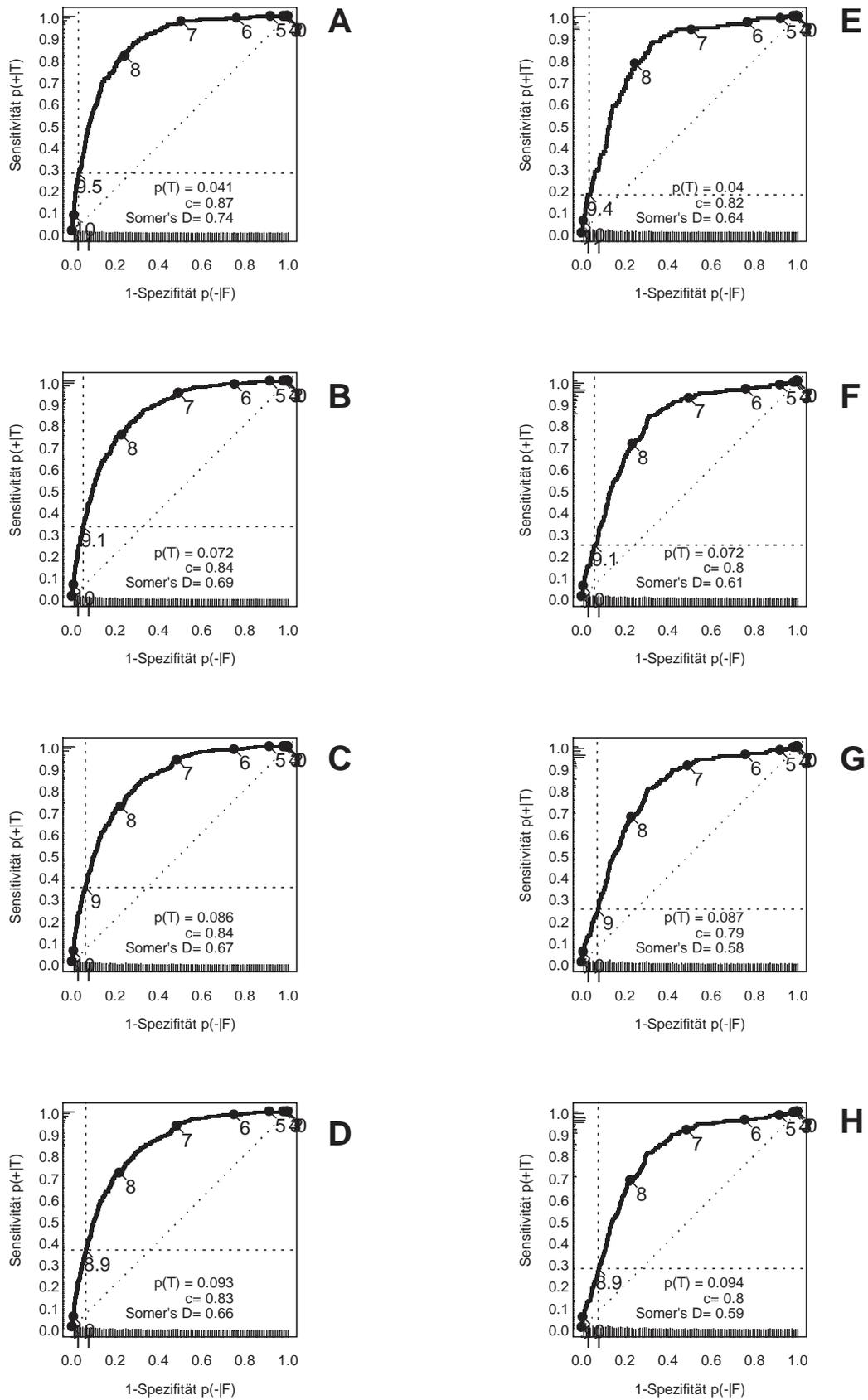


Abbildung 49: ROC-Kurven für Prognoseschätzung bei vervollständigten Registerdaten mit einem künstlichen neuronalen Netz

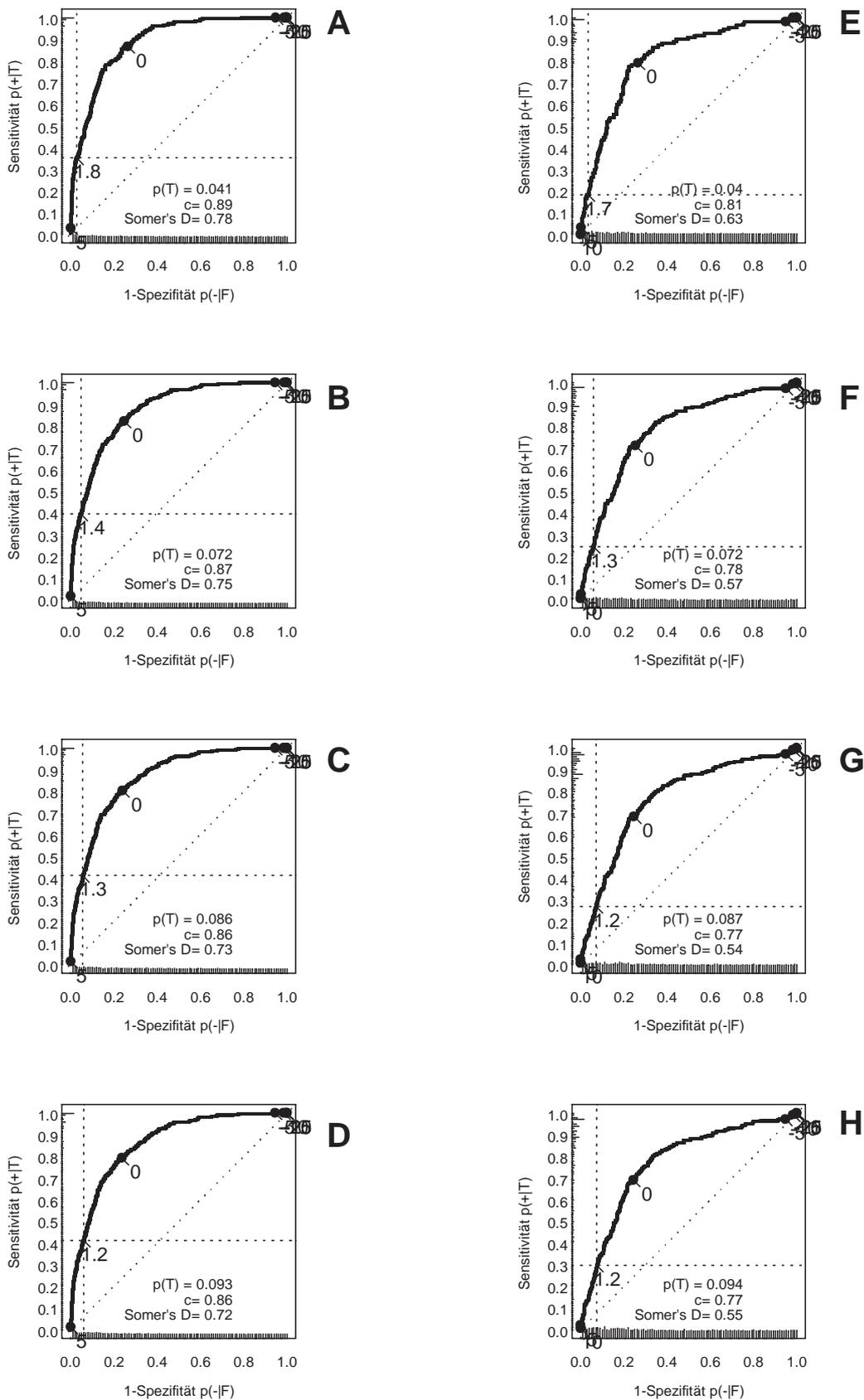


Abbildung 50: ROC-Kurven für Prognoseschätzung bei vervollständigten Registerdaten mit einem komplexen künstlichen neuronalen Netz

Bei den Trainingsdaten lieferte das komplexere künstliche neuronale Netz die beste Anpassung an die Daten. Die schlechteste Anpassung erfolgte durch die alleinige Verwendung der Tumordicke und durch das CART-Modell.

Tabelle 24: Rang-Korrelationen für die Trainingsdaten und die verschiedenen Modelle bei den vervollständigten Registerdaten (5460 Fälle, davon 506 nicht zensiert, 3416962 relevante Paare)

Gruppe	C	Index	S.D.	Konkordante Paare
Tumordicke	0,81	0,62	0,017	2766946
CART	0,82	0,63	0,017	2792147
COX-PH	0,82	0,65	0,017	2817385
KNN	0,83	0,67	0,016	2844958
KNN2	0,86	0,72	0,014	2939895

Bei Überprüfung der Modelle mit den Testdaten ist eine deutliche Überanpassung durch das komplexere künstliche neuronale Netz zu erkennen. Die besten Prognoseeigenschaften bei den Testdaten lieferte das COX-PH-Modell. Die alleinige Verwendung der Tumordicke führt aber nur zu einem sehr kleinen Unterschied gegenüber dem COX-PH-Modell.

Tabelle 25: Rang-Korrelationen für die Testdaten und die verschiedenen Modelle bei den vervollständigten Registerdaten (2743 Fälle, davon 259 nicht zensiert, 828370 relevante Paare)

Gruppe	C	Index	S.D.	Konkordante Paare
Tumordicke	0,82	0,64	0,021	678422
CART	0,80	0,61	0,025	665475
COX-PH	0,82	0,64	0,022	680655
KNN	0,80	0,60	0,027	661343
KNN2	0,78	0,56	0,029	645503