

Entwicklung eines neuen prognostischen Index für Patienten mit fortgeschrittenem Mantelzell-Lymphom

Eva Hoster

2008

**Aus dem Institut für Medizinische
Informationsverarbeitung, Biometrie und
Epidemiologie der Ludwig-Maximilians-Universität
München**

Direktor: Prof. Dr. rer. nat. U. Mansmann

**Entwicklung eines neuen prognostischen Index für
Patienten mit fortgeschrittenem
Mantelzell-Lymphom**

Dissertation

zum Erwerb des Doktorgrades der Humanbiologie
an der Medizinischen Fakultät
der Ludwig-Maximilians-Universität zu München

vorgelegt von

Eva Hoster

aus München

2008

Mit Genehmigung der Medizinischen Fakultät
der Universität München

Berichterstatter: Prof. Dr. med. Joerg Hasford
2. Berichterstatter: Prof. Dr. Stefan K. Bohlander

Mitberichterstatter: Priv. Doz. Dr. Dr. Fuat Oduncu
Priv. Doz. Dr. Simon Rothenfuß

Mitbetreuung durch den
promovierten Mitarbeiter:

Dekan: Prof. Dr. med. Dr. h.c. M. Reiser, FACR

Tag der mündlichen Prüfung: 10.11.2008

Inhaltsverzeichnis

Zusammenfassung	4
Danksagung	6
1 Hintergrund und Fragestellung	7
1.1 Das Mantelzell-Lymphom	7
1.1.1 Diagnose	7
1.1.2 Epidemiologie und Symptome	8
1.1.3 Therapie	11
1.1.4 Prognose	11
1.2 Prognosemodelle bei Lymphomen	11
1.2.1 Ann-Arbor-Stadium	12
1.2.2 International Prognostic Index	13
1.2.3 Follicular Lymphoma International Prognostic Index	16
1.2.4 Prognosefaktoren beim Mantelzell-Lymphom	18
1.3 Fragestellung und Ziele der Arbeit	20
2 Daten und statistische Methoden	21
2.1 Datenbasis	21
2.1.1 Studien der GLSG und des European MCL Network	21
2.1.2 Patienten	25
2.1.3 Baselinecharakteristika	25
2.1.4 Zielgrößen	26
2.2 Beschreibung des Patientenkollektivs	26
2.3 Validierung der bekannten Prognosemodelle	27
2.4 Entwicklung eines neuen prognostischen Index	27
2.4.1 Potenzielle Prognosefaktoren	27
2.4.2 Regressionsmodell	27
2.4.3 Prognostischer Index	28
2.4.4 Vereinfachter prognostischer Index	31
2.5 Schätzung der Modellgüte	33

2.5.1	Maßzahlen der Modellgüte	34
2.5.2	Das Bootstrap-Verfahren	37
2.6	Überprüfung der Modellannahmen	42
2.6.1	Proportional-Hazards-Annahme	42
2.6.2	Linearität	43
2.6.3	Extrembeobachtungen	43
2.7	Weitere explorative Analysen	43
2.7.1	Effekt der Auswahl vollständiger Datensätze	43
2.7.2	Wechselwirkungen	44
2.7.3	Alter und Allgemeinzustand	44
2.7.4	Zeit bis zum Therapieversagen	44
3	Ergebnisse	45
3.1	Beschreibung des Patientenkollektivs	45
3.1.1	Patientenzahl	45
3.1.2	Baselinecharakteristika	45
3.1.3	Therapie und Therapieerfolg	50
3.1.4	Gesamtüberleben	51
3.2	Validierung der bekannten Prognosemodelle	51
3.2.1	Ann-Arbor-Stadium	51
3.2.2	International Prognostic Index	52
3.2.3	Follicular Lymphoma International Prognostic Index	54
3.2.4	Vergleich und Bewertung	55
3.3	Entwicklung eines neuen prognostischen Index	57
3.3.1	Potenzielle Prognosefaktoren	57
3.3.2	Regressionsmodell	62
3.3.3	Prognostischer Index (MIPI)	65
3.3.4	Vereinfachter prognostischer Index	68
3.4	Schätzung der Modellgüte	73
3.4.1	Regressionsmodell	73
3.4.2	Prognostischer Index	73
3.5	Überprüfung der Modellannahmen	74
3.5.1	Proportional-Hazards-Annahme	74
3.5.2	Linearität	76
3.5.3	Extrembeobachtungen	76
3.6	Weitere explorative Analysen	78
3.6.1	Effekt der Auswahl vollständiger Datensätze	78
3.6.2	Wechselwirkungen	81
3.6.3	Alter und Allgemeinzustand	82

3.6.4	Zeit bis zum Therapieversagen	83
4	Diskussion	85
4.1	Kurzzusammenfassung der Ergebnisse	85
4.2	Vergleich mit publizierten Ergebnissen	86
4.2.1	Ergebnisse zum IPI	86
4.2.2	Ergebnisse zum FLIPI	87
4.2.3	Prognosefaktoren	88
4.3	Mögliche Limitationen	89
4.3.1	Patientenkollektiv und Verallgemeinerbarkeit	89
4.3.2	Methodik	91
4.3.3	Mögliche Confounder	93
4.4	Schlussfolgerungen und Ausblick	94
	Literaturverzeichnis	95

Zusammenfassung

Zielsetzung: Das Mantelzell-Lymphom nimmt unter den Non-Hodgkin-Lymphomen mit einem medianen Gesamtüberleben von unter fünf Jahren einen besonders ungünstigen Verlauf. Für Patienten mit Mantelzell-Lymphom gibt es bislang keinen prognostischen Index, mit dessen Hilfe man Patienten auf Grund von bei Diagnosestellung erhobenen Parametern in Gruppen unterschiedlich schweren Krankheitsverlaufs einteilen könnte. Der für Patienten mit Mantelzell-Lymphom meist verwendete International Prognostic Index (IPI) und der Follicular Lymphoma International Prognostic Index (FLIPI) beruhen auf Daten von Patienten mit diffus-großzelligem bzw. follikulärem Lymphom, zu denen das Mantelzell-Lymphom nicht zählt. Darüber hinaus geben die publizierten Ergebnisse zu Prognosefaktoren beim Mantelzell-Lymphom ein unklares Bild und basieren überwiegend auf kleinen Patientenzahlen von unter 130. Im Rahmen dieser Arbeit wurden die Daten von 455 Patienten mit fortgeschrittenem Mantelzell-Lymphom aus drei Primärtherapiestudien der German Low-Grade Lymphoma Study Group (GLSG) und des European Mantle Cell Lymphoma Network analysiert, um die prognostische Relevanz von IPI, FLIPI und weiterer potenzieller Prognosefaktoren beim Mantelzell-Lymphom zu klären und gegebenenfalls einen neuen prognostischen Index zu entwickeln.

Methoden: Zielgröße war das Gesamtüberleben von der Studienaufnahme bis zum Tod jeglicher Ursache. Potenzielle Prognosefaktoren waren die folgenden bei Diagnosestellung und vor Therapiebeginn erhobenen Parameter: Alter, Geschlecht, Allgemeinzustand, Tumorstadium nach Ann Arbor, B-Symptomatik, Anzahl extranodaler Befälle, Anzahl befallener Lymphknotenregionen, Knochenmarksbefall, maximale Lymphknotengröße, Milzbefall, Laktat-Dehydrogenase-Aktivität (LDH), Leukozyten-, Lymphozyten-, Granulozyten-, Monozyten-, und Thrombozytenzahl, Hämoglobin, Albumin, β_2 -Mikroglobulin, sowie die Immunglobuline IgM, IgG und IgA. Die Überprüfung der prognostischen Relevanz von IPI, FLIPI und der weiteren Parameter erfolgte mit der Kaplan-Meier-Schätzung, dem Logrank-Test bzw. der Cox-Regression. Die Identifizierung unabhängiger Prognosefaktoren erfolgte mit Rückwärts-Selektion im Rahmen einer multiplen Cox-Regression. Für den prognostischen Index wurden durch Kategorisierung des Prognosescores aus dem endgültigen Regressionsmodell Risikogruppen nach der „Minimum-p-Value“-Methode für die Logrank-Statistik definiert. Die interne Validierung erfolgte mit dem Bootstrap-Verfahren.

Ergebnisse: Der IPI klassifizierte mehr als zwei Drittel der Patienten in die beiden Gruppen mittleren Risikos, deren Überlebenskurven schlecht getrennt waren. Die Hochrisikogruppe war mit unter 10% der Patienten sehr klein. Der FLIPI klassifizierte 6% der Patienten als Low Risk, die Überlebenskurven der Low und Intermediate Risk Gruppen verliefen nicht getrennt und die High Risk Gruppe umfasste knapp zwei Drittel der Patienten bei relativ günstigem Verlauf. Von den potenziellen Prognosefaktoren zeigten Alter, Allgemeinzustand, B-Symptome, Milzbefall, maximale Lymphknotengröße, LDH, Hämoglobin, Leukozyten-, Lymphozyten-, Granulozyten- und Monozytenzahl, sowie β_2 -Mikroglobulin univariaten Einfluss auf das Gesamtüberleben. Dagegen hatten Geschlecht, Stadium, Knochenmarksbefall, Anzahl extranodaler Befälle, Anzahl befallener Lymphknotenregionen, Thrombozytenzahl, Albumin, sowie die Immunglobuline IgM, IgG und IgA keinen univariaten Einfluss auf das Gesamtüberleben. Mit den vier neu identifizierten unabhängigen Prognosefaktoren Alter, Allgemeinzustand, LDH und Leukozytenzahl konnte ein neuer prognostischer Index (Mantle Cell Lymphoma International Prognostic Index, MIPI) definiert werden. Der MIPI teilt die Patienten in drei Risikogruppen mit unterschiedlichem Gesamtüberleben ein, eine Low Risk Gruppe (44% der Patienten, medianes Gesamtüberleben nicht erreicht), eine Intermediate Risk (35%, medianes Gesamtüberleben 51 Monate) und eine High Risk Gruppe (21%, medianes Gesamtüberleben 29 Monate). Die Bootstrap-Validierung zeigte die Stabilität des Prognosemodells und bestätigte die signifikante Auftrennung der Überlebenskurven nach dem neuen prognostischen Index.

Schlussfolgerung: Die nur mäßige prognostische Relevanz von IPI und FLIPI bei Patienten mit fortgeschrittenem Mantelzell-Lymphom unterstrich die mangelnde Übertragbarkeit von Ergebnissen einzelner Lymphomentitäten auf andere, sowie die Notwendigkeit eigener Prognosestudien für die einzelnen biologischen Entitäten. Im Rahmen dieser Arbeit konnte der erste prognostische Index speziell für Patienten mit fortgeschrittenem Mantelzell-Lymphom unter Verwendung des bisher größten publizierten Kollektivs entwickelt werden. Der neue Index definierte drei im Gesamtüberleben gut getrennte Risikogruppen auf der Basis von vier in der Praxis leicht verfügbaren klinischen Parametern. Die Ergebnisse dieser Arbeit können die Beurteilung der Vergleichbarkeit von Studienergebnissen erleichtern und in Beobachtungsstudien risikoadjustierte Analysen sowie in klinischen Studien die stratifizierte Randomisierung ermöglichen. Darüber hinaus könnten sie zum Entscheidungsprozess im Rahmen einer risikostratifizierten Therapie dieser schweren Erkrankung beitragen.

Danksagung

Bei allen, die mich während der Zeit meiner Promotion gefördert, unterstützt und ermutigt haben, möchte ich mich an dieser Stelle von ganzem Herzen bedanken.

Mein allererster Dank geht an Herrn Prof. Dr. Joerg Hasford, der meine Begeisterung für die medizinische Biometrie weckte und förderte, der sich stets Zeit nahm, meine vielen Fragen ausführlich und fundiert zu beantworten, und mir mit seiner konstruktiven Kritik jederzeit wertvolle Hilfestellung gab.

Herrn Prof. Dr. Wolfgang Hiddemann danke ich sehr herzlich dafür, mich von Anfang an in die German Low-Grade Lymphoma Study Group und das European MCL Network als Biometrikerin eingebunden zu haben. Darüber hinaus bin ich sehr dankbar für seine stete Unterstützung bei der Durchführung und Präsentation meiner Arbeit.

Herrn Dr. Michael Unterhalt danke ich sehr herzlich für die geduldige Beantwortung meiner vielen Fragen, für die ausführliche und bereichernde Diskussion meiner Ergebnisse und die kontinuierliche Ermutigung bei meiner Arbeit.

Bei Herrn Prof. Dr. Martin Dreyling bedanke ich mich für die Unterstützung bei der Publikation meiner Arbeit sowie für die Möglichkeit, meine Ergebnisse im Rahmen des European MCL Network zu präsentieren.

Allen Mitgliedern der German Low-Grade Lymphoma Study Group (GLSG) und des European MCL Network danke ich für die Patientenrekrutierung und die Dokumentation im Rahmen der Studien. Durch diese gemeinsame Anstrengung konnte der bislang größte Datensatz von Patienten mit Mantelzell-Lymphom gewonnen werden, ohne den diese Arbeit nicht möglich gewesen wäre. Den Teilnehmern des European MCL Meeting in Kopenhagen 2006 danke ich sehr herzlich für die ausgiebige und sehr fruchtbare Diskussion meiner Ergebnisse.

Meiner Kollegin Frau Monika Volk danke ich sehr herzlich für ihre freundliche Unterstützung und Begleitung.

Meiner Kollegin Marietta Rottenkolber danke ich herzlich dafür, dass sie mich während meiner Elternzeit vertreten hat, und mir dadurch den Rücken für die Fertigstellung meiner Arbeit freihielt.

Meinem Mann Dr. Matthias Hoster danke ich von ganzem Herzen für seinen Zuspruch, seine Geduld und sein Verständnis. Ganz besonders danke ich ihm dafür, dass er mich auch nach der Geburt unserer Tochter kräftig dabei unterstützte, meine Arbeit fertigzustellen.

Kapitel 1

Hintergrund und Fragestellung

1.1 Das Mantelzell-Lymphom

1.1.1 Diagnose

Maligne Lymphome sind bösartige Tumoren, die von lymphatischen Geweben oder Zellen ausgehen. Zu den lymphatischen Geweben zählen die Lymphknoten, der Thymus, die Milz, die Mandeln, sowie die lymphatischen Gewebe der Schleimhäute. Die lymphatischen Zellen, von denen maligne Lymphome ausgehen, sind Zellen des spezifischen Abwehrsystems und umfassen B- und T-Lymphozyten sowie die natürlichen Killer (NK)-Zellen [1]. Die Klassifikation der malignen Lymphome war lange Zeit uneinheitlich und schwierig [2]. Während in Europa seit den siebziger Jahren die Kiel-Klassifikation [3],[4] verwendet wurde, galt in den USA seit den achtziger Jahren die Working Formulation [5]. Anfang der neunziger Jahren entwickelte ein internationales Gremium erstmals eine international gültige Klassifikation, die Revised European-American Classification of Lymphoid Neoplasms (REAL-Klassifikation [2]) von 1994, welche die zur damaligen Zeit bestehenden Klassifikationssysteme zusammenführte und ablöste. Die REAL-Klassifikation wurde 1997 zur WHO-Klassifikation [6] modifiziert und in der aktuell gültigen ICD-O-3 Klassifikation [7] der WHO fortgeführt.

Die malignen Lymphome werden in das Hodgkin-Lymphom und die heterogene Gruppe der Non-Hodgkin-Lymphome (NHL) unterteilt, die wiederum in B-, T- und NK-Zell-Lymphome unterschieden werden. Die Abgrenzung der einzelnen Entitäten von NHL stützt sich auf morphologische, immunphänotypische und genetische Charakteristika der Tumorzellen und -gewebe. Das Mantelzell-Lymphom (ICD-O-3: 9673/3) zählt zu den B-Zell-NHL und wurde bereits in der Kiel-Klassifikation unter dem Namen „zentrozytisches Lymphom“ charakterisiert. Erst die REAL-Klassifikation [2] von 1994 etablierte das Mantelzell-Lymphom unter diesem Namen als eigenständige Entität des NHL auf internationaler Ebene. Der Name weist auf die Mantelzone des Lymphfollikels hin, nach deren Wachstumsmuster sich die Tumorzellen häufig anordnen. Die Tumorzellen sind typi-

scherweise kleine bis mittlere lymphoide Zellen mit atypischen Zellkernen. In der Mehrzahl der Fälle liegt bei den Tumorzellen eine erworbene chromosomale Translokation $t(11;14)$ vor, die zu einer Überexpression des normalerweise nicht exprimierten Proteins Cyclin-D1 führt [2]. Abzugrenzen ist das Mantelzell-Lymphom unter anderem vom folliculären Lymphom (ICD-O-3: 9690/3, Grad 1 bis 3: 9695/3, 9691/3, 9698/3), das in der Regel von den Follikelzentrumszellen des Lymphfollikels ausgeht [2]. Immunhistochemisch erlaubt das Fehlen CD23-positiver Zellen eine Abgrenzung von der chronischen lymphatischen Leukämie und das Vorhandensein CD5-positiver Zellen vom folliculären und dem Marginalzonenlymphom.

Häufigstes Symptom, das zur Entdeckung eines malignen Lymphoms führt, ist eine Lymphknotenvergrößerung, manchmal auch eine Vergrößerung der Milz. Mit bildgebenden Verfahren werden möglichst alle befallenen Lymphknotenstationen sichtbar gemacht, um den Ausbreitungsgrad des Lymphoms zu bestimmen. Die endgültige Diagnose wird dann ausschließlich durch Entnahme und histopathologische Charakterisierung von Tumorgewebe, meist von Lymphknoten oder Knochenmark, gestellt.

1.1.2 Epidemiologie und Symptome

Aus vier bevölkerungsbasierten Studien [8], [9], [10], [11] ließ sich die jährliche Inzidenzrate des Mantelzell-Lymphoms und der Anteil an allen NHL ermitteln (Tabelle 1.1). Eine bevölkerungsbasierte Studie aus Dänemark [8] berichtete eine jährliche Neuerkrankungsrate von 0,42 pro 100.000 Einwohner, 0,70 pro 100.000 Männer und 0,20 pro 100.000 Frauen. Aus einer niederländischen Studie [11] ließ sich eine jährliche Neuerkrankungsrate von ca. 0,30 pro 100.000 Einwohner schätzen. Für British Columbia, Kanada [10], konnte eine jährliche Inzidenzrate von ca. 0,32 pro 100.000 Einwohner ermittelt werden. Aus den Daten des Tumorzentrums München [12] von 2004 ergab sich eine Zahl von ca. 0,40 Neuerkrankungen pro Jahr und 100.000 Einwohner im Raum München und Umgebung. Der Anteil der Mantelzell-Lymphome an allen Non-Hodgkin-Lymphomen lag in den genannten Arbeiten zwischen 3,5% und 4,5%. In einer internationalen retrospektiven Analyse [13] betrug der Anteil der neu diagnostizierten Patienten mit Mantelzell-Lymphom 6% an allen NHL (83 von 1378 Patienten). Das Mantelzell-Lymphom stellt somit eine relativ seltene Tumorentität dar.

Das mediane Erkrankungsalter wurde in den bevölkerungsbasierten Publikationen zwischen 65 und 68 Jahren angegeben (Tabelle 1.1). Mit einem Anteil zwischen 61% und 75% erkrankten mehr Männer als Frauen. Bei Diagnose waren 20% bis 28% der Patienten arbeitsunfähig oder bettlägerig (ECOG Performance Status 2 - 4, s. Abschnitt 2.1.3). Der Ausbreitungsgrad des Lymphoms wird mittels der Stadieneinteilung nach Ann Arbor (s. Abschnitt 1.2.1) festgestellt. Die meisten Patienten befanden sich bei Diagnosestellung bereits in einem fortgeschrittenem Ann-Arbor-Stadium III oder IV. In den bevölkerungs-

bezogenen Studien lag der Anteil der Patienten mit den lokal begrenzten Stadien I und II zwischen 14% und 20%, der Anteil der Patienten mit Stadium IV zwischen 72% und 80%.

Der hohe Anteil von Patienten im Stadium IV spiegelte den hohen Anteil von Patienten mit Knochenmarksbefall zwischen 63% und 81% wider. Die Milz war in [8] bei 36% der Patienten befallen, der Gastrointestinaltrakt bei 10% bzw. 12% der Patienten. 49% bzw. 24% hatten B-Symptomatik (Gewichtsverlust, Fieber oder Nachtschweiß, s. Abschnitt 1.2.1), und die Laktat-Dehydrogenase-Aktivität (LDH) war in 21% - 55% der Patienten erhöht.

Erstautor Referenz	Andersen [8]	Møller [9]	Argatoff [10]	Velders [11]
Land	Dänemark	Dänemark	Kanada	Niederlande
Patientenzahl	105	93	80	41
Zeitraum	1992 - 2000	1992 - 1998	1988 - 1995	1981 - 1989
jährl. Inzidenzrate /10 ⁵	0,42	ca. 0,52	ca. 0,32	ca. 0,30
Anteil an NHL	3,5%	4,5%	4% der ML	3,7%
medianes Alter (Jahre)	66	68	65	68
Range	38 - 89	38 - 92	39 - 86	39 - 85
männlich	75%	71%	70%	61%
ECOG 2 - 4	28%	25%	21%	20%
Stadium I+II	16%	17%	14%	20%
Stadium IV	80%	—	72%	78%
Knochenmarksbefall	72%	—	63%	81%
Milzbefall	36%	—	—	—
gastrointestinaler Befall	12%	—	10%	—
Bulk (> 5cm)	—	—	—	42%
B-Symptome	49%	—	24%	—
erhöhte LDH	21%	25%	39%	55%
CR-Rate	43%	42%	—	32%
mediane CRD	—	38 Monate	—	25 Monate
medianes PFS	15 Monate	—	—	—
medianes OS	30 Monate	37 Monate	43 Monate	31,5 Monate

Tabelle 1.1: Epidemiologische und klinische Daten von Patienten mit Mantelzell-Lymphom aus vier bevölkerungsbezogenen Studien. ML: Malignes Lymphom, —: keine Angabe, Bulk: größter Lymphknoten > 5cm.

Neun weitere institutionsbezogene Studien [14], [15], [16], [17], [18], [19], [20], [21], [22] konnten zur Beschreibung der klinischen Charakteristika bei Diagnose herangezogen werden (Tabelle 1.2), wobei hier die heterogene Auswahl der Patienten beachtet werden muss. Mit medianem Alter bei Diagnose zwischen 59 und 66 Jahren waren in diesen Studien die Patienten etwas jünger und meist in einem geringeren Anteil zwischen 14% und 20% arbeitsunfähig oder bettlägerig (ECOG Performance Status 2 - 4). Die Verteilung der Patienten zu den Stadien und der Anteil von Patienten mit Knochenmarksbefall waren ähnlich wie in den bevölkerungsbezogenen Studien. Die Milz war bei ca. 30% - 40% befallen, der Gastrointestinaltrakt bei ca. 20% der Patienten und ca. 20% der Patienten wiesen einen größten Lymphknoten von über 10cm im Durchmesser (Bulk) auf. Der Anteil von Patienten mit B-Symptomatik lag meist zwischen 20% und 40%, und der Anteil von Patienten mit erhöhter LDH meist zwischen 30% und 40%.

Erstautor	Tiemann	Räty	Samaha	Oinonen	Schrader
Referenz	[14]	[15]	[16]	[17]	[18]
Land	Deutschland	Finnland	Frankreich	Finnland	Deutschland
Patientenzahl	304	127	121	94	95
Zentrenzahl	Multicenter	Single center	Single center	Single center	Multicenter
Zeitraum	1972 - 1994	1980 - 2000	1979 - 1997	1980 - 1996	1975 - 1985
medianes Alter (Jahre)	64	65	63	66	62% > 60
Range	27 - 86	30 - 90	32 - 84	44 - 87	—
männlich	76%	64%	67%	59%	77%
ECOG 2 - 4	16%	15%	19%	14%	24%
Stadium I+II	8%	17%	13%	24%	8%
Stadium IV	—	—	83%	68%	—
Knochenmarksbefall	72%	63%	79%	61%	63%
Milzbefall	—	41%	47%	31%	—
gastrointestinaler Befall	—	21%	—	19%	—
Bulk (> 10cm)	—	21%	20%	20%	—
B-Symptome	41%	35%	14%	35%	42%
erhöhte LDH	29%	46%	40%	38%	16%
CR-Rate	—	45%	28%	34%	—
mediane CRD	—	16 Monate	—	18 Monate	—
medianes PFS	—	18 Monate	11 Monate	28 Monate	—
medianes OS	ca. 36 Monate	34 Monate	37 Monate	41 Monate	ca. 30 Monate

Erstautor	Weisenburger	Zucca	Bosch	Decaudin
Referenz	[19]	[20]	[21]	[22]
Land	USA	Schweiz	Spanien	Frankreich
Patientenzahl	68	65	59	45
Zentrenzahl	Multicenter	Multicenter	Multicenter	Single center
Zeitraum	1985 - 1996	1979 - 1993	1983 - 1995	1978 - 1995
medianes Alter (Jahre)	64	64	63	59
Range	40 -	27 - 85	39 - 83	29 - 76
männlich	75%	67%	74%	78%
ECOG 2 - 4	19%	20%	51%	2%
Stadium I+II	25%	22%	5%	13%
Stadium IV	—	72%	86%	69%
Knochenmarksbefall	60%	58%	81%	67%
Milzbefall	—	35%	44%	27%
gastrointestinaler Befall	—	—	17%	—
Bulk (> 10cm)	—	—	5%	20%
B-Symptome	32%	31%	25%	38%
erhöhte LDH	40%	30%	40%	16%
CR-Rate	42%	51%	19%	58%
mediane CRD	—	44 Monate	17 Monate	ca. 24 Monate
medianes PFS	12 Monate	—	—	15 Monate
medianes OS	38 Monate	42 Monate	49 Monate	ca. 56 Monate

Tabelle 1.2: Klinische Daten von Patienten mit Mantelzell-Lymphom aus neun instituti-
onsbezogenen Studien. ML: Malignes Lymphom, —: keine Angabe, Bulk: größter Lymph-
knotendurchmesser > 10cm.

1.1.3 Therapie

In den untersuchten Studien erhielt der größte Teil der Patienten Chemotherapie-Kombinationen mit Anthrazyklinen oder Mitoxantron. Wenige Patienten erhielten darüber hinaus bei Ansprechen auf die Chemotherapie eine autologe oder allogene Blutstammzell-Transplantation. Ein Teil der Patienten erhielt Chemotherapien ohne Anthrazykline. In Einzelfällen, meist bei den lokal begrenzten Stadien I oder II, wurde der Tumor bestrahlt und/oder operativ entfernt.

Die therapeutisch erzielten Raten kompletter Remissionen (CR, vollständiges Verschwinden aller vor Therapie bestehenden Krankheitszeichen) lag zwischen 19% und 58% (Median über alle Studien 42%). Die mediane Zeit in kompletter Remission (CRD) lag zwischen 16 und 44 Monaten (Median 21 Monate). Die mediane Zeit von Diagnose bis zur ersten Progression oder Tod (progressionsfreies Überleben, PFS) wurde zwischen 11 und 28 Monaten angegeben (Median 18 Monate).

1.1.4 Prognose

Das Mantelzell-Lymphom hat unter allen NHL die schlechteste Prognose [13]. Die publizierten Angaben zum medianen Gesamtüberleben (Zeit von Diagnose bis zum Tod jeder Ursache, OS) lagen zwischen 30 und 43 Monaten in den bevölkerungsbasierten Studien, und zwischen 30 und ca. 56 Monaten in allen untersuchten Studien (Median 37 Monate).

1.2 Prognosemodelle bei Lymphomen

Der klinische Verlauf der Lymphomkrankung kann von Patient zu Patient sehr unterschiedlich sein. Diese interindividuelle Variabilität kann mit bestimmten Charakteristika des Patienten oder dessen Erkrankung assoziiert sein, die man dann als Prognosefaktoren bezeichnet. Ein Teil der Variabilität wird letztlich nicht erklärbar bleiben, man spricht von zufälliger oder nicht-deterministischer Variabilität. Prognosemodelle dienen unter anderem dazu, die Abhängigkeit des Krankheitsverlauf von messbaren Parametern zu identifizieren und quantifizieren. Der folgende Abschnitt gibt eine Übersicht über die aktuell bei Lymphomen verwendeten Prognosemodelle sowie deren Relevanz für Patienten mit Mantelzell-Lymphom auf der Basis publizierter Ergebnisse. Zielgröße, auf die sich das Wort „Prognose“ bezieht, ist hierbei immer das Gesamtüberleben, das heißt die Zeit von Diagnose bis zum Tod jeglicher Ursache.

1.2.1 Ann-Arbor-Stadium

Beschreibung

Wie für viele Tumorerkrankungen existiert auch für die Lymphome eine Einteilung nach dem Ausbreitungsstadium bei Diagnosestellung. Die 1971 formulierte Ann-Arbor-Klassifikation [23] ist die aktuell gültige und international anerkannte Stadieneinteilung für alle Lymphome [24]. Sie wurde ursprünglich für das Hodgkin-Lymphom entwickelt und unterscheidet die folgenden vier Stadien unter Berücksichtigung der Anzahl befallener Lymphknotenregionen oder nichtlymphatischer Organe und Gewebe sowie deren Lokalisation in Bezug auf das Zwerchfell.

Stadium I: Befall einer einzigen Lymphknotenregion (I) oder lokalisierter Befall eines einzigen nichtlymphatischen Organs oder Gewebes (I_E)

Stadium II: Befall von 2 oder mehr Lymphknotenregionen auf der gleichen Seite des Zwerchfells (II) oder lokalisierter Befall extralymphatischer Organe oder Gewebe und Befall von einer oder mehrerer Lymphknotenregionen auf der gleichen Seite des Zwerchfells (II_E)

Stadium III: Befall von Lymphknotenregionen auf beiden Seiten des Zwerchfells (III), welcher ebenfalls von lokalisiertem extralymphatischem Organ- und Gewebefall (III_E) oder Milzbefall (III_S) oder beidem (III_{ES}) begleitet sein kann

Stadium IV: Diffuser oder disseminierter Befall von einem oder mehrerer extralymphatischer Organe oder Gewebe mit oder ohne Lymphknotenvergrößerung

Zusätzlich wird jedes Stadium in zwei Klassen A und B eingeteilt, je nach Nicht-Vorhandensein (A) oder Vorhandensein (B) mindestens eines der drei folgenden Allgemeinsymptome (B-Symptome)

- unerklärter Gewichtsverlust von mehr als 10% des Körpergewichts in den vorausgehenden 6 Monaten,
- unerklärtes Fieber über 38°C ,
- Nachtschweiß.

Da sich das Ausbreitungsmuster von Hodgkin- und Non-Hodgkin-Lymphomen im allgemeinen unterscheidet, sind für die Anwendung der Ann-Arbor-Klassifikation bei Non-Hodgkin-Lymphomen mehrere Modifikationen zur Bewertung des extranodalen Befalls vorgeschlagen [25], [26], jedoch bislang nicht allgemein akzeptiert worden. Zur Stadieneinteilung nach Ann Arbor ist eine genaue Definition der Lymphknotenregionen notwendig.

Nach Armitage [27], [28] sollen auf dem Rye-Symposium 1965 die folgenden 14 anatomischen Regionen als zu unterscheiden definiert worden sein: rechts zervikal (zervikal, supraclavikulär, okzipital, präaurikulär), links zervikal, rechts axillär, links axillär, rechts infraklavikulär, links infraklavikulär, mediastinal, hilär, periaortal, mesenterial, rechts iliakal, links iliakal, rechts inguinal und links inguinal.

Die Stadieneinteilung der Lymphome dient vor allem der Stratifikation der Patienten zur optimalen Therapiewahl. Bei lokalisierten Stadien I und II dient die Ann-Arbor-Klassifikation der optimalen Durchführung einer lokalisierten Strahlentherapie, bei fortgeschrittenen Stadien III und IV wird dagegen meist eine systemische Therapie durchgeführt.

Relevanz beim Mantelzell-Lymphom

In den bereits genannten Publikationen zum Mantelzell-Lymphom wurde stets das Stadium nach Ann Arbor untersucht und als Prognosefaktor evaluiert. Hierbei zeigte sich zunächst, dass der überwiegende Anteil der Patienten (bis zu 86% in [21]) erst im Stadium IV diagnostiziert worden war. Zur Bewertung der prognostischen Relevanz wurde meist nicht die Einteilung in vier Stadien untersucht, sondern die Einteilung in lokalisiertes Stadium (I oder II) einerseits und fortgeschrittenes Stadium (III oder IV) andererseits. In sechs von elf Studien zeigten Patienten mit fortgeschrittenem Stadium kein signifikant schlechteres Gesamtüberleben (s. Tabelle 1.7). Die prognostische Relevanz der Stadieneinteilung nach Ann Arbor beim Mantelzell-Lymphom scheint daher begrenzt zu sein.

1.2.2 International Prognostic Index

Beschreibung

Der 1993 veröffentlichte International Prognostic Index (IPI, [29]) war das erste Prognosemodell bei Lymphomen und wurde mit einem Datensatz von 2031 Patienten mit aggressiven Non-Hodgkin-Lymphomen entwickelt. Auf der Basis von fünf vor Therapiebeginn bestimmten klinischen Parametern unterscheidet er vier Risikogruppen in Bezug auf das Gesamtüberleben. Als prognostisch ungünstig erwiesen sich ein Alter von 60 Jahren oder mehr, ein fortgeschrittenes Ann-Arbor-Stadium III oder IV, ein erhöhter LDH-Wert, ein ECOG Performance Status 2 - 4 (s. Abschnitt 2.1.3) sowie mehr als ein extranodaler Befall. Je nach Anzahl der Risikofaktoren wurden Patienten eingeteilt in eine Low Risk-Gruppe (LR, 0 oder 1 Risikofaktor), eine Low Intermediate Risk- (LIR, 2 Risikofaktoren), High Intermediate Risk- (HIR, 3 Risikofaktoren) und eine High Risk-Gruppe (HR, 4 oder 5 Risikofaktoren). Tabelle 1.3 zeigt die Patienteneinteilung und die Gesamtüberlebensraten in den Risikogruppen nach dem IPI auf dem Gesamtdatensatz aus [29].

Bei der Entwicklung des IPI wurden die folgenden Charakteristika auf ihren poten-

Risikogruppe	Anzahl RF	Patienten	2-Jahres OS	5-Jahres OS
Low Risk	0 oder 1	35%	84%	73%
Low Intermediate Risk	2	27%	66%	51%
High Intermediate Risk	3	22%	54%	43%
High Risk	4 oder 5	16%	34%	26%

Tabelle 1.3: Patienteneinteilung und Gesamtüberleben (OS) in den Risikogruppen nach IPI auf dem Gesamtdatensatz der Originalpublikation [29] von 2031 Patienten. RF: Risikofaktoren.

ziellen prognostischen Einfluss geprüft: Geschlecht, Alter, Tumorstadium, Allgemeinzustand nach dem ECOG Performance Status, B-Symptome, Anzahl extranodaler Befälle, maximale Tumorgröße, und die Serum-Konzentrationen von LDH, Albumin, und β_2 -Mikroglobulin. Bei der Variablenselektion wurde das Geschlecht auf Grund fehlenden univariaten Einflusses auf das Gesamtüberleben, sowie Albumin und β_2 -Mikroglobulin auf Grund eines hohen Anteils fehlender Werte nicht mitberücksichtigt.

Alle Risikofaktoren wurden dichotomisiert verwendet. Bei der Bildung von Risikogruppen auf Basis der Anzahl vorhandener Risikofaktoren wurden die fünf Risikofaktoren gleich gewichtet, da die im multivariaten Modell geschätzten relativen Risiken (Tabelle 1.4) vergleichbar waren. Kategorien mit ähnlichem Gesamtüberleben wurden zusammengefasst, so dass sich vier Risikogruppen ergaben. Eine weitere Zusammenfassung zu drei oder weniger Risikogruppen überließen die Autoren den Ärzten, um je nach Zweck der Anwendung eine Flexibilität im Umgang mit dem IPI zu ermöglichen.

Risikofaktor	Vergleich	HazR	p-Wert
Alter	≥ 60 vs. < 60	1,96	$< 0,001$
LDH	$> \text{normal}$ vs. $\leq \text{normal}$	1,85	$< 0,001$
ECOG Performance Status	2 - 4 vs. 0 - 1	1,80	$< 0,001$
Stadium	III/IV vs. I/II	1,47	$< 0,001$
extranodale Befälle	> 1 vs. ≤ 1	1,48	$< 0,001$

Tabelle 1.4: Einfluss der fünf Risikofaktoren des IPI auf das Gesamtüberleben im multiplen Cox-Modell auf dem Trainingsdatensatz der Originalpublikation aus [29] mit 1385 Patienten. HazR: Hazard-Ratio.

Relevanz beim Mantelzell-Lymphom

Auf Grund der Einschlusskriterien waren in dem bei der Entwicklung des IPI verwendeten Datensatz Patienten mit Mantelzell-Lymphom nicht enthalten. Daher stellt sich die Frage, inwiefern der IPI auch zur prognostischen Klassifikation von Patienten mit Mantelzell-Lymphom geeignet ist. Møller et al. [9] untersuchten die prognostische Relevanz des IPI an einem bevölkerungsbezogenen Datensatz von 93 Patienten mit Mantelzell-Lymphom, die zwischen 1992 und 1998 in West-Dänemark diagnostiziert worden waren. Nach dem IPI gehörten 22% dieser Patienten zur LR-Gruppe, 41% zur LIR-Gruppe, 22% zur HIR-Gruppe und 16% zur HR-Gruppe. Der Logrank-Test zeigte einen signifikanten Einfluss des IPI auf das Gesamtüberleben ($p < 0,0001$) mit 2-Jahres-Überlebensraten von 85%, 73%, 45% und 27% in den vier Risikogruppen. Die Autoren kritisierten am IPI vor allem

den geringen Anteil von Patienten in der Hochrisikogruppe, die schlechte Trennung der drei ungünstigsten Risikogruppen sowie die fehlende prognostische Relevanz der Anzahl extranodaler Befälle.

Auch die anderen Autoren überprüften die prognostische Relevanz des IPI bei Patienten mit Mantelzell-Lymphom (s. Tabelle 1.7). Meist zeigte der IPI einen signifikanten Einfluss auf das Gesamtüberleben [8], [11], [16], [17], [18], [19], [20], [21]. Fast immer wurden jedoch die vier Risikogruppen des IPI zu drei [19], [17] oder zwei [14], [18], [8], [21], [22] zusammengefasst. Darüber hinaus verliefen die Überlebenskurven verschiedener Risikogruppen oft sehr ähnlich. So waren in [16] die Überlebenskurven der drei Gruppen LR, LIR und HIR nicht getrennt. Sowohl in [17] als auch in [19] und [20] verliefen die Überlebenskurven der mittleren beiden Risikogruppen sehr ähnlich. In [11] konnte nur die LR-Gruppe von den übrigen drei Gruppen signifikant abgetrennt werden. In [10], [14] und [22] zeigte der IPI dagegen keinen signifikanten Einfluss auf das Gesamtüberleben.

Die einzelnen Risikofaktoren des IPI sind des Weiteren auf ihre univariate prognostische Relevanz bei Patienten mit Mantelzell-Lymphom untersucht worden (s. Tabelle 1.7), oft aber mit modifizierten Grenzwerten. So wurde eine Altersgrenze von 65 [8], [19], [20] oder 70 [16] Jahren verwendet, oder der ECOG Performance Status [17] oder die Anzahl extranodaler Manifestationen [14], [18], [11] anders gruppiert. In sechs univariaten Analysen war das Alter signifikant [14], [16], [8], [17], [18], [20], nicht aber in [21] und [10]. Das Ann-Arbor-Stadium war in fünf Analysen univariat signifikant [14], [16], [8], [17], [11], in sechs dagegen nicht [18], [10], [19], [20], [21], [22]. Der ECOG Performance Status war in neun [14], [16], [8], [17], [9], [10], [19], [21], [11] von elf (nicht in [18], [20]) und die LDH in sieben [14], [16], [8], [17], [18], [20], [21] von zehn Analysen (nicht in [10], [22], [11]) signifikant. Die Anzahl extranodaler Manifestationen war in drei [14], [16], [11] von zehn Analysen (nicht in [8], [17], [9], [18], [10], [21], [22]) signifikant. Die berichteten Ergebnisse beziehen sich auf ein Signifikanzniveau von 5%, auch wenn in [19] offenbar ein Signifikanzniveau von 10% verwendet wurde.

Nur in [17] wurde in einem multiplen Modell die prognostische Relevanz der fünf Einzelfaktoren des IPI analysiert. Hierbei erwiesen sich nur Alter und Stadium als unabhängige Prognosefaktoren. Die übrigen multiplen Modelle enthielten weitere potenzielle Prognosefaktoren. Der IPI selbst war in fünf multiplen Regressionsmodellen enthalten [8], [9], [18], [20], [21], und hierbei immer zusammen mit mehreren Einzelfaktoren des IPI, was die Interpretation der Ergebnisse erschwert. In den multiplen Modellen zeigte der IPI nur in [18] unabhängige prognostische Relevanz.

Zusammenfassend scheint der IPI bei Patienten mit Mantelzell-Lymphom zwar von prognostischer Relevanz für das Gesamtüberleben zu sein, aber auf Grund einer ungünstigen Verteilung der Patienten zu den Risikogruppen, einer schlechten Auftrennung von vier Überlebenskurven und der mangelnden prognostischen Relevanz von Stadium und Anzahl extranodaler Manifestationen nicht optimal zu sein. Mehrere Autoren [9], [15],

[17] forderten daher die Entwicklung eines eigenen Prognosemodells für Patienten mit Mantelzell-Lymphom.

1.2.3 Follicular Lymphoma International Prognostic Index

Beschreibung

Seit 1993 wurde der IPI in der klinischen Praxis auch für Patienten mit follikulärem Lymphom eingesetzt, obwohl bei seiner Entwicklung keine Daten von Patienten mit dieser Lymphomentität verwendet wurden. Solal-Celigny et al. veröffentlichten 2004 den Follicular Lymphoma International Prognostic Index (FLIPI), der auf Daten von 1795 Patienten mit follikulärem Lymphom basiert [30]. Der FLIPI ist der erste speziell für eine einzige Lymphomentität entwickelte prognostische Index.

Der FLIPI definiert drei Risikogruppen auf Basis der Anzahl vorhandener Risikofaktoren. Die fünf Risikofaktoren des FLIPI sind Alter über 60 Jahre, Stadium III oder IV, Hämoglobin (Hb) unter 120g/l, LDH über dem oberen Normwert und mehr als vier befallene Lymphknotenregionen. Patienten mit höchstens einem Risikofaktor gehören der Low Risk (LR)-Gruppe an, Patienten mit zwei Risikofaktoren der Intermediate Risk (IR)-Gruppe und Patienten mit mehr als zwei Risikofaktoren bilden die High Risk (HR)-Gruppe. Tabelle 1.5 zeigt die Patientenverteilung und Gesamtüberlebensraten der FLIPI-Risikogruppen im Datensatz der Modellentwicklung mit 1795 Patienten. Bei der externen Validierung auf 919 später rekrutierten Patienten ließ sich die Trennung der drei Risikogruppen reproduzieren.

Risikogruppe	Anzahl RF	Patienten	5-Jahres OS	10-Jahres OS
Low Risk	0 - 1	36%	90,6%	70,7%
Intermediate Risk	2	37%	77,6%	50,9%
High Risk	3 - 5	27%	52,5%	35,5%

Tabelle 1.5: Patienteneinteilung und Gesamtüberleben (OS) in den Risikogruppen nach FLIPI auf dem Datensatz der Modellentwicklung von 1795 Patienten aus der Originalpublikation [30]. RF: Risikofaktoren.

Tabelle 1.6 zeigt die relativen Risiken (Hazard-Ratios) der fünf Risikofaktoren für das Gesamtüberleben im multiplen Cox-Modell.

Risikofaktor	Vergleich	HazR (95%-KI)
Alter	≥ 60 vs. < 60	2,38 (2,04 - 2,78)
Stadium	III/IV vs. I/II	2,00 (1,56 - 2,58)
Hämoglobin	$< 120\text{g/l}$ vs. $\geq 120\text{g/l}$	1,55 (1,30 - 1,88)
LDH	$> \text{normal}$ vs. $\leq \text{normal}$	1,50 (1,27 - 1,77)
nodale Befälle	> 4 vs. ≤ 4	1,39 (1,18 - 1,64)

Tabelle 1.6: Einfluss der fünf Risikofaktoren des FLIPI auf das Gesamtüberleben im multiplen Cox-Modell auf dem Datensatz der Modellentwicklung [30] mit 1795 Patienten. HazR: Hazard-Ratio, KI: Konfidenzintervall.

Auf dem Gesamtdatensatz von 4167 Patienten zeigten von 17 potenziellen Prognosefaktoren alle bis auf den Zelltyp univariaten Einfluss auf das Gesamtüberleben. Für die multiple Regression nicht berücksichtigt wurden Zelltyp, der ECOG Performance Status, weil er in Europa und den USA unterschiedlich vollständig dokumentiert war, die nur in europäischen Zentren dokumentierte Blutsenkungsgeschwindigkeit, sowie β_2 -Mikroglobulin und Albumin auf Grund des hohen Anteils fehlender Werte (17% bzw. 51%). Im multiplen Cox-Modell auf 1795 vollständigen Datensätzen in Bezug auf 12 Parameter zeigten B-Symptomatik, die Anzahl extranodaler Manifestationen (außer Knochenmark), Milzbefall und Thrombozytenzahl keine unabhängige prognostische Relevanz. Von den übrigen acht unabhängigen Prognosefaktoren wurden die drei Parameter Geschlecht, Knochenmarksbefall und Lymphozytenzahl eliminiert, um schließlich die a priori gewünschte Anzahl von maximal fünf Prognosefaktoren zu erhalten.

Ein Problem bei der Bestimmung des FLIPI ist die Anzahl befallener Lymphknotenregionen als Risikofaktor, da die Regionen nach einem neuen, nicht standardisierten Verfahren gezählt werden müssen. Im Gegensatz zu der etablierten Methode nach Ann Arbor (s. Abschnitt 1.2.1) werden paraaortale und iliakale Befälle zusammengefasst und nicht nach links und rechts unterschieden. Außerdem wird nicht zwischen hochzervikal und zervikal unterschieden. Diese Unterschiede erschweren die retrospektive Bestimmung des FLIPI und führen möglicherweise auch zu uneinheitlicher Bestimmung bei zukünftigen Patienten.

Relevanz beim Mantelzell-Lymphom

Møller et al. validierten in [9] als Einzige auch den FLIPI bei Patienten mit Mantelzell-Lymphom. Nach dem FLIPI ließen sich 23% der Patienten in die LR-Gruppe klassifizieren, 25% der Patienten in die IR-Gruppe und 52% der Patienten bildeten die HR-Gruppe. Der Logrank-Test zeigte einen signifikanten Einfluss des FLIPI auf das Gesamtüberleben ($p < 0,0001$). Die 2-Jahres-Überlebensraten waren 86%, 77% und 45% für die LR-, die IR- und die HR-Gruppe. Univariat zeigten die Anzahl der Lymphknotenregionen, Hb und ECOG Performance Status einen signifikanten Einfluss aufs Gesamtüberleben, die Anzahl extranodaler Manifestationen dagegen nicht. Im multiplen Cox-Modell mit IPI, FLIPI sowie den Einzelfaktoren zu IPI und FLIPI und der Therapie (CHOP vs. non-CHOP) waren FLIPI, ECOG Performance Status, Anzahl extranodaler Befälle sowie die Therapie unabhängig voneinander signifikant. IPI, Alter, Stadium, Anzahl extranodaler Befälle, LDH und Hb waren in diesem Gesamtmodell nicht signifikant.

Die Überlebenskurven der LR- und der IR-Gruppe erschienen nicht gut getrennt, und die HR-Gruppe umfasste über die Hälfte der Patienten. Die Ergebnisse der multiplen Cox-Regression sind schwer zu interpretieren, da die beiden Indizes zusammen mit ihren Einzelfaktoren im Modell enthalten waren und damit Korrelationen der Kovariaten sehr

wahrscheinlich sind. Unabhängig vom FLIPI scheint aber vor allem der ECOG Performance Status von prognostischer Relevanz zu sein, der aus technischen Gründen nicht im FLIPI berücksichtigt worden war. Møller et al. schlossen aus ihren Ergebnissen, dass der FLIPI dem IPI bei Patienten mit Mantelzell-Lymphom überlegen, aber dennoch nicht ideal sei.

1.2.4 Prognosefaktoren beim Mantelzell-Lymphom

Seitdem das Mantelzell-Lymphom mit der REAL-Klassifikation von 1994 als eigenständige Entität unter den Non-Hodgkin-Lymphomen international anerkannt wurde, ist eine Reihe von Publikationen zu klinischen und histopathologischen Charakteristika, Therapie und Prognose des Mantelzell-Lymphoms erschienen, die auch potenzielle Prognosefaktoren auf ihre prognostische Relevanz untersuchten. Diese Publikationen wurden bereits in Kapitel 1.1 aufgeführt, in den Tabellen 1.1 und 1.2 sind die Patientenkollektive charakterisiert.

Tabelle 1.7 gibt eine Übersicht über die Relevanz der untersuchten Prognosefaktoren. Als potenzielle Prognosefaktoren wurden klinische und histopathologische Parameter auf ihren Einfluss auf das Gesamtüberleben geprüft. Von den klinischen Parametern wurden meistens Alter, Stadium, Allgemeinzustand, LDH, Knochenmarksbefall, Anzahl extranodaler Befälle und B-Symptomatik untersucht. Häufig kamen der Befall der Milz oder des peripheren Blutes, Hämoglobin sowie die Therapie in Betracht. Seltener untersucht wurden die Laborparameter Leukozyten-, Lymphozyten- und Thrombozytenzahl, Albumin und β_2 -Mikroglobulin, sowie die Tumorgröße und der Befall des Gastrointestinaltraktes oder der Leber. Von den histopathologischen Parametern interessierte meist die Form der Tumorzellen (Zytologie) und die Architektur des Tumorgewebes, sowie Proliferationsmaße wie die Anzahl der Mitosen und die Expression von Ki-67 in den Tumorzellen. Bis auf wenige Ausnahmen wurden alle metrischen Merkmale bereits dichotomisiert analysiert, wobei die verwendeten Grenzwerte sich oft willkürlich unterschieden.

In den univariaten Analysen zeigten von den häufig untersuchten klinischen Parametern Allgemeinzustand, Hb, Therapie, LDH, und Alter am konsistentesten einen Einfluss auf das Gesamtüberleben. Weniger konsistent waren Befall von Knochenmark oder peripherem Blut, B-Symptomatik, Milzbefall und Stadium. Kaum oder nie zeigten die Anzahl extranodaler Befälle und das Geschlecht einen Einfluss auf das Gesamtüberleben. Von den häufig untersuchten histopathologischen Parametern zeigten die Proliferationsmaße Mitosen und Ki-67 am konsistentesten einen Einfluss auf das Gesamtüberleben, weniger konsistent waren die Ergebnisse zu Zytologie und Architektur (Tabelle 1.7).

Insgesamt sind die publizierten Ergebnisse sehr heterogen, und auf Grund meist relativ kleiner Fallzahlen unter 130 und unterschiedlicher Definitionen und Methodik schwer vergleichbar. Die Ergebnisse der multiplen Regressionen sind besonders schlecht vergleich-

Erstautor	n	univariate Analyse	multiple Cox-Regression	EPV
Tiemann [14]	304	Alter , Geschlecht, ECOG , Stadium , B-Symptome , extranodal , KM , LDH , Architektur , Zytologie, Mitosen , Ki-67 , IPI	Alter , Geschlecht, ECOG , Stadium , LDH , Mitosen , Ki-67	< 8
Räty [15]	127	Architektur , Zytologie, Mitosen , Ki-67 , p53	Alter , ECOG , Stadium , extranodal, LDH , Architektur , Mitosen , Ki-67	< 7,8
Samaha [16]	121	Alter , ECOG , Stadium , B-Symptome, Bulk , extranodal , GI , Milz , PB , Hb , LDH , β_2 - Mikroglobulin , Albumin , Zytogenetik, Architektur , Zytologie, IPI , Therapie	Alter , ECOG , Stadium , B-Symptome, Bulk , extranodal, GI , Milz , PB , Hb , LDH , β_2 - Mikroglobulin , Albumin	< 4,3
Andersen [8]	97	Alter , Geschlecht, ECOG , Stadium , B-Symptome, extranodal, KM , Milz , Leukozyten , Thrombozyten , Hb , LDH , Albumin , IgG , Ki-67 , IPI , extranodal dominierend	Alter , ECOG , Stadium , Milz , Leukozyten , Thrombozyten , Hb , LDH , IPI , extranodal dominierend	ca. 6,1
Oinonen [17]	94	Alter , Geschlecht, ECOG , Stadium , B-Symptome , Bulk , extranodal, KM , Milz , Blut , Leukozyten , Lymphozyten , Thrombozyten , Hb , LDH , Thymidin-Kinase , ESR , IPI , Therapie	Alter , Geschlecht, ECOG , Stadium , B-Symptome, KM , Milz , PB , Leukozyten , Lymphozyten , Hb , LDH , ESR , Therapie	ca. 3,9
Møller [9]	93	ECOG , nodal , extranodal, Hb , IPI , FLIPI , Therapie	Alter , ECOG , Stadium , extranodal, Hb , LDH , nodal , IPI , FLIPI , Therapie	ca. 5
Schrader [18]	93	Alter , Geschlecht, ECOG , Stadium , B-Symptome, extranodal, KM , LDH , Ki-67 , TopoIIα , IPI	Alter , Geschlecht, ECOG , Stadium , B-Symptome, extranodal, KM , LDH , Ki-67 , TopoIIα , IPI	ca. 5,3
Argatoff [10]	80	Alter , ECOG , Stadium , extranodal, PB , LDH , Architektur , Zytologie , Mitosen , IPI	ND	ND
Weisenburger [19]	68	ECOG , Stadium , B-Symptome , KM , Architektur , Zytologie , Mitosen , IPI	ND	ND
Zucca [20]	65	Alter , Geschlecht, ECOG , Stadium , B-Symptome, KM , Leber , Milz , LDH , β_2 - Mikroglobulin , IPI , Therapie	Alter , ECOG , Stadium , LDH , IPI , Therapie	ca. 5
Bosch [21]	59	Alter , Geschlecht, ECOG , Stadium , B-Symptome , extranodal, Milz , PB , Leukozyten , LDH , Zytologie , Mitosen , IPI , Therapie	Alter , ECOG , Stadium , B-Symptome, Milz , Leukozyten , LDH , Zytologie , Mitosen , IPI , Therapie	< 2,7
Decaudin [22]	45	Stadium , extranodal, KM , Leber , Milz , PB , Thrombozyten , Hb , LDH , Zytologie , IPI	ND	ND
Velders [11]	41	ECOG , Stadium , extranodal, KM , LDH , IPI , Ki-67 , Cyclin D1	ND	ND

Tabelle 1.7: Publierte Daten zu Prognosefaktoren für das Gesamtüberleben bei Patienten mit Mantelzell-Lymphom. Auf dem 5%-Niveau signifikante Parameter sind fett markiert. n: Patientenzahl, EPV: events per variable bei der multiplen Cox-Regression, extranodal: Anzahl extranodaler Befälle, KM: Knochenmarksbefall, GI: gastrointestinaler Befall, Milz: Milzbefall, PB: Befall des peripheren Blutes, ESR: Erythrozyten-Senkungsgeschwindigkeit, nodal: Anzahl befallener Lymphknotenregionen, Leber: Leberbefall, ND: nicht durchgeführt.

bar, da unterschiedliche Prognosefaktoren eingebracht wurden und die Methodik der Variablenselektion sich oft unterschied oder nicht beschrieben war. Auch war die Fallzahl in den meisten Fällen für eine aussagekräftige multiple Cox-Regression zu klein. In keiner der multiplen Analysen lag die Zahl der Ereignisse pro Variable (events per variable, EPV) über dem in [31], [32] empfohlenen Minimalwert von 10. Bislang wurde kein prognostischer Index speziell für Patienten mit Mantelzell-Lymphom entwickelt.

1.3 Fragestellung und Ziele der Arbeit

Mantelzell-Lymphome sind unter den Non-Hodgkin-Lymphomen relativ selten, weisen aber eine besonders schlechte Prognose auf. Das Mantelzell-Lymphom ist trotz einer breiten Auswahl an Therapiemöglichkeiten derzeit nicht heilbar. Die Patienten sterben im Median nach drei bis vier Jahren, wobei der individuelle Verlauf der Erkrankung zwischen einem Überleben von wenigen Monaten und mehreren Jahren krankheitsfreier Zeit variieren kann. Prognostische Klassifikationssysteme können wichtige Instrumente zur Auswahl geeigneter individualisierter Therapiestrategien auf Grund unterschiedlichen vorhergesagten Verlaufes sein. Für Patienten mit Mantelzell-Lymphom existiert derzeit kein Prognosemodell und die dennoch in der klinischen Praxis angewendeten Klassifikationssysteme wurden für andere Lymphomentitäten entwickelt und weisen bei Patienten mit Mantelzell-Lymphom, wie in der Literatur mehrfach dargestellt, nur limitierte prognostische Relevanz auf. Darüber hinaus sind bezüglich der Relevanz potenzieller Prognosefaktoren in mehreren Arbeiten divergierende Ergebnisse publiziert worden, die meist auf limitierten Patientenzahlen unter 130 beruhen.

Die Deutsche Studiengruppe für Niedrig-Maligne Lymphome (German Low-Grade Lymphoma Study Group, GLSG) und das European MCL Network, ein europäisches Netzwerk von Studiengruppen und klinischen Zentren, führten seit 1996 drei große randomisierte Primärtherapiestudien durch, im Rahmen derer Patienten mit fortgeschrittenem Mantelzell-Lymphom behandelt und regelmäßig möglichst lebenslang nachbeobachtet wurden. Ziel der vorliegenden Arbeit war es, an dem relativ großen Datensatz dieser Studien die Relevanz der bisher in der klinischen Praxis eingesetzten prognostischen Klassifikationssysteme sowie der potenziellen Prognosefaktoren zu klären. Falls auf Grund dieser Ergebnisse erforderlich, sollte schließlich ein neues prognostisches Klassifikationssystem speziell für Patienten mit fortgeschrittenem Mantelzell-Lymphom entwickelt werden.

Kapitel 2

Daten und statistische Methoden

2.1 Datenbasis

2.1.1 Studien der GLSG und des European MCL Network

Die vorliegende Analyse beruht auf den Daten der randomisierten, multizentrischen Primärtherapiestudien „CHOP vs. MCP“ [33] und „CHOP vs. R-CHOP“ [34], [35], [36] der German Low-Grade Lymphoma Study Group (GLSG), sowie „European MCL Trial 1“ [37] des European MCL Network. Die Rekrutierung zur Studie „CHOP vs. MCP“ erfolgte von März 1996 bis Juli 2000. Seit Mai 2000 konnten Patienten zur Nachfolgestudie „CHOP vs. R-CHOP“ gemeldet werden. Die Rekrutierung zur Europäischen MCL-Studie erfolgte seit September 1996. Die Studien waren in Übereinstimmung mit der Deklaration von Helsinki in der jeweils aktuellen Fassung durchgeführt und von den zuständigen Ethikkommissionen zustimmend bewertet worden. Alle Patienten hatten der Studienteilnahme nach Aufklärung schriftlich zugestimmt. Die drei Studien waren bis auf die zu prüfenden Therapien im Wesentlichen gleich aufgebaut. Tabelle 2.1 gibt eine Übersicht über die wichtigsten Eigenschaften der Studien, welche in den folgenden Abschnitten genauer ausgeführt werden.

Patienten

Alle drei Studien nahmen erwachsene Patienten mit Mantelzell-Lymphom auf, die bei den GLSG-Studien darüber hinaus Patienten mit follikulärem Lymphom oder lymphoplasmazytischem oder lymphoplasmocytoidem Immunozytom. Die histologische Diagnose erfolgte nach der Kiel- oder der REAL-Klassifikation, später nach der aktuell gültigen WHO-Klassifikation. Zentrale Referenzpathologen überprüften die in der Regel vom lokalen Pathologen des meldenden Zentrums gestellte histologische Diagnose.

Einschlusskriterium war weiter ein fortgeschrittenes Stadium III oder IV nach Ann Arbor. Bei Patienten in lokalisierten Stadien I oder II sollte eine möglicherweise kurative

Studie	CHOP vs. MCP	CHOP vs. R-CHOP	European MCL 1
Referenz	[33]	[34], [35], [36]	[37]
Studiengruppe	GLSG	GLSG	Europ. MCL Network
Rekrutierung	03/1996 - 07/2000	ab 05/2000	09/1996 - 07/2004
Alter	ab 18 Jahre	ab 18 Jahre	18 - 66 Jahre
Histologie	MCL, FL, IC	MCL, FL, IC	MCL
Ann-Arbor-Stadium	III, IV	III, IV	III, IV
Aufnahmezeitpunkt	Diagnose (MCL)	Diagnose (MCL)	Diagnose
	Therapiebedürftigkeit (FL, IC)	Therapiebedürftigkeit (FL, IC)	
1. Randomisierung	CHOP vs. MCP	CHOP vs. R-CHOP	ASCT vs. IFN
2. Randomisierung	ASCT vs. IFN (< 60 Jahre)	ASCT vs. IFN (< 60 Jahre)	
Induktionstherapie	MCP, CHOP	CHOP, R-CHOP	CHOP-like
Postremissionstherapie	ASCT, IFN	ASCT, IFN	ASCT, IFN

Tabelle 2.1: Beschreibung der drei Studien von GLSG und European MCL Network, welche die Daten zur vorliegenden Analyse lieferten. MCL: Mantelzell-Lymphom, FL: folliculäres Lymphom, IC: lymphoplasmazytisches oder -cytoides Immunozytom, MCP: Mitoxantron + Chlorambucil + Prednison, CHOP: Cyclophosphamid + Hydroxydaunorubicin + Vincristin + Prednison, R-CHOP: CHOP + Rituximab, ASCT: autologe Stammzelltransplantation, IFN: Erhaltungstherapie mit Interferon- α .

Bestrahlung einer systemischen Chemotherapie vorgezogen werden. Bei Patienten mit folliculärem Lymphom oder Immunozytom musste Therapiebedürftigkeit bestehen, während bei Patienten mit Mantelzell-Lymphom auf Grund der schlechten Prognose Studienaufnahme und Therapiebeginn unmittelbar nach Diagnose erfolgten. Zur Europäischen MCL-Studie waren Patienten bis zu einem Alter von 66 Jahren zugelassen, während das Alter bei den beiden GLSG-Studien nicht nach oben begrenzt war. Von der Studienteilnahme ausgeschlossen waren vorbehandelte Patienten sowie Patienten mit einem schlechten Allgemeinzustand (ECOG Performance Status 2 - 4, außer lymphombedingt) oder mit Begleiterkrankungen oder anderen Kontraindikationen gegen eine studiengerechte Therapiedurchführung.

Randomisierung und Behandlungsplan

Die Studien zielten auf eine Verbesserung der Primärtherapie des Lymphoms ab. Die Primärtherapie beinhaltete zunächst eine Kombinations-Chemotherapie (Induktion) zur Reduktion der Tumormasse. Bei Patienten mit ausreichendem Ansprechen auf die Induktion erfolgte im Anschluss eine Erhaltungstherapie. Die beiden Studien der GLSG hatten ein faktorielles Design und verglichen randomisiert sowohl zwei Induktionstherapieschemata als auch zwei Erhaltungstherapieschemata. Die Europäische MCL-Studie verglich zwei Erhaltungstherapieschemata.

So verglich die GLSG-Studie von 1996 die Induktions-Chemotherapieschemata CHOP (Cyclophosphamid, Hydroxydaunorubicin, Vincristin = Oncovin, Prednison) und MCP (Mitoxantron, Chlorambucil, Prednison). Die Nachfolgestudie der GLSG untersuchte ab 2000 die zusätzliche Wirksamkeit des monoklonalen Antikörpers Rituximab in Kombination mit CHOP (R-CHOP).

Patienten beider GLSG-Studien, die jünger als 60 Jahre alt waren, wurden nach 2 Zyklen Induktionstherapie zwischen einer autologen Stammzelltransplantation (ASCT) und einer Erhaltungstherapie mit Interferon- α (IFN- α) randomisiert. Die Erhaltungstherapie wurde dann nach Erreichen einer Tumorreduktion von mindestens 50% begonnen. Für ältere Patienten der GLSG-Studien mit einer Tumorreduktion von mindestens 50% war durchweg eine Erhaltungstherapie mit IFN- α vorgesehen. Im Rahmen der Europäischen MCL-Studie wurden die Patienten vor Beginn der Induktionstherapie zu einem der beiden Erhaltungstherapiearme ASCT oder IFN- α randomisiert und erhielten dann eine anthrazyklinhaltige Kombinations-Chemotherapie („CHOP-like“). Im Falle einer Tumorreduktion von mindestens 50% schloss sich daran die Erhaltungstherapie gemäß dem Randomisierungsergebnis an. Die Erhaltungstherapieschemata für jüngere Patienten der GLSG-Studien und in der Europäischen MCL-Studie waren identisch. Daher konnte die Induktionstherapie bei Patienten der Europäischen MCL-Studie in Deutschland in der Regel im Rahmen der jeweils laufenden GLSG-Studie erfolgen.

Um möglichst vergleichbare Risikostrukturen der Therapiegruppen zu erreichen, waren die Randomisierungen stratifiziert nach den IPI-Risikofaktoren, bei den beiden GLSG-Studien auch nach histologischer Diagnose. Die Randomisierungen zwischen den Erhaltungstherapien waren zusätzlich stratifiziert nach der Art der initialen Therapie. Die Nachbeobachtung aller Patienten erfolgte halbjährlich möglichst zeitlebens.

Zielgrößen

Die Beurteilung des Ansprechens auf die Induktionstherapie erfolgte nach den Kriterien der International Working Group [38]. Die Kategorien sind hierbei komplette Remission (CR, vollständiges Verschwinden aller Krankheitszeichen), partielle Remission (PR, Rückgang der Tumormasse um mindestens 50%), Progression (progressive disease, PD, Vergrößerung der Lymphommasse um mindestens 25% oder Auftreten neuer Tumorläsionen), Tod und stabile Erkrankung (stable disease, SD, keine der anderen Kategorien).

Die Effektivität der initialen Chemotherapie wurde an Hand der Rate kompletter Remissionen und der Gesamtremissionsrate (Rate von CR oder PR) bestimmt. Der Vergleich der Erhaltungstherapien erfolgte mit der Remissionsdauer (RD), dem Zeitintervall zwischen dem Ende einer erfolgreichen Induktionstherapie (PR oder CR) und einer Progression des Lymphoms oder dem Tod des Patienten. Die Zeit bis zum Therapieversagen (time to treatment failure, TTF) war die Zeit zwischen Beginn der Induktionstherapie und dem Eintreten eines Therapieversagens. Ein Therapieversagen lag vor, wenn die initiale Chemotherapie nicht zu einem Rückgang der Tumormasse um mindestens 50% führte, oder bei Progression oder beim Tod des Patienten. Das Gesamtüberleben (overall survival, OS) war die Zeit zwischen Studienaufnahme und dem Tod des Patienten. Die Zeit bis zum Therapieversagen, die Remissionsdauer und das Gesamtüberleben konnten Zen-

sierungen aufweisen. TTF und RD wurden bei Patienten ohne Therapieversagen während der Beobachtungszeit zum Zeitpunkt des letzten Tumorstagings zensiert, OS bei noch lebenden Patienten zum Zeitpunkt des letzten Kontakts.

Hauptzielgröße für den Vergleich der Induktionstherapien war bei der Studie „CHOP vs. MCP“ die Rate kompletter Remissionen (CR-Rate) und bei der Nachfolgestudie „CHOP vs. R-CHOP“ die Gesamtremissionsrate. Für den Vergleich der Erhaltungstherapien war in allen Studien die Remissionsdauer die Hauptzielgröße. Die jeweils übrigen Parameter wurden als Nebenzielgrößen ausgewertet. Für den statistischen Vergleich der Therapien wurden sequenzielle Verfahren nach Whitehead [39] angewendet mit geplanten Zwischenauswertungen nach jedem neu auswertbaren Patienten oder jedem neuen Ereignis. Sobald eine der Fragestellungen zur Induktionstherapie oder Erhaltungstherapie entschieden war, wurde die Randomisierung beendet. Die Rekrutierung wurde für die jeweils andere Fragestellung fortgeführt und neu gemeldete Patienten erhielten nach Entscheidung der Studiengruppe die jeweils überlegene Therapie.

Ergebnisse

Die Studie „CHOP vs. MCP“ zeigte für die Induktionstherapie im Gesamtkollektiv von 415 Patienten zwar keinen Unterschied für die Hauptzielgröße, die Rate kompletter Remissionen, (CHOP: 18% vs. MCP: 14%, $p = 0,066$), aber eine signifikant höhere Gesamtremissionsrate nach CHOP (86% vs. 77%, $p = 0,0094$, [33]). Für die 86 Patienten mit Mantelzell-Lymphom waren mit einer Rate kompletter Remissionen von 15% vs. 20% und einer Gesamtremissionsrate von 87% vs. 73% die Ergebnisse vergleichbar. Für CHOP und MCP konnten keine Unterschiede in Bezug auf Zeit bis zum Therapieversagen, Remissionsdauer und Gesamtüberleben entdeckt werden.

Die Studie „CHOP vs. R-CHOP“ zeigte für die Hauptzielgröße Gesamtremissionsrate einen signifikanten Vorteil für R-CHOP (94% vs. 75% für CHOP, $p = 0,0054$) sowie eine signifikant höhere Rate kompletter Remissionen (34% vs. 7%, $p < 0,0001$) bei 121 Patienten mit fortgeschrittenem Mantelzell-Lymphom [34]. Auch die mediane Zeit bis zum Therapieversagen konnte signifikant von 14 Monaten auf 21 Monate verlängert werden, für Remissionsdauer und Gesamtüberleben konnten jedoch keine Unterschiede entdeckt werden.

Im Rahmen der Europäischen MCL-Studie [37] führte die autologe Stammzelltransplantation zu einer signifikanten Verlängerung der Remissionsdauer von 17 Monaten unter Interferon- α -Erhaltung auf 39 Monate ($p = 0,0108$, $n = 122$ Patienten). Auch hier konnte kein Vorteil für das Gesamtüberleben gezeigt werden.

Auf Grund der Ergebnisse wurde die Randomisierung zwischen CHOP und MCP im Dezember 1998 beendet und alle nachfolgenden Patienten dieser Studie erhielten CHOP. Ähnlich wurde die Randomisierung zwischen CHOP und R-CHOP im August 2003 been-

det, alle nachfolgenden Patienten erhielten R-CHOP. In Bezug auf die Remissionsdauer zeigte der statistische Test bei Patienten mit follikulärem Lymphom im April 1999 und für die Patienten der Europäischen MCL-Studie im Oktober 2000 die Überlegenheit der Stammzelltransplantation gegenüber der Interferon-Erhaltung. In allen Studien wurde jedoch die Randomisierung jüngerer Patienten zwischen ASCT und IFN- α fortgeführt, um die Frage einer möglichen Überlebenszeitverlängerung durch die ASCT zu beantworten.

2.1.2 Patienten

Die vorliegende Auswertung verwendete die Daten aller Patienten mit fortgeschrittenem Mantelzell-Lymphom aus den drei genannten Studien der GLSG und des European MCL Network. Bei der Auswahl wurden Histologie und Ann-Arbor-Stadium überprüft.

2.1.3 Baselinecharakteristika

Die folgenden klinischen Parameter wurden im Rahmen der Studiendokumentation vor Therapiebeginn erhoben und kamen als potenzielle Prognosefaktoren in Betracht (Tabelle 2.2).

Parameter
Alter bei Studienaufnahme
Geschlecht
Allgemeinzustand (ECOG Performance Status)
Ann-Arbor-Stadium
Anzahl befallener Lymphknotenregionen
Tumorgröße (maximaler Durchmesser des größten befallenen Lymphknotens)
Anzahl extranodaler Befälle
Knochenmarksbefall
Milzbefall
B-Symptome
Leukozytenzahl im Blut
Granulozytenzahl im Blut
Lymphozytenzahl im Blut
Monozytenzahl im Blut
Thrombozytenzahl im Blut
Hämoglobinkonzentration im Blut
LDH-Enzymaktivität im Serum
Albumin-Konzentration im Serum
β_2 -Mikroglobulin-Konzentration im Serum
IgG-Konzentration im Serum
IgA-Konzentration im Serum
IgM-Konzentration im Serum

Tabelle 2.2: Potenzielle Prognosefaktoren.

Allgemeine Parameter waren das Alter bei Studienaufnahme, das Geschlecht und der Allgemeinzustand des Patienten nach dem Eastern Cooperative Oncology Group (ECOG) Performance Status ([40], Tabelle 2.3).

Als tumorspezifische Parameter dokumentierte der behandelnde Arzt im Rahmen der primären Diagnostik das Ausbreitungsstadium des Lymphoms nach Ann Arbor [23], das Lymphknoten-Befallsmuster, den größten Durchmesser des größten befallenen Lymphknotens (Tumorgröße), das Vorhandensein und die Lokalisation extranodaler Befälle, vor

ECOG	Zustand des Patienten
0	vollständig aktiv, keine Einschränkung
1	eingeschränkte körperliche Aktivität und Arbeitsfähigkeit, nicht bettlägerig
2	arbeitsunfähig, selbständige Lebensführung, weniger als 50% der Zeit bettlägerig
3	weitgehend unfähig, sich selbst zu versorgen, mehr als 50% der Zeit bettlägerig
4	vollständig bettlägerig und pflegebedürftig
5	tot

Tabelle 2.3: Beurteilung des Allgemeinzustands nach dem ECOG Performance Status [40].

allen des Knochenmarks, sowie das Vorhandensein von Milzbefall oder von B-Symptomen. Aus dem peripheren Blut wurden die Zahl der Leukozyten, Granulozyten, Lymphozyten, Monozyten und Thrombozyten sowie der Hämoglobinwert (Hb) dokumentiert. Im Serum wurden die Aktivität der Laktat-Dehydrogenase (LDH), sowie die Konzentrationen von Albumin, β_2 -Mikroglobulin und der Immunglobuline IgG, IgA und IgM erhoben.

2.1.4 Zielgrößen

Die Hauptzielgröße zur Validierung der bekannten Prognosemodelle, zur Beurteilung der prognostischen Relevanz der potenziellen Prognosefaktoren sowie zur Entwicklung des neuen Prognosemodells war das Gesamtüberleben, d.h. die Zeit von Studienaufnahme bis zum Tod jeglicher Ursache. Das neu entwickelte Prognosemodell wurde auch hinsichtlich seiner prognostischen Relevanz für die Zeit bis zum Therapieversagen untersucht.

2.2 Beschreibung des Patientenkollektivs

Die Daten zu Baselinecharakteristika, Therapie, Therapieerfolg und Gesamtüberleben wurden deskriptiv statistisch ausgewertet. Für kategoriale Baselineparameter, Therapie und Therapieansprechen wurden absolute und relative Häufigkeiten bestimmt, für metrische Baselineparameter Median, sowie Minimum und Maximum. Die Verteilung der metrischen Baselineparameter wurde mit Histogrammen grafisch dargestellt. Bei stark rechtsschiefer Verteilung wurde eine Logarithmustransformation durchgeführt. Zur Untersuchung möglicher paarweiser Korrelationen wurden für ausgewählte metrische Baselinecharakteristika Streudiagramme dargestellt und jeweils der Korrelationskoeffizient nach Pearson geschätzt. Für die Zeit-Ereignis-Daten zu Gesamtüberleben und Zeit bis zum Therapieversagen wurden Kaplan-Meier-Kurven geschätzt, und die 5-Jahres-Raten mit 95%-Konfidenzintervallen nach der Greenwood-Formel berichtet. Die mediane Nachbeobachtungszeit wurde an den Patienten ohne Ereignis sowie mit der Zensierungsverteilung [41] geschätzt.

2.3 Validierung der bekannten Prognosemodelle

Die für Lymphome bekannten Klassifikationsmodelle Ann-Arbor-Stadium, IPI und FLIPI wurden an dem Studienkollektiv von Patienten mit fortgeschrittenem Mantelzell-Lymphom auf ihre prognostische Relevanz in Bezug auf das Gesamtüberleben überprüft. Hierfür interessierte zunächst die Verteilung der Patienten auf die einzelnen Risikogruppen (absolute und relative Häufigkeiten), sowie die nach Kaplan-Meier geschätzten Überlebenskurven der einzelnen Risikogruppen. Die prognostische Relevanz wurde jeweils mit dem Logrank-Test ausgewertet. Für IPI und FLIPI wurde darüber hinaus in einem multiplen Cox-Regressionsmodell die prognostische Relevanz der einzelnen Risikofaktoren auf das Gesamtüberleben untersucht. Schließlich wurden jeweils die in Abschnitt 2.5.1 näher beschriebenen Maßzahlen der prognostischen Güte (Konkordanzindex, Separationsmaß und integrierter Brier Score) berechnet und verglichen.

2.4 Entwicklung eines neuen prognostischen Index

Die Entwicklung des neuen prognostischen Index erfolgte in zwei Stufen. Zunächst wurden aus den potenziellen Prognosefaktoren die unabhängig relevanten identifiziert und damit ein Cox-Regressionsmodell geschätzt. Auf der Basis dieses Prognosemodells wurde dann ein prognostischer Index entwickelt, der die Patienten auf Grund der Werte der Prognosefaktoren in unterschiedliche Risikogruppen einteilt. Die folgenden Abschnitte beschreiben die Methoden, die hierbei zur Anwendung kamen.

2.4.1 Potenzielle Prognosefaktoren

Alle potenziellen Prognosefaktoren wurden zunächst mittels univariater Cox-Regression auf ihre prognostische Relevanz für das Gesamtüberleben überprüft. Bei metrischen Parametern mit einer stark rechtsschiefen Verteilung wurde darüber hinaus überprüft, inwiefern eine Logarithmus-Transformation zu einer symmetrischeren Verteilung führte, bezüglich derer die Regressionsparameter besser interpretiert werden könnten.

2.4.2 Regressionsmodell

Datensatz zur Modellentwicklung

Das Regressionsmodell kann nur vollständige Datensätze (complete cases) bezüglich der Zielgröße und aller potenzieller Prognosefaktoren verwenden. Ein relativ geringer Anteil fehlender Daten einzelner potenzieller Prognosefaktoren kann hierbei zu einer massiven Reduktion der Zahl vollständiger Datensätze führen. Im vorliegenden Datensatz waren erwartungsgemäß nicht alle Baseline-Charakteristika vollständig dokumentiert. Des weite-

ren wurde in Simulationsstudien [32] gezeigt, dass für eine aussagekräftige Cox-Regression mindestens ein Verhältnis 10:1 auswertbarer Ereignisse zur Zahl der Kovariaten (10 events per variable, EPV) notwendig ist. Bei Verwendung aller dokumentierter Parameter als potenzielle Prognosefaktoren wäre dieses Verhältnis deutlich unterschritten worden.

Aus diesen Gründen wurden für die multiple Cox-Regression einige potenzielle Prognosefaktoren a priori ausgeschlossen. Hierbei erfolgte die Auswahl nach dem Anteil fehlender Werte in aufsteigender Reihenfolge. Das Ergebnis der univariaten Analyse wurde bei der Auswahl nicht berücksichtigt, da es durchaus sein kann, dass ein univariat nicht signifikanter Parameter sich unter Berücksichtigung weiterer Parameter als Prognosefaktor herausstellt. Bei stark korrelierten Parametern wurde der jeweils vollständiger dokumentierte ausgewählt. Verschiedene in Abschnitt 2.7.1 beschriebene explorative Analysen sollten die Beurteilung eines möglichen Selektionsbias durch die Auswahl vollständiger Datensätze erleichtern.

Multiple Cox-Regression

Von den in die multiple Analyse eingegangenen potenziellen Prognosefaktoren wurden die voneinander unabhängigen Parameter durch Variablenselektion ermittelt und das Prognosemodell damit auf die unabhängigen Prognosefaktoren reduziert. Die Variablenselektion erfolgte rückwärts ausgehend vom vollständigen Modell mit allen potenziellen Prognosefaktoren mit der Wald-Statistik bei einem Signifikanzniveau von $\alpha = 0,05$. Zur Kontrolle wurde auch eine Vorwärts-Selektion durchgeführt sowie ein Signifikanzniveau von 0,10 verwendet.

Das endgültige Regressionsmodell war das Cox-Modell, das nach Variablenselektion nur die unabhängigen Prognosefaktoren enthielt. In diesem Modell wurden Regressionskoeffizienten und relative Risiken (Hazard-Ratios) mit 95%-Konfidenzintervallen geschätzt und p -Werte der Wald-Statistik bestimmt. Der Prognosescore war dann der lineare Prädiktor in diesem endgültigen Cox-Modell, errechnete sich also als die mit den Regressionskoeffizienten gewichtete Summe der Werte der Prognosefaktoren. Mit kontinuierlichen Prognosefaktoren war der Prognosescore ebenfalls ein kontinuierlicher Parameter, der mit dem Gesamtüberleben in der Form assoziiert war, dass ein höherer Wert des Prognosescores ein höheres Risiko bedeutete, früher zu versterben.

2.4.3 Prognostischer Index

Motivation

Das Regressionsmodell vermag einzelne Patienten bezüglich ihrer geschätzten Prognose zu vergleichen. Es definiert aber bei mindestens einer kontinuierlichen Kovariate keine und bei ausschließlich kategorialen Kovariaten in der Regel zu viele Risikogruppen. Eine

Zuteilung der Patienten zu Risikogruppen (Klassifikation) ist für die klinische Praxis jedoch von hohem Interesse, wenn zum Beispiel das Prognosemodell als Entscheidungshilfe zur Therapiewahl dienen soll. Ausgehend von dem Regressionsmodell sollte daher in der vorliegenden Arbeit ein Klassifikationsmodell (prognostischer Index) definiert werden.

Eine Möglichkeit der Klassenbildung bietet die Kategorisierung der kontinuierlichen Prognosefaktoren. Bei der Entwicklung des IPI [29] und des FLIPI [30] verwendeten die Autoren für die Regression bereits dichotomisierte Parameter mit Hilfe entweder „klinisch üblicher“ Grenzwerte oder optimaler Grenzwerte in Bezug auf die Martingalresiduen (LDH beim IPI). In einer aktuellen Publikation [42] wiesen Altman et al. darauf hin, dass eine a-priori-Kategorisierung von klinischen Parametern in der klinischen Forschung aus statistischer Sicht nicht zu empfehlen sei. Zum einen reduziere die Kategorisierung die statistische Power der Analysen, was angesichts der oft sehr knappen Fallzahlen nicht erwünscht sei. Zum anderen könne dadurch auch die Anzahl falsch positiver Ergebnisse bei Regressionsmodellen erhöht werden. Schließlich existierten für die wenigsten Parameter anerkannte Grenzwerte, die für eine Kategorisierung ohne Begründung verwendet werden könnten. Die Abhängigkeit der Zielgröße von der kontinuierlichen Einflussgröße könne durch eine Dichotomisierung nur unzureichend abgebildet werden. Die Dichotomisierung am Median der Verteilung in einer Studie führe zu unterschiedlichen Grenzwerten in unterschiedlichen Studien mit Verlust von Reproduzierbarkeit und Vergleichbarkeit der Ergebnisse. Die Suche nach „optimalen“ Grenzwerten im Sinne größter statistischer Signifikanz berge ein hohes Risiko für falsch positive Ergebnisse und eine Überschätzung der Effekte. Hierbei sind aber Korrekturen für multiples Testen möglich, z.B. nach Bonferroni oder nach Lausen [43].

Während es bei der Entwicklung eines Regressionsmodells keinen angemessenen Grund für eine Kategorisierung kontinuierlicher Einflussgrößen gibt, stellt die Bildung von Risikogruppen ausgehend von einem Regressionsmodell eine vor allem für die klinische Praxis sinnvolle Vorgehensweise dar, wie Royston et al. in ihrer aktuellen Publikation [44] darlegen. Dabei empfehlen die Autoren die Diskretisierung in der letzten Stufe der Modellbildung, nämlich ausgehend von einem kontinuierlichen Prognosescore, der in einem multiplen Regressionsmodell entwickelt wurde. Für die Wahl von Grenzwerten für den Prognosescore geben die Autoren dabei aber keine Empfehlungen.

Methoden

Bei der Identifizierung der unabhängigen Prognosefaktoren wurden gemäß den Empfehlungen von Altman et al. [42] und Royston et al. [44] die potenziellen Prognosefaktoren möglichst in dem Datentyp mit dem höchsten Informationsgehalt belassen, insbesondere wurden die kontinuierlichen Parameter als solche in das Regressionsmodell aufgenommen. Ausgehend von dem Cox-Modell mit den durch Variablenselektion identifizierten

unabhängigen Prognosefaktoren wurden dann Risikogruppen durch Kategorisierung des Prognosescores definiert.

Die Auswahl der Grenzwerte für den Prognosescore erfolgte nach der „Minimum-p-Value“-Methode, bei der die Logrank-Teststatistik für die Zielgröße Gesamtüberleben maximiert bzw. deren p -Wert minimiert wird. Zunächst wurde hierfür überprüft, inwiefern die Verteilung des Prognosescores im Datensatz, insbesondere die Anzahl Modi, einen Hinweis auf eine biologisch begründete Anzahl von Risikogruppen sowie die Lage von Grenzwerten gab. War dies nicht der Fall, z.B. auf Grund einer unimodalen Verteilung, so war die Zahl der Risikogruppen auf Grund der Linearitätsannahme für den Prognosescore im Cox-Modell letztendlich nur durch die Größe des Datensatzes nach oben beschränkt. Die lineare Abhängigkeit der logarithmierten Hazardfunktion von der Höhe des Prognosescores hat nämlich zur Folge, dass bei nicht allzu kleinen Fallzahlen zwischen beliebig definierten Risikogruppen signifikante Unterschiede im Gesamtüberleben zu sehen sein müssen.

Falls möglich, sollten drei Risikogruppen identifiziert werden, eine Gruppen niedrigen, eine mittleren und eine hohen Risikos. Eine Einteilung in mindestens drei Risikogruppen erschien für die klinische Praxis als vorteilhaft, wenn zwei Extremgruppen vorhanden sind, die möglicherweise klare Behandlungsoptionen haben, und eine oder mehrere mittlere Gruppen, bei denen auf jeden Fall weitere Faktoren zur Entscheidungsbildung mitberücksichtigt werden sollten. Eine Einteilung in mehr als drei Risikogruppen erschien allerdings für die Anwendung des prognostischen Index beispielsweise in klinischen Studien als weniger sinnvoll, da es sich beim Mantelzell-Lymphom um eine relativ seltene Erkrankung handelt, so dass mit den bislang üblichen Patientenkollektiven von um die 100 Patienten die absolute Zahl von Patienten in den einzelnen Risikogruppen unerwünscht klein würde. Dies wird auch durch die Beobachtung unterstützt, dass in den Publikationen zur prognostischen Wertigkeit des IPI beim Mantelzell-Lymphom die ursprünglich vier Risikogruppen fast immer zu drei oder zwei zusammengefasst wurden (s. Abschnitt 1.2.2). Darüber hinaus hätte trotz des im Vergleich zu den bislang publizierten Ergebnissen großen Datensatzes, der in der vorliegenden Arbeit zur Modellentwicklung zur Verfügung stand, eine Einteilung in mehr als drei Risikogruppen zwangsläufig zu kleinen absoluten Fallzahlen einzelner Risikogruppen geführt. Somit wäre die Gefahr einer Überanpassung des Modells an die vorliegenden Daten und einer statistischen Inferenz auf Grund weniger Beobachtungen sehr groß gewesen. Aus diesem Grund wurde die minimale Größe jeder Risikogruppe nach unten beschränkt. In Anbetracht eines Gesamtdatensatzes von ca. 400 Patienten mit ca. 140 Events erschien eine Risikogruppe von 10% mit 40 Patienten als zu klein, um statistisch zuverlässige Schlussfolgerungen zu ziehen. Daher sollte keine Risikogruppe weniger als 20% der Patienten umfassen. Außerdem sollte keine Risikogruppe mehr als 50% der Patienten umfassen, damit keine zahlenmäßig dominierende Gruppe entstand und eine ausreichende Differenzierung der Patienten auf Grund der Prognose

erfolgen konnte.

Unter Einhaltung dieser Randbedingungen wurden mit der „Minimum-p-Value“-Methode zwei Cutpoints für den Prognosescore aus dem endgültigen Cox-Modell gewählt. Als mögliche Cutpoints kamen Werte zwischen dem ersten und neunten Dezil der Verteilung des Prognosescores in Betracht. Zur Definition von drei Risikogruppen wurde schrittweise zunächst ein Cutpoint ausgewählt, und mit diesem festen ersten Cutpoint dann ein weiterer Grenzwert. Zur Kontrolle wurden auch sämtliche Paare von Grenzwerten unter Berücksichtigung der Gruppengrößen auf maximale Signifikanz überprüft.

Die „Minimum-p-Value“-Methode wird im Allgemeinen von statistischer Seite her stark kritisiert [42], [44], [43], [45]. Das Hauptproblem ist das des multiplen Testens bei wiederholten Signifikanztests für verschiedene Grenzwerte, das zu falsch positiven Resultaten führen kann. Lausen et al. [43] gaben eine Korrekturformel zur Adjustierung des p -Wertes für die Nullhypothese der stochastischen Unabhängigkeit von Einfluss- und Zielgröße an. Für den Prognosescore ist dagegen nach Wahl des endgültigen Regressionsmodells der Einfluss auf die Zielgröße bereits getestet und angenommen worden und es sollen nun Gruppen definiert werden. Ist die Linearitätsannahme des Cox-Modells erfüllt, so kann prinzipiell jeder Cutpoint verwendet werden, vorausgesetzt, die Fallzahlen der einzelnen Gruppen werden nicht zu klein. So wäre eine Aufteilung nach Terzilen der Verteilung des Prognosescores oder eine Aufteilung mit dem Ziel von Hazard-Ratios zwischen den Gruppen, die signifikant von 1 abweichen, wie in [44] vorgeschlagen, mögliche und sinnvolle Varianten. Mir erschien als Ziel die maximale statistische Diskrimination mit der Logrank-Statistik als Maßzahl am relevantesten, da diese Größe üblicherweise bei klinischen Studien verwendet und zur Bewertung prognostischer Indizes herangezogen wird. Zur Kontrolle wurden die p -Werte der Logrank-Statistik nach Bonferroni für multiples Testen korrigiert.

2.4.4 Vereinfachter prognostischer Index

Im Gegensatz zur klaren Haltung der Biometriker gegen eine Kategorisierung der klinischen Parameter kritisierten viele Mediziner die mangelnde praktische Handhabbarkeit eines kontinuierlichen Prognosescores mit Grenzwerten zur Einteilung der Patienten in Risikogruppen. Wie bei der Entwicklung des IPI und des FLIPI erschien den Klinikern in der Diskussion die Verwendung von dichotomisierten Einflussgrößen zur Bestimmung von Risikogruppen als einziges praktikables Mittel. In diesem Dilemma zwischen methodischer Validität und praktischer Anwendbarkeit sind bisher wenige konkrete Auswege vorgeschlagen und umgesetzt worden.

Eine mögliche Umsetzung eines von statistischer Seite optimalen Modells in die klinische Praxis besteht in der Einrichtung einer Seite im World Wide Web, auf der ein Online-Rechner zur Verfügung steht. Für den „New CML Score“ [46], der Patienten mit

chronisch myeloischer Leukämie (CML) in drei Risikogruppen einteilt, findet sich der Online-Rechner unter

<http://www.pharmacoepi.de/cgi-bin/pharmacoepi/cmlscore.cgi>.

Hier kann der behandelnde Arzt die bei Diagnose bestimmten Werte der sechs kontinuierlichen Prognosefaktoren eingeben und erhält als Ergebnis sowohl den Wert des Prognosescores als auch die Risikogruppe für den entsprechenden Patienten. Auch der neue prognostische Index für Patienten mit fortgeschrittenem Mantelzell-Lymphom wurde in dieser Form auf der Homepage des European MCL Network der Öffentlichkeit zur Verfügung gestellt unter

http://www.european-mcl.net/de/clinical_mipi.php.

Um darüber hinaus eine Möglichkeit zu bieten, die Klassifikation der Patienten ohne Taschenrechner und Internet durchführen zu können, und damit Klinikern die Bedeutung der Prognosefaktoren besser zu veranschaulichen, wurde explorativ versucht, die Zuteilung der Patienten zu Risikogruppen des neuen prognostischen Index durch Kategorisierung der Prognosefaktoren zu approximieren. Ziel war hierbei nicht die Entwicklung eines neuen Prognosemodells, sondern eine Vereinfachung der ursprünglichen Gruppenzuteilung. Zunächst wurde analog zu IPI und FLIPI eine Einteilung in Risikogruppen nach Anzahl Risikofaktoren versucht, die durch Dichotomisierung der Prognosefaktoren mit Hilfe von Standard-Grenzwerten definiert wurden. Um die Approximation an das ursprüngliche Klassifikationsmodell zu optimieren, wurde daraufhin versucht, der kontinuierlichen Natur der Prognosefaktoren durch Kategorisierung in jeweils vier Gruppen besser Rechnung zu tragen. Der mittlere Grenzwert war hierbei jeweils der Standard-Wert, und zwei weitere Grenzwerte wurden gewählt, im Falle einer schiefen Verteilung auch asymmetrisch um den mittleren Grenzwert.

So berechnete sich der vereinfachte Risikoscore durch die Summe der Risikopunkten für die einzelnen Prognosefaktoren. Damit dieses Verfahren statistisch gerechtfertigt war, sollten in einem multiplen Cox-Modell der kategorisierten Prognosefaktoren die Regressionskoeffizienten ungefähr vergleichbar sein. Unter Berücksichtigung der Gruppengrößen erfolgte dann eine Einteilung in drei Risikogruppen je nach Summe der Risikopunkte. Unter den möglichen Varianten war das Modell das beste, das den ursprünglichen prognostischen Index am besten reproduzierte. Die hierfür verwendeten statistischen Maßzahlen waren der Spearmansche Korrelationskoeffizient ρ für den Zusammenhang zwischen neuem Risikoscore und ursprünglichem Prognosescore sowie Cohen's gewichtetes κ [47] für die Übereinstimmung zwischen dem vereinfachten und dem ursprünglichen prognostischen Index.

2.5 Schätzung der Modellgüte

Bei der Entwicklung eines Prognosemodells ist eine Beurteilung der prognostischen Güte von zentraler Bedeutung. Diese Validierung erstrebt unter anderem eine Bewertung der Güte der Anpassung an die zur Modellentwicklung verwendeten Daten („goodness of fit“), eine Beurteilung der Vorhersagekraft des Modells für zukünftige Beobachtungen, sowie einen Vergleich des neuen Modells mit möglicherweise bereits vorhandenen Modellen.

Die prognostische Güte eines Modells wird meist an Hand statistischer Maßzahlen geschätzt. Welche Maßzahlen zur Verfügung stehen, hängt maßgeblich vom Datentyp der Zielgröße ab. Ist die Zielgröße metrisch und wird das Modell im Rahmen einer multiplen linearen Regression bestimmt, so ist das klassische Maß der Modellgüte das R^2 -Maß, der Anteil der durch das Modell erklärten Varianz (der vorhergesagten Werte von den beobachteten Werten) an der Gesamtvarianz.

In der vorliegenden Arbeit wiesen die Daten der Zielgröße Zensierungen auf, wenn Patienten im Rahmen der Beobachtungszeit nicht verstorben waren. Daher war für einige Patienten die wahre Überlebenszeit nicht bekannt, und konnte nicht mit der geschätzten Überlebenszeit verglichen werden. Darüber hinaus schätzt das Cox-Modell zwar den relativen Einfluss von Kovariaten auf die Überlebenszeit, nicht aber die absoluten Überlebenskurven und damit die individuell erwartete Überlebenszeit (s. [48], Abschnitt 7.). Daher sind traditionelle R^2 -Maße in diesem Fall nicht verfügbar und die Schätzung der prognostischen Güte gestaltet sich schwierig.

Verschiedene Maßzahlen für die prognostische Güte bei zensierter Zielgröße sind entwickelt worden und Graf et al. [48] geben hierüber eine umfassende Übersicht. Die in der vorliegenden Arbeit verwendeten Maßzahlen werden in Abschnitt 2.5.1 vorgestellt.

Im Rahmen der Modellvalidierung stellt sich weiter die Frage, an welchem Datensatz die prognostische Güte bestimmt werden soll. Bei der Modellentwicklung interessiert zunächst die prognostische Güte am verwendeten Datensatz. Da das Modell an die zur Entwicklung verwendeten Daten angepasst sein wird, wird diese Schätzung der prognostischen Güte als Maß für die prognostische Relevanz an zukünftigen Daten zu optimistisch sein. Dieses Phänomen wird als Überanpassung (Overfitting) bezeichnet. Mit der Komplexität eines Modells steigt im Allgemeinen die Überanpassung an die verwendeten Daten.

Um die prognostische Relevanz für zukünftige Beobachtungen zu schätzen, müssen daher andere Strategien verwendet werden. Hierbei kann man interne und externe Validierungsverfahren unterscheiden. Die interne Validierung verwendet nur die Daten der Modellentwicklung. Bei der externen Validierung schätzt man die prognostische Güte an einem möglichst völlig unabhängigen Datensatz (insb. zeitlich und örtlich). In der vorliegenden Arbeit erfolgte die interne Validierung durch ein Bootstrap-Verfahren, das von Efron [49] entwickelt wurde und im folgenden Abschnitt 2.5.2 beschrieben ist. Abschnitt 2.5.2 gibt dann eine Übersicht über weitere interne Validierungsverfahren und

eine Begründung für das gewählte Verfahren. Eine externe Validierung war an Hand des Datensatzes nicht möglich. Es ist aber geplant, die externe Validierung an prospektiven Studiendaten der französischen, holländischen und nordischen Lymphomstudiengruppen sowie der laufenden Studien des European MCL Network durchzuführen.

2.5.1 Maßzahlen der Modellgüte

Teststatistiken

Die bei der Auswahl eines Prognosemodells oft verwendeten Teststatistiken können auch als Maßzahlen für die prognostische Güte verwendet werden. Im Fall von Überlebensdaten mit Zensierungen sind dies meist die Likelihood-Ratio-Statistik für ein Regressionsmodell und die Logrank-Statistik für ein Klassifikationsmodell. In der Testtheorie geben diese Teststatistiken bzw. ihre damit bestimmten p -Werte an, wie plausibel die vorliegenden Daten unter der Nullhypothese keines Einflusses des Prognosemodells sind.

Steigung der Kalibrierungsgeraden

Ergebnis der Maximum-Likelihood-Schätzung aus generalisierten linearen Modellen und auch der Cox-Regression ist der lineare Prädiktor, der im Falle von Prognosestudien den Prognosescore darstellt. Der lineare Prädiktor ist eine Linearkombination aus den Regressionskoeffizienten und den Kovariablen des Modells, eine Zahl, die für jede Beobachtung oder jeden Patienten berechnet werden kann. Führt man die Regression statt mit den einzelnen Kovariablen mit dem Prognosescore auf demselben Datensatz durch, so ergibt sich für dessen Regressionskoeffizienten immer ein Wert von 1. Bei externer Validierung des Prognosescores auf einem unabhängigen Datensatz ist dieser Koeffizient in der Regel kleiner als 1 und wird als Steigung der Kalibrierungsgeraden bezeichnet. Je näher diese Steigung bei 1 liegt, umso besser ist das Prognosemodell auf den neuen Datensatz kalibriert.

Konkordanzindex

Der Konkordanzindex quantifiziert die Wahrscheinlichkeit, dass bei zwei beliebigen unabhängigen Beobachtungen der erwartete und der beobachtete Wert der Zielgröße konkordant sind. Konkordanz bedeutet, dass die Beobachtung mit dem größeren erwarteten Wert auch den größeren beobachteten Wert der Zielgröße hat. Bei Prognosestudien ist der erwartete Wert der mit dem Prognosemodell auf Grund von Prognosefaktoren geschätzte Wert der Zielgröße.

Im Fall einer binären, ordinalen oder metrischen Zielgröße schätzt sich der Konkordanzindex als Anteil konkordanter Paare an allen möglichen Paaren eines Datensatzes mit verschiedenem Wert der Zielgröße. Paare von Daten mit gleicher erwarteter Zielgröße

(Bindungen) werden mit dem Gewicht $\frac{1}{2}$ einbezogen. Der Konkordanzindex c wurde von Harrell in [50] für Zeit-Ereignisdaten mit Zensierungen verallgemeinert. Nach [51] schätzt sich der Konkordanzindex c als der Anteil der konkordanten Paare an allen bewertbaren Paaren. Bewertbar sind die Paare von Beobachtungen, bei denen entweder in beiden Fällen das Ereignis zu verschiedenen Zeiten eingetreten ist, oder bei denen in einem Fall das Ereignis aufgetreten ist und die Beobachtungszeit des anderen mindestens so lange ist wie die Überlebenszeit des einen. Ein bewertbares Paar ist konkordant, wenn bei der Beobachtung mit der kürzeren Überlebenszeit die erwartete Überlebenszeit ebenfalls kürzer ist als die der anderen. Bewertbare Paare von Daten mit gleicher erwarteter Überlebenszeit (Bindungen) werden als konkordante Paare mit Gewicht $\frac{1}{2}$ gezählt.

Der Konkordanzindex c ist als Wahrscheinlichkeit eine Zahl zwischen 0 und 1. Vollständige Diskordanz liegt bei einem Wert von 0,5 vor. Ein Konkordanzindex von $c = 0,5$ bedeutet, dass das Prognosemodell so gut ist wie ein Münzwurf zur Entscheidung, welche von zwei Beobachtungen die längere Überlebenszeit haben wird. Ein Konkordanzindex von 1 bedeutet vollständig korrekte Ordnung und ein Konkordanzindex von 0 vollständig falsche Ordnung der erwarteten Überlebenszeiten. Bei einem Konkordanzindex kleiner als 0,5 sollte man daher die Ordnung der erwarteten Überlebenszeit umdrehen. Nach dieser Konvention liegt dann der Konkordanzindex immer zwischen 0,5 und 1, und je größer c , umso besser die Diskriminationsfähigkeit des prognostischen Modells.

Im Fall einer binären Zielgröße, wie bei diagnostischen Tests oder bei der logistischen Regression, stellt sich c als die Fläche unter der Receiver Operating Characteristic (ROC)-Kurve dar, wie Hanley und McNeil in [52] zeigten. Weiter gilt nach [52], dass das Produkt cn_1n_2 gleich der Mann-Whitney-U-Teststatistik für den Vergleich der erwarteten Werte der Zielgröße bzw. des Prognosescores in den beiden Outcome-Gruppen der Größe n_1, n_2 ist.

Für die Berechnung des Konkordanzindex ist die absolute Bestimmung der erwarteten Überlebenszeit nicht notwendig. Es genügt, wenn für zwei Beobachtungen entschieden werden kann, bei welcher die erwartete Überlebenszeit größer ist. Das leistet zum Beispiel der lineare Prädiktor $\hat{\beta}x$ der Cox-Regression. Mit der Baseline-Überlebensfunktion $S_0(t)$ wird die erwartete Überlebenswahrscheinlichkeit in Abhängigkeit von $\hat{\beta}x$ durch $S(t, x) = S_0(t)^{\exp(\hat{\beta}x)}$ modelliert. Wegen $0 \leq S_0(t) \leq 1$ ist daher die Überlebenswahrscheinlichkeit umso größer, je kleiner der lineare Prädiktor ist. Für die mediane oder erwartete Überlebenszeit gilt der gleiche Zusammenhang.

Der Konkordanzindex c nach Harrell stellt ein leicht zu berechnendes Diskriminationsmaß für Überlebenszielgrößen dar. Er wurde in dieser Arbeit mit der Funktion „rcorr.cens“ aus dem R-Paket „Hmisc“ von Harrell zur Bewertung von Prognosescores aus der Cox-Regression berechnet. Für die Beurteilung eines Klassifikationsmodells mit wenigen Risikogruppen ist er weniger gut geeignet, weil hier die Zahl der Bindungen naturgemäß sehr groß ist. Zur Beurteilung der Vorhersagekraft des Modells im Sinne einer absoluten

Abschätzung der erwarteten Überlebenszeiten ist er eher ungeeignet. Außerdem wird kritisiert, die Methoden diagnostischer Tests auf die Prognose von Zeit-Ereignis-Daten zu übertragen, deren prospektiver Natur nicht Rechnung getragen werden kann [48].

Separationsmaß

Bei einem Klassifikationsmodell mit Überlebenszielgröße ist das Separationsmaß (SEP, [53]) ein Maß für die Auftrennung der Überlebenskurven der Risikogruppen. Er berechnet sich als mit den Fallzahlen n_i der Risikogruppen gewichtetes geometrisches Mittel der „absoluten“ relativen Risiken $\exp(|\hat{\beta}_i|)$ der g einzelnen Risikogruppen zum Gesamtkollektiv der Größe n .

$$SEP = \exp\left(\sum_{i=1}^g \frac{n_i}{n} |\hat{\beta}_i|\right)$$

Zur Berechnung der $\hat{\beta}_i$ wurde zunächst in [53] vorgeschlagen, im Datensatz die Beobachtungen der Risikogruppe i zu verdoppeln und den Cox-Regressionsschätzer dieser Gruppe gegenüber der Gesamtgruppe zu berechnen. Diese Berechnungsweise ist auf Grund abhängiger Daten methodisch problematisch [48]. Stattdessen erfolgte die Berechnung nach [48] mit Hilfe von je einer Dummyvariablen pro Risikogruppe. Die Dummyvariable d_i wurde so gewählt, dass sie über den Gesamtdatensatz den Mittelwert 0 und für die Patienten der Risikogruppe i den Wert 1 hatte. $\hat{\beta}_i$ war dann der Regressionsschätzer der Dummyvariablen d_i im univariaten Cox-Modell.

Nach Definition ist SEP immer ≥ 1 . Im ungünstigsten Fall identischer Überlebenskurven errechnet sich SEP zu 1. Je weiter SEP von 1 abweicht, umso stärker die Auftrennung der Risikogruppen in Bezug auf die Zielgröße. Die Gewichtung mit der Fallzahl der Risikogruppen trägt der Anforderung an ein Prognosemodell Rechnung, dass die Risikogruppen einen substanziellen Anteil von Patienten enthalten sollten. Die Berechnung des SEP setzt eine Einteilung in Risikogruppen voraus. Es gelten alle Voraussetzungen des Cox-Modells, also die Annahme zeitunabhängiger Risiken und der Linearität. Das Separationsmaß ist ein reines Diskriminationsmaß und gibt keine absolute Information über die Vorhersagekraft des Prognosemodells.

Brier Score

Der Brier Score [54] wurde 1950 von Brier für die Beurteilung der Prognosegüte von Wettervorhersagen entwickelt. Im Kontext einer logistischen Regression mit binärer Zielgröße ist er das Maß der quadratischen Abweichung der erwarteten Wahrscheinlichkeit für das Eintreten des Zielereignis $\hat{P}(Y = 1)$ vom tatsächlichen Auftreten ($Y = 1$). Er wird nach der folgenden Formel geschätzt.

$$BS = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{P}(Y_i = 1))^2$$

Der Brier Score ist immer positiv und als Maß der quadratischen Abweichung sollte er möglichst klein sein. Ist die erwartete Wahrscheinlichkeit für das Auftreten des Ereignisses $\frac{1}{2}$, so hat der Brier Score den Wert $\frac{1}{4}$.

Zur Beurteilung eines Prognosemodells kann die erwartete Auftretenswahrscheinlichkeit $\hat{P}(Y = 1|X)$ in Abhängigkeit von einem Kovariatenvektor X geschätzt werden, der so in die Berechnung des Brier Scores eingeht.

$$BS = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{P}(Y_i = 1|X_i))^2$$

Im Kontext der Analyse von Zeit-Ereignis-Daten ohne Zensierungen kann der Brier Score für jeden Zeitpunkt t berechnet werden, wobei $Y = 1$ so definiert wird, dass das Zielereignis bis zum Zeitpunkt t noch nicht eingetreten ist. Mit der Indikatorfunktion I , die den Wert 1 annimmt, wenn das Argument wahr ist, und ansonsten 0, und den Ereigniszeitpunkten T_i berechnet sich der Brier Score für den Zeitpunkt t dann als

$$BS(t) = \frac{1}{n} \sum_{i=1}^n (I(T_i > t) - \hat{P}(T_i > t|X_i))^2.$$

Um ein zeitunabhängiges Maß zu erhalten, kann der Brier Score über die Zeit integriert werden.

Graf et al. [48] entwickelten schließlich eine Möglichkeit, Zensierungen in die Berechnung des Brier Scores einzubeziehen. Das Hauptproblem besteht darin, dass für zensierte Beobachtungen bei Zeitpunkten nach der Zensierung nicht entschieden werden kann, ob das Ereignis bis dahin eingetreten ist. Für diese Beobachtungen kann der Summand im Brier Score nicht berechnet werden. Stattdessen werden die Beiträge der unzensierten Beobachtungen und der nach dem entsprechenden Zeitpunkt zensierten Beobachtungen gewichtet. Die Gewichtung erfolgt mit dem Kehrwert $1/\hat{G}(t)$ der nach Kaplan-Meier geschätzten Zensierungsverteilung $\hat{G}(t)$ im Gesamtkollektiv. Beiträge unzensierter Beobachtungen werden mit $\min(1/\hat{G}(T_i), 1/\hat{G}(t))$ und Beiträge nach t zensierter Beobachtungen mit $1/\hat{G}(t)$ gewichtet.

Zur Berechnung des Brier Scores müssen die erwarteten Wahrscheinlichkeiten $\hat{P}(T_i > t|X_i)$ für das Auftreten des Zielereignisses in Abhängigkeit von Kovariaten geschätzt werden. Bei der Beurteilung eines Klassifikationsschemas, bei dem Beobachtungen in Risikogruppen eingeteilt werden, kann man dafür einfach eine Kaplan-Meier-Schätzung für jede Risikogruppe durchführen. Im Fall der Cox-Regression muss eine Schätzung der Baseline-Überlebensfunktion erfolgen, oder die Anpassung eines parametrischen Modells.

2.5.2 Das Bootstrap-Verfahren

Die Schätzung der prognostischen Güte des neuen Modells wurde in der vorliegenden Arbeit in Anlehnung an Steyerberg et al. [55] mit einem Bootstrap-Verfahren durchgeführt.

Die Maßzahlen Logrank- bzw. Likelihood-Ratio-Statistik, Steigung der Kalibrierungsgeraden, Konkordanzindex c , Separationsmaß SEP und Brier Score wurden hierbei bootstrap-korrigiert geschätzt. Bei dem Verfahren wird zunächst die überoptimistische Schätzung am Datensatz der Modellentwicklung bestimmt. Die Differenz dieser Schätzung zum theoretischen Erwartungswert an einem unabhängigen zukünftigen Datensatz, der sog. Optimismus, wird dann unter Verwendung von Bootstrap-Stichproben nichtparametrisch geschätzt. Schließlich wird der überoptimistisch geschätzte Wert um den geschätzten Optimismus korrigiert. Bootstrap-Stichproben sind Stichproben der gleichen Größe wie die Originalstichprobe, die aus der Originalstichprobe durch Ziehen mit Zurücklegen entstehen. In der vorliegenden Auswertung wurden jeweils $B = 1000$ Bootstrap-Stichproben gezogen. Die Schritte des Bootstrap-Verfahrens sind im Einzelnen:

1. Modellentwicklung am Originaldatensatz,
2. (überoptimistische) Schätzung der Modellgüte am Originaldatensatz,
3. Ziehung von B Bootstrap-Stichproben,
4. Modellentwicklung an jeder Bootstrap-Stichprobe,
5. für jede Bootstrap-Stichprobe Schätzung der Modellgüte an der Bootstrap-Stichprobe und am Originaldatensatz,
6. für jede Bootstrap-Stichprobe Berechnung der Differenz der beiden Werte der Modellgüte,
7. Mittelung der Differenz über die B Bootstrap-Stichproben (Optimismus),
8. Korrektur der überoptimistischen Schätzung um den Optimismus.

Theorie des Bootstrap-Verfahrens

Bootstrap-Verfahren geben in ganz allgemeinen Situationen die Möglichkeit zur nichtparametrischen Maximum-Likelihood-Schätzung und sind in [56] ausführlich beschrieben. Im Folgenden werden die wichtigsten Prinzipien der Bootstrap-Verfahren dargestellt.

Besteht die Möglichkeit, aus einer unbekanntem Verteilung beliebig viele Zufallsstichproben zu ziehen, so kann man den unbekanntem Erwartungswert durch den Mittelwert einer Zufallsstichprobe oder durch den über mehrere Zufallsstichproben gemittelten Mittelwert schätzen. Das Bootstrap-Verfahren bietet diese Möglichkeit mit nur einer Zufallsstichprobe aus einer unbekanntem Verteilung. Hierbei approximiert die empirische Verteilungsfunktion der Stichprobe die Verteilung der Grundgesamtheit. Aus dieser empirischen Verteilungsfunktion kann man nun beliebig viele Stichproben (Bootstrap-Stichproben) ziehen, an Hand derer man die gesuchten Parameter wie oben beschrieben schätzt. Der

Schätzwert des gesuchten Parameters ist dann der Mittelwert über alle B Bootstrap-Stichproben und der Standardfehler wird als Standardabweichung der Schätzwerte in den B Bootstrap-Stichproben angegeben. Das Ziehen von Stichproben aus der empirischen Verteilungsfunktion der Originalstichprobe ist dasselbe wie das Ziehen mit Zurücklegen aus dem ursprünglichen Datensatz. Meist werden Bootstrap-Stichproben derselben Länge wie die Original-Stichprobe gezogen. Das Prinzip, nach dem die empirische Verteilungsfunktion \hat{F} der Daten die wahre Verteilungsfunktion F der Grundgesamtheit und die Bootstrap-Stichprobe x^* die Original-Stichprobe x ersetzt, nennt man auch das „plug-in principle“ beim Bootstrap-Verfahren. Wie viele Bootstrap-Stichproben gezogen werden sollen, kann man nicht theoretisch begründen; empfohlen werden generell mindestens 100.

Efron stellte 1983 in [49] verschiedene Bootstrap-Verfahren zur Schätzung des Prädiktionsfehlers eines Klassifikationsmodells vor. Die verwendete nichtparametrische Bootstrap-Schätzung des Prädiktionsfehlers eines Prognosemodells, der „regular bootstrap“ wird nun im Einzelnen in Anlehnung an [49] beschrieben.

Mit $X = (C, Y)$ seien Einflussgröße C und Zielgröße Y mit Verteilung F bezeichnet, für die der Datensatz eine Stichprobe $x_i = (c_i, y_i), i \in \{1, \dots, n\}, x = (x_1, \dots, x_n)$ darstellt. Mit $x_0 = (c_0, y_0)$ sei eine beliebige zukünftige Beobachtung aus derselben Verteilung F bezeichnet. Das am Datensatz x entwickelte Prognosemodell η_x schätze den erwarteten Wert der Zielgröße durch $\hat{y}_0 = \eta_x(c_0)$. Der Prädiktionsfehler für x_0 sei mit $Q(y_0, \hat{y}_0) = Q(y_0, \eta_x(c_0))$ als Maß für die Abweichung des durch das Modell vorhergesagten Wertes vom wahren Wert bezeichnet. Der zu schätzende Prädiktionsfehler des Modells errechnet sich dann als Erwartungswert des Prädiktionsfehlers über beliebige zukünftige Beobachtungen

$$err(x, F) = E_{x_0 \sim F}(Q(y_0, \eta_x(c_0))), \quad (2.1)$$

wobei $E_{x_0 \sim F}$ darauf hinweist, dass der Erwartungswert über zukünftige Beobachtungen x_0 mit Verteilung F geht. Die überoptimistische Schätzung des Prädiktionsfehlers am Datensatz der Modellentwicklung lässt sich schreiben als der mittlere Prädiktionsfehler für die Beobachtungen des Original-Datensatzes

$$err(x, \hat{F}) = E_{x_0 \sim \hat{F}}(Q(y_0, \eta_x(c_0))) = \frac{1}{n} \sum_{i=1}^n Q(y_i, \eta_x(c_i)). \quad (2.2)$$

Der sog. Optimismus wird nun als Erwartungswert der Differenz zwischen dem wahren und dem überoptimistisch geschätzten Prädiktionsfehler definiert

$$\omega(F) = E_{x \sim F}(err(x, F) - err(x, \hat{F})). \quad (2.3)$$

Nach dem plug-in principle ist der Bootstrap-Schätzer des wahren Optimismus

$$\omega(\hat{F}) = E_{x^* \sim \hat{F}}(err(x^*, \hat{F}) - err(x^*, \hat{F}^*)), \quad (2.4)$$

der approximiert wird durch Mittelung über B Bootstrap-Stichproben

$$\begin{aligned}\hat{\omega}(\hat{F}) &= \hat{E}_{x^* \sim \hat{F}}(err(x^*, \hat{F}) - err(x^*, \hat{F}^*)) = \frac{1}{B} \sum_{b=1}^B (err(x_b^*, \hat{F}) - err(x_b^*, \hat{F}_b^*)) \\ &= \frac{1}{B} \sum_{b=1}^B \left(\frac{1}{n} \sum_{i=1}^n Q(y_i, \eta_{x_b^*}(c_i)) - \frac{1}{n} \sum_{i=1}^n Q(y_{i,b}^*, \eta_{x_b^*}(c_{i,b}^*)) \right),\end{aligned}\quad (2.5)$$

wobei $x_b^* = (c_b^*, y_b^*)$ die b -te Bootstrap-Stichprobe bezeichnet, die durch Ziehen mit Zurücklegen aus x entsteht, \hat{F}_b^* die empirische Verteilungsfunktion von x_b^* , und B die Anzahl Bootstrap-Stichproben. Der Term $err(x_b^*, \hat{F})$ ist der auf der Original-Stichprobe geschätzte Prädiktionsfehler des Modells, das auf der b -ten Bootstrap-Stichprobe entwickelt wurde. Der Term $err(x_b^*, \hat{F}_b^*)$ ist der überoptimistisch geschätzte Prädiktionsfehler des Modells, das auf der b -ten Bootstrap-Stichprobe entwickelt wurde.

Mit dem so geschätzten Optimismus kann man nun den wahren Prädiktionsfehler bootstrap-korrigiert schätzen nach

$$\widehat{err}(x, F) = err(x, \hat{F}) + \hat{\omega}(\hat{F}). \quad (2.6)$$

Neben der Schätzung des wahren Prädiktionsfehlers können auch andere Maßzahlen der Modellgüte mit den genannten Validierungsverfahren geschätzt werden. Ein solches Vorgehen wird z.B. von Harrell in [51] empfohlen. Das Validierungsverfahren bewertet einerseits, in welchem Maße die verwendeten Verfahren der Modellentwicklung eine Überanpassung an die Daten erlaubten. Andererseits ermöglichen die damit geschätzten Maßzahlen der Modellgüte einen unverzerrteren Vergleich zu bereits bekannten Prognosemodellen, die an denselben Daten extern validiert werden und daher kein Overfitting aufweisen.

Weitere Bootstrap-Verfahren

Die einfachste Möglichkeit für ein Bootstrap-Verfahren zur Schätzung des Prädiktionsfehlers wäre, für jede Bootstrap-Stichprobe den Prädiktionsfehler des an der Bootstrap-Stichprobe entwickelten Modells an der Originalstichprobe zu validieren und den mittleren Prädiktionsfehler über alle Bootstrap-Stichproben zu mitteln:

$$\hat{E}_{\hat{F}}(err(x^*, \hat{F})) = \frac{1}{B} \sum_{b=1}^B \sum_{i=1}^n Q(y_i, \eta_{x_b^*}(c_i)) / n. \quad (2.7)$$

Nach [56] unterschätzt man mit diesem Vorgehen den wahren Prädiktionsfehler, da das zu untersuchende Modell, das mit den Daten x entwickelt wurde, in (2.7) nicht fixiert ist, sondern jeweils an einer Bootstrap-Stichprobe entwickelt wird.

Efron führte in [49] des Weiteren den .632-Bootstrap-Schätzer ein. Hierbei liegt die Idee zu Grunde, dass bei der Schätzung des Prädiktionsfehlers des an der Bootstrap-

Stichprobe entwickelten Modells an der Original-Stichprobe ein Teil der Daten sowohl zur Modellentwicklung als auch zur Schätzung des Prädiktionsfehlers verwendet werden. Der bootstrap-korrigierte Schätzer des Prädiktionsfehlers nach (2.6) könnte daher den wahren Prädiktionsfehler unterschätzen. Die Wahrscheinlichkeit einer Beobachtung der Original-Stichprobe, in die Bootstrap-Stichprobe gezogen zu werden, beträgt ca. 63,2%. Mit ϵ_0 , dem mittleren Prädiktionsfehler des an einer Bootstrap-Stichprobe entwickelten Modells an den Daten der Originalstichprobe, die nicht in der Bootstrap-Stichprobe enthalten sind, wird der .632-Bootstrap-Schätzer des Optimismus definiert als

$$\hat{\omega}^{.632} = 0,632(\epsilon_0 - err(x, \hat{F})) \quad (2.8)$$

und damit der .632-Schätzer des wahren Prädiktionsfehlers als

$$\widehat{err}(x, F)^{.632} = err(x, \hat{F}) + 0,632(\epsilon_0 - err(x, \hat{F})). \quad (2.9)$$

Die Motivation für die Definition des .632-Schätzers ist, wie die Autoren schreiben, heuristisch. Daher gibt es bis heute kein theoretisches Argument für die Überlegenheit des .632-Schätzers über den einfachen Schätzer. In mehreren Simulationen schien jedoch dieser Schätzer in einigen sehr kleinen Datensätzen überlegen zu sein. Die vorgestellten Bootstrap-Schätzer sind einander jedoch asymptotisch äquivalent [56]. Efron et al. führten darüber hinaus eine weitere Modifikation, den .632+ - Schätzer [57], auf den hier nicht weiter eingegangen werden soll.

Weitere mögliche Verfahren der Modellvalidierung

Da die einfache Schätzung der Modellgüte am Datensatz der Modellentwicklung überoptimistisch ist, wurden verschiedene Strategien zur Verbesserung eingeführt. In der medizinischen Forschung sehr beliebt ist die zufällige Aufteilung des Datensatzes in zwei Teile (random data split). Auf dem sog. Trainingsdatensatz wird das Modell entwickelt, dessen Güte auf dem sog. Validierungsdatensatz geschätzt wird. Dieses Verfahren ist vor allem dann sinnvoll, wenn genügend Daten zur Verfügung stehen. Im Allgemeinen stellt die Aufteilung des Datensatzes einen Powerverlust sowohl für die Modellentwicklung als auch für die Validierung dar (s. auch [58], [51]). Besser als eine zufällige Aufteilung der Daten sei nach [51] eine zeitliche oder örtliche. Auch dieses Verfahren stellt eine interne Validierung dar und kann die externe Validierung an einem unabhängigen Datensatz nicht ersetzen.

Die sog. Kreuzvalidierung stellt ein internes Validierungsverfahren zur unverzerrten Schätzung des wahren Prädiktionsfehlers dar [49], [56]. Hierbei wird der Datensatz zufällig in k Teile zerlegt, für den k -ten Teil das Modell auf den Daten ohne diesen Teil entwickelt und auf dem k -ten Teil validiert. Der Wert der prognostischen Güte wird schließlich über die k Werte gemittelt. Typischerweise wird $k = 10$ gewählt. Die Kreuzvalidierung

wurde in Simulationsstudien als unverzerrt bestätigt, allerdings mit hoher Variabilität der Schätzung. Bootstrap-Verfahren erwiesen sich in Simulationen der Kreuzvalidierung in der Regel als überlegen [49], [56], [57]. Daher wurde das Bootstrap-Verfahren für die interne Validierung gewählt.

Steyerberg et al. [55] verglichen in einer Simulationsstudie die verschiedenen internen Validierungsverfahren für ein Prognosemodell, das mit logistischer Regression entwickelt wurde. Sie verwendeten als Maßzahlen den Konkordanzindex c nach Harrell, die Steigung der Kalibrierungsgeraden, den Brier Score, die skalierte χ^2 -Statistik des Modells, sowie Nagelkerke's R^2 -Maß. Die Ergebnisse zeigten eine Überlegenheit der Bootstrap-Verfahren gegenüber der überoptimistischen Schätzung, dem Datensplit und der Kreuzvalidierung. Unter den Bootstrap-Verfahren schnitt der einfache Bootstrap nicht schlechter ab als die 632 und .632+ Varianten. Darüber hinaus zeigten die Ergebnisse eine starke Abhängigkeit des Overfitting von der Fallzahl bzw. der Zahl von Ereignissen pro Variable (EPV). Bei EPV-Werten von 5, 10 oder 20 zeigten die im Datensatz der Modellentwicklung geschätzten Maßzahlen der Modellgüte eine deutliche Überschätzung mit einer hohen Variabilität. Bei EPV von 40 oder 80 waren die Abweichungen der verschiedenen Validierungsverfahren minimal bei deutlich reduzierter Variabilität.

2.6 Überprüfung der Modellannahmen

Die im Cox-Modell vorausgesetzten Annahmen wurden mit Hilfe von Residuen für das endgültige Cox-Regressionsmodell überprüft. Die überprüften Annahmen waren einerseits die zeitliche Unabhängigkeit der relativen Risiken bei fester Kovariablen-Konstellation (Proportional-Hazards-Annahme) sowie die lineare Abhängigkeit der logarithmierten relativen Risiken von den Kovariaten. Darüber hinaus wurden Beobachtungen mit besonders starkem Einfluss auf die Regressionsparameter identifiziert.

2.6.1 Proportional-Hazards-Annahme

Zur grafischen Überprüfung der Proportional-Hazards-Annahme wurden die skalierten Schoenfeld-Residuen [59], [60] der Prognosefaktoren des endgültigen Cox-Modells gegen die Zeit geplottet und jeweils eine Glättung mit Splines durchgeführt. Bei erfüllter Proportional-Hazards-Annahme sollten die geglätteten Kurven annähernd konstant sein. Die Berechnung der skalierten Schoenfeld-Residuen und die Glättung erfolgten mit der Funktion „cox.zph“ des R-Pakets „survival“. Darüber hinaus wurde der in Therneau et al. [60] dargestellte statistische Test auf Abweichungen von der Proportional-Hazards-Annahme mit der Funktion „cox.zph“ des R-Pakets „survival“ durchgeführt.

2.6.2 Linearität

Zur Überprüfung der Linearitätsannahme im Cox-Modell wurden Martingalresiduen nach Therneau et al. [61] zu Hilfe genommen. Ohne Kenntnis der funktionalen Form für die Abhängigkeit der logarithmierten Hazard-Quotienten von den Kovariaten wurden die Martingalresiduen für das Modell ohne Kovariaten berechnet. Zur grafischen Darstellung wurden die Martingalresiduen gegen die in Betracht gezogenen Kovariaten geplottet. Die durch LOWESS-Glättung geschätzte Kurve gibt einen Anhaltspunkt für die funktionale Abhängigkeit zwischen dem Wert der Kovariaten und den logarithmierten Risikoquotienten. Bei erfüllter Linearität sollte die geglättete Kurve daher annähernd linear sein. Die Notwendigkeit eventueller Variablentransformationen (Logarithmierung) wurde ebenfalls mit Hilfe der Martingalresiduen bewertet. Die Berechnung der Martingalresiduen erfolgte mit der Funktion „residuals.coxph“ des R-Pakets „survival“, die LOWESS-Glättung mit der Funktion „lowess“ des R-Pakets „stats“.

2.6.3 Extrembeobachtungen

Zur Analyse des Einflusses von Extrembeobachtungen auf die Regressionsparameter und damit auf die Stabilität des Cox-Modells wurden die Scorerresiduen nach Therneau et al. [61] im endgültigen Modell berechnet und für jede Kovariate gegen deren Wert geplottet. Die Scorerresiduen geben für jede Beobachtung und jede Kovariate an, wie sich der Wert des Regressionsparameter für die entsprechende Kovariate prozentual ändert, wenn die jeweilige Beobachtung aus dem betrachteten Cox-Modell weggelassen wird. Positive Scorerresiduen bedeuten hierbei eine prozentuale Verkleinerung des Schätzwertes für den jeweiligen Regressionsparameter. Durch Schätzung der Scorerresiduen konnten mögliche Datenfehler identifiziert und der Einfluss von Extrembeobachtungen auf das Regressionsmodell beurteilt werden. Die Berechnung der Scorerresiduen erfolgte ebenfalls mit der Funktion „residuals.coxph“ des R-Pakets „survival“.

2.7 Weitere explorative Analysen

2.7.1 Effekt der Auswahl vollständiger Datensätze

Das neue Prognosemodell wurde auf einem reduzierten Datensatz von Patienten mit multivariat vollständigen Daten zu einer Auswahl von univariat fast vollständigen Parametern entwickelt. Um einen möglichen Selektionsbias bei dieser Auswahl zu entdecken, wurde zunächst mittels Kaplan-Meier-Schätzung und Logrank-Test überprüft, wie sich die Überlebenskurven für das Gesamtüberleben der Patienten mit vollständigen Daten von denen mit unvollständigen Daten unterscheiden. Darüber hinaus wurde die univariate Cox-Regression für die wichtigsten potenziellen Prognosefaktoren mit dem für die

multiple Regression verwendeten Datensatz wiederholt.

Die bei der multiplen Cox-Regression nicht verwendeten potenziellen Prognosefaktoren wurden auf ihre zu dem entwickelten Prognosemodell eventuelle zusätzliche prognostische Relevanz überprüft. Das Cox-Modell dieser Analysen enthielt die Prognosefaktoren des neu entwickelten Modells und jeweils einen der weggelassenen potenziellen Prognosefaktoren. Zur Kontrolle eines möglichen Selektionsbias, der durch Reduktion auf die bezüglich des bislang weggelassenen Prognosefaktors vollständigen Datensätze eintreten könnte, wurde zusätzlich ein Cox-Modell auf diesen vollständigen Datensätzen mit den Prognosefaktoren des neuen Prognosemodells geschätzt.

2.7.2 Wechselwirkungen

Um das Modell möglichst einfach zu halten, wurden Wechselwirkungen zwischen den potenziellen Prognosefaktoren für die Modellentwicklung nicht berücksichtigt. Die paarweisen Wechselwirkungen zwischen den Prognosefaktoren des endgültigen Modells wurden dann mittels Cox-Regression analysiert.

2.7.3 Alter und Allgemeinzustand

Da ein höheres Alter und ein schlechterer Allgemeinzustand meist Kontraindikationen gegen aggressivere, experimentelle Therapien darstellen, welche für die Hochrisikogruppe des Prognosemodells in Betracht gezogen werden können, wurden verschiedene explorative Analysen zur Bedeutung dieser beiden Parameter durchgeführt. Im Falle einer Auswahl dieser beiden Parameter als unabhängige Prognosefaktoren wurde zum einen mittels multipler Cox-Regression mit Rückwärts-Selektion überprüft, ob bei Weglassen dieser beiden so genannten Host-Charakteristika andere Prognosefaktoren an deren Stelle treten würden. Darüber hinaus wurde die Relevanz des neuen Klassifikationsmodells in der Gruppe von Patienten unter 65 Jahren überprüft. Dafür wurden zunächst Kaplan-Meier-Kurven für das Gesamtüberleben in den Risikogruppen geschätzt und mit dem Logrank-Test verglichen, sowie die prognostische Relevanz der Prognosefaktoren in einem multiplen Cox-Modell geschätzt. Schließlich wurde explorativ auch die Rückwärts-Selektion im Cox-Modell für die Untergruppe dieser jüngeren Patienten wiederholt.

2.7.4 Zeit bis zum Therapieversagen

Für das neu entwickelte Klassifikationsmodell wurden auch Überlebenskurven für die Zeit bis zum Therapieversagen nach Risikogruppe geschätzt, um die prognostische Bedeutung des Modells für diesen stark therapieabhängigen Parameter zu beurteilen. Des Weiteren wurde die Bedeutung der Prognosefaktoren des neuen Modells auf die Zeit bis zum Therapieversagen in einem multiplen Cox-Modell untersucht.

Kapitel 3

Ergebnisse

3.1 Beschreibung des Patientenkollektivs

3.1.1 Patientenzahl

Zwischen Mai 1996 und Oktober 2004 waren 492 Patienten mit Mantelzell-Lymphom in eine der Primärtherapiestudien "CHOP vs. MCP" und "CHOP vs. R-CHOP" der GLSG oder in die Europäische MCL-Studie eingebracht worden. Bei 19 Patienten war auf Grund kurz zurück liegender Meldung noch keine Dokumentation verfügbar, 8 Patienten waren lost-to-follow-up. Somit waren Daten von 465 Patienten verfügbar, von denen 455 Patienten ein dokumentiertes klinisches Stadium III oder IV nach Ann Arbor hatten. Bei 6 Patienten war das Stadium nicht dokumentiert, ein Patient war im Stadium I und drei Patienten hatten ein Stadium II.

3.1.2 Baselinecharakteristika

Das mediane Alter der Patienten bei Studienaufnahme lag bei 60 Jahren (Spannweite 34 bis 86 Jahre), 76% der Patienten waren männlich und 91% waren weder bettlägerig noch arbeitsunfähig (ECOG Performance Status 0 oder 1). Im Stadium IV nach Ann Arbor waren 84% der Patienten, 43% zeigten B-Symptome und bei 54% lag ein Milzbefall vor. Bei 79% der Patienten war das Knochenmark befallen und 32% hatten mehr als einen extranodalen Befall. Die mediane Anzahl befallener Lymphknotenregionen war 8 (0 - 11) und die mediane maximale Lymphknotengröße 4cm (0,5cm - 24cm); bei den zwei Patienten mit reinem Knochenmarksbefall war die maximale Lymphknotengröße auf 0,5cm festgesetzt worden. Die LDH war bei 32% der Patienten erhöht, die mediane Leukozytenzahl lag bei 7.900/ μ l (Spannweite 1.000/ μ l - 764.000/ μ l), und die mediane Lymphozytenzahl bei 2.007/ μ l (141/ μ l - 62.5869/ μ l). Die mediane Thrombozytenzahl war 188.000/ μ l (3.000/ μ l - 1.346.000/ μ l), der mediane Hämoglobinwert bei den männlichen Patienten 133g/l (55g/l - 175g/l) und bei den Patientinnen 124g/l (30g/l - 149g/l). Tabelle 3.1 zeigt die klinischen

Parameter	n	% fehlend	Wert	Häufigkeit (%) bzw. Median (Min. - Max.)
Alter	455	0	Jahre	60 (34 - 86)
			≥ 60 Jahre	240 (53)
Geschlecht	455	0	w	111 (24)
			m	344 (76)
Ann-Arbor-Stadium	455	0	III	71 (16)
			IV	384 (84)
B-Symptomatik	451	1	ja	196 (43)
Milzbefall	420	8	ja	225 (54)
extranodale Befälle	445	2	0	53 (12)
			1	249 (56)
			2	111 (25)
			3	26 (6)
			4	6 (1)
Knochenmarksbefall	453	0	ja	360 (79)
Lymphknotenregionen	432	5	Zahl	8 (0 - 11)
maximale Lymphknotengröße	387	15	cm	4 (0,5 - 24)
ECOG Performance Status	452	1	0	147 (33)
			1	263 (58)
			2	38 (8)
			3	3 (1)
			4	1 (0)
LDH-Quotient*	443	3		0,86 (0,15 - 5,3)
LDH-Wert			> obere Norm	142 (32)
Leukozyten	451	1	Zahl/ μ l	7.900 (1.000 - 764.000)
Lymphozyten	422	7	Zahl/ μ l	2.007 (141 - 625.860)
Granulozyten	413	9	Zahl/ μ l	4.200 (187 - 26.352)
Monozyten	408	10	Zahl/ μ l	491 (14 - 10.900)
Thrombozyten	452	1	Zahl/ μ l	188.000 (3.000 - 1.346.000)
Hb - Frauen	109	2	g/l	124 (30 - 149)
Hb - Männer	342	1	g/l	133 (55 - 175)
Albumin-Quotient*	268	41		0,80 (0,36 - 1,26)
β_2 -Mikroglobulin-Quotient*	285	37		1,06 (0,06 - 8)
IgA	286	37	g/l	2,0 (0,1 - 84,5)
IgG	289	36	g/l	10,55 (0,86 - 82,3)
IgM	287	37	g/l	0,77 (0,04 - 11,1)

Tabelle 3.1: Deskription der Baseline-Charakteristika. n: Anzahl nicht fehlender Werte.
* Quotienten zum oberen Normalwert

Charakteristika der Patienten bei Studienaufnahme im Einzelnen.

Abbildung 3.1 zeigt Histogramme für die metrischen Baselinecharakteristika. Die metrischen Parameter LDH-Quotient, Leukozyten-, Granulozyten-, Lymphozyten-, Monozytenzahl, β_2 -Mikroglobulin, sowie die Immunglobuline hatten stark rechtsschiefe Verteilungen, die durch Logarithmustransformation deutlich symmetrischer wurden (Abbildung 3.2). Abbildung 3.3 zeigt Streudiagramme für ausgewählte metrische Parameter. Die Leukozytenzahl war erwartungsgemäß stark korreliert mit der Lymphozytenzahl ($r = 0,83$), und ebenfalls korreliert mit Granulozytenzahl ($r = 0,50$) und Monozytenzahl ($r = 0,53$). Keine wesentlichen Korrelationen bestanden zwischen Leukozytenzahl einerseits und Thrombozytenzahl ($r = 0,01$), Hb ($r = -0,25$), LDH-Quotient ($r = 0,19$) bzw. Alter ($r = -0,02$) andererseits. Auch Alter und LDH-Quotient waren nicht korreliert ($r = 0,02$).

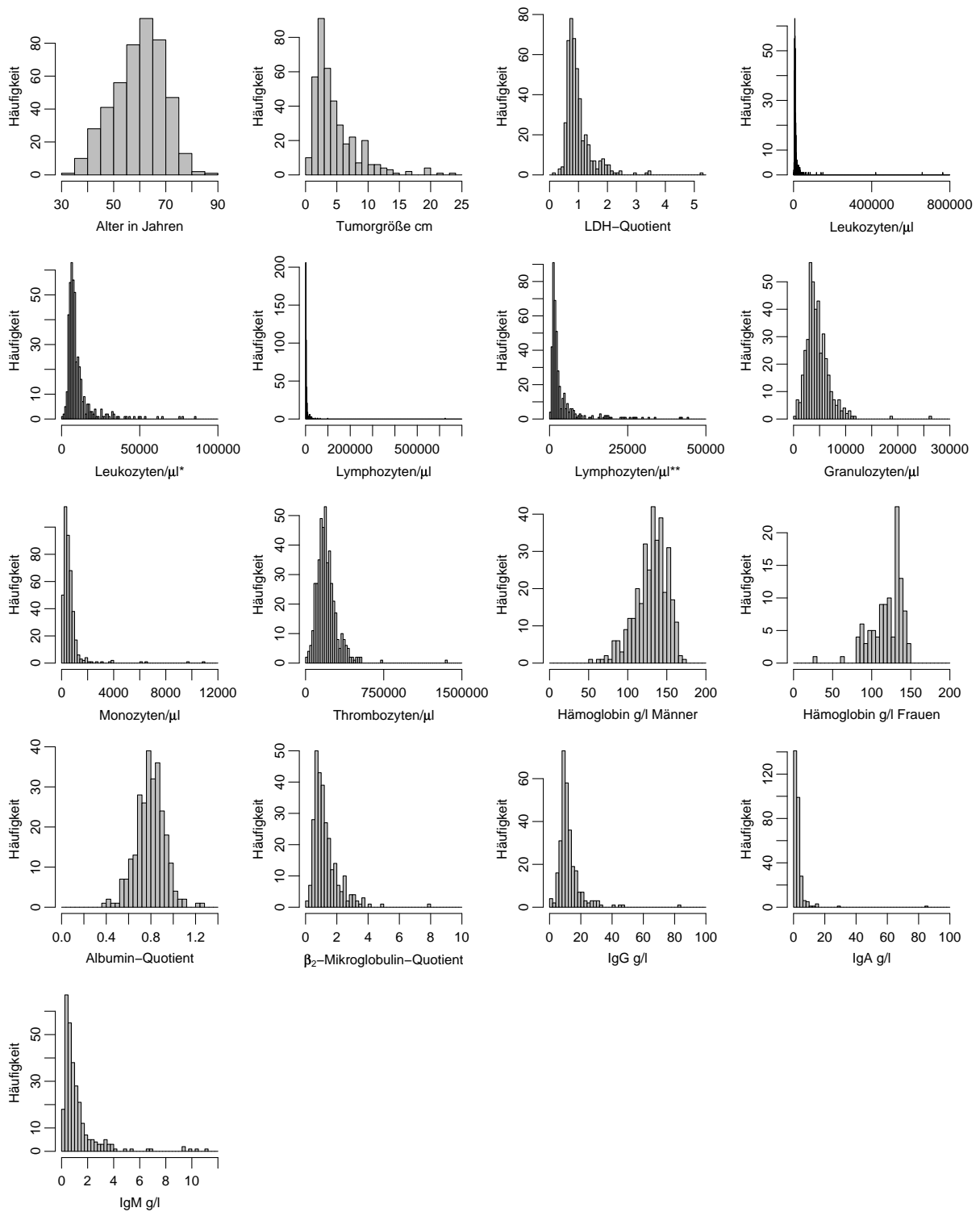


Abbildung 3.1: Verteilung der metrischen klinischen Charakteristika.

* unter Ausschluss der Werte über $100.000/\mu\text{l}$

** unter Ausschluss der Werte über $50.000/\mu\text{l}$

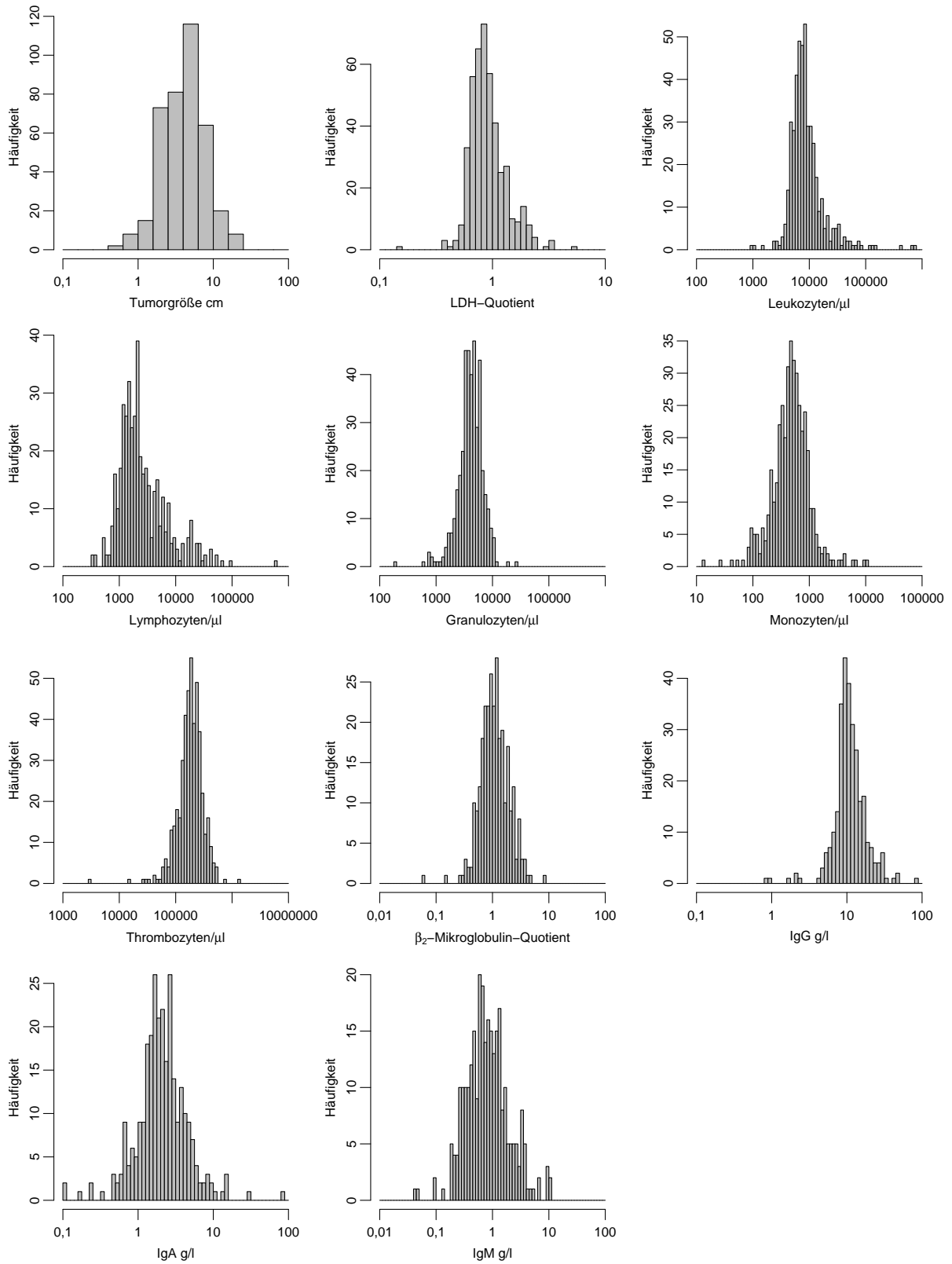


Abbildung 3.2: Verteilung der logarithmierten klinischen Charakteristika.

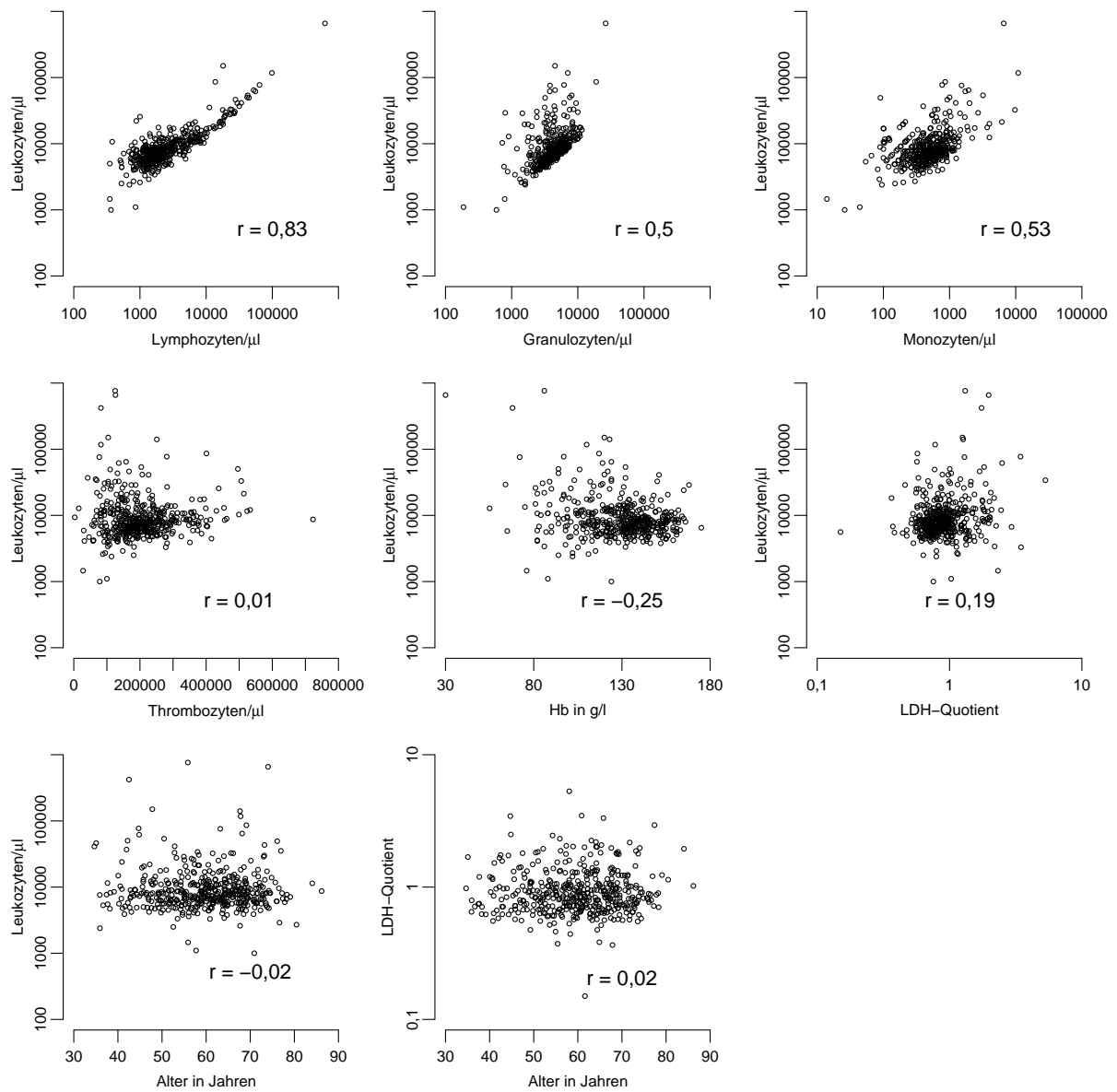


Abbildung 3.3: Korrelation metrischer klinischer Charakteristika.

3.1.3 Therapie und Therapieerfolg

Die überwiegende Mehrheit der Patienten (56%) war initial mit CHOP behandelt worden, 11% der Patienten mit MCP und 2% mit anderen Chemotherapieschemata. Die Kombination R-CHOP mit dem Antikörper Rituximab hatten 31% der Patienten erhalten. Das Ansprechen auf die initiale Therapie war bei 438 Patienten auswertbar. Davon hatten 80 Patienten eine komplette Remission (CR-Rate 18%) und 351 mindestens eine partielle Remission erreicht (Gesamtremissionsrate 80%).

Die mediane Zeit bis zum Therapieversagen war 18 Monate seit Therapiebeginn (Abbildung 3.4). Die mediane Beobachtungszeit der 123 Patienten ohne Therapieversagen war 20 Monate, die aus der Zensierungsverteilung geschätzte mediane Beobachtungszeit lag bei 45 Monaten. Von den 351 Patienten, die auf die initiale Therapie angesprochen hatten, war bei 80 Patienten eine autologe Stammzelltransplantation eingeleitet worden, bei 199 Patienten war eine Interferon- α -Erhaltungstherapie und bei 72 Patienten keine Erhaltungstherapie gegeben worden. Alle Patienten mit Stammzelltransplantation waren jünger als 66 Jahre. Von den 114 Patienten über 66 Jahre war bei 83 Patienten eine Interferon- α -Erhaltungstherapie und bei 31 Patienten keine Erhaltungstherapie eingeleitet worden.

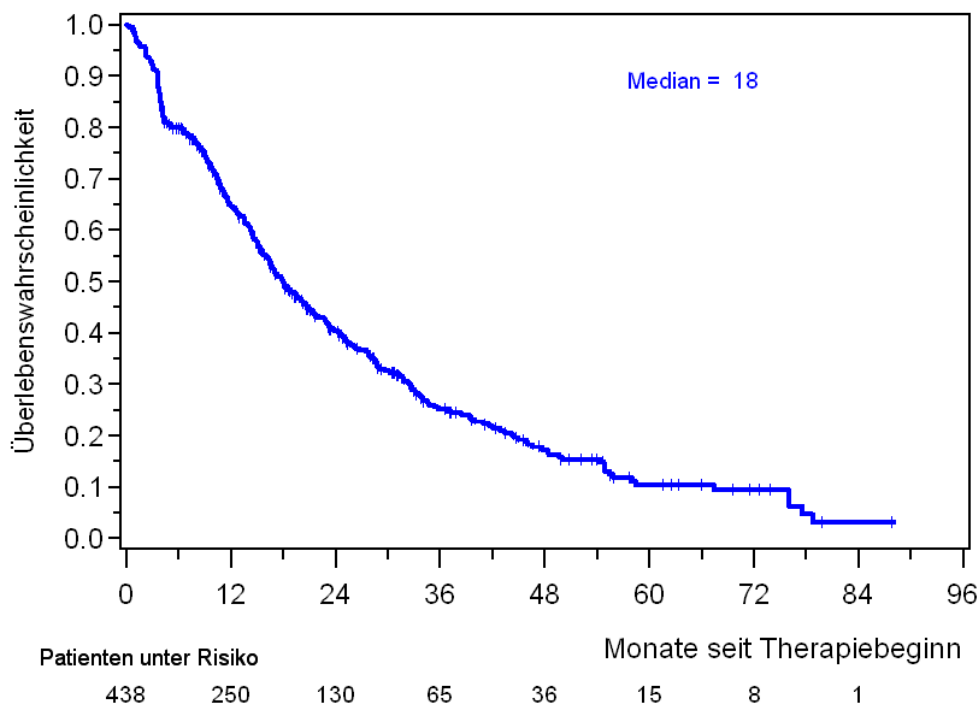


Abbildung 3.4: Zeit bis zum Therapieversagen

3.1.4 Gesamtüberleben

Von den 455 Patienten waren im Beobachtungszeitraum 159 Patienten verstorben, damit waren 65% der Patienten für die Analyse des Gesamtüberlebens zensiert. Die mediane Beobachtungszeit der überlebenden Patienten betrug 32 Monate, die mit der Zensierungsverteilung geschätzte mediane Beobachtungszeit 43 Monate. Die mediane Überlebenszeit lag bei 57 Monaten, die 5-Jahres-Überlebensrate bei 47% (95%-Konfidenzintervall 40% bis 53%). Abbildung 3.5 zeigt die Kaplan-Meier-Kurve für das Gesamtüberleben im Studienkollektiv.

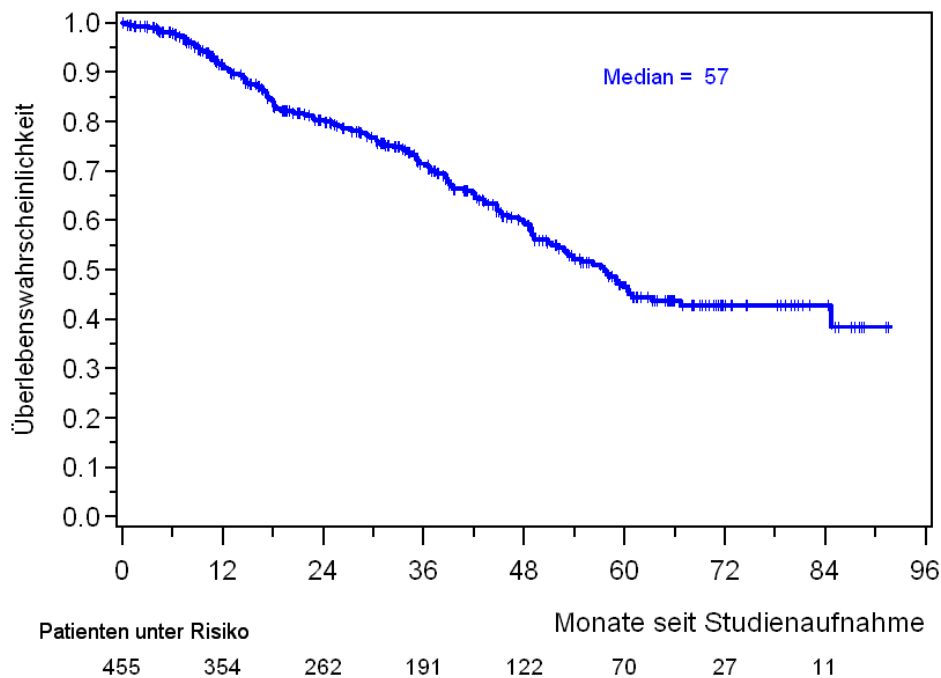


Abbildung 3.5: Gesamtüberleben

3.2 Validierung der bekannten Prognosemodelle

3.2.1 Ann-Arbor-Stadium

Von den 455 Patienten waren bei Studienaufnahme 71 (16%) im klinischen Stadium III nach Ann Arbor. Die überwiegende Mehrheit der Patienten, 384 (84%), befand sich im generalisierten Stadium IV. Mit einem medianen Gesamtüberleben von 57 und 58 Monaten zeigte das klinische Stadium in diesem Kollektiv von Patienten nur fortgeschrittenen Stadiums keine prognostische Relevanz (Logrank- $\chi^2 = 0,1460$, $p = 0,70$, Abbildung 3.6).

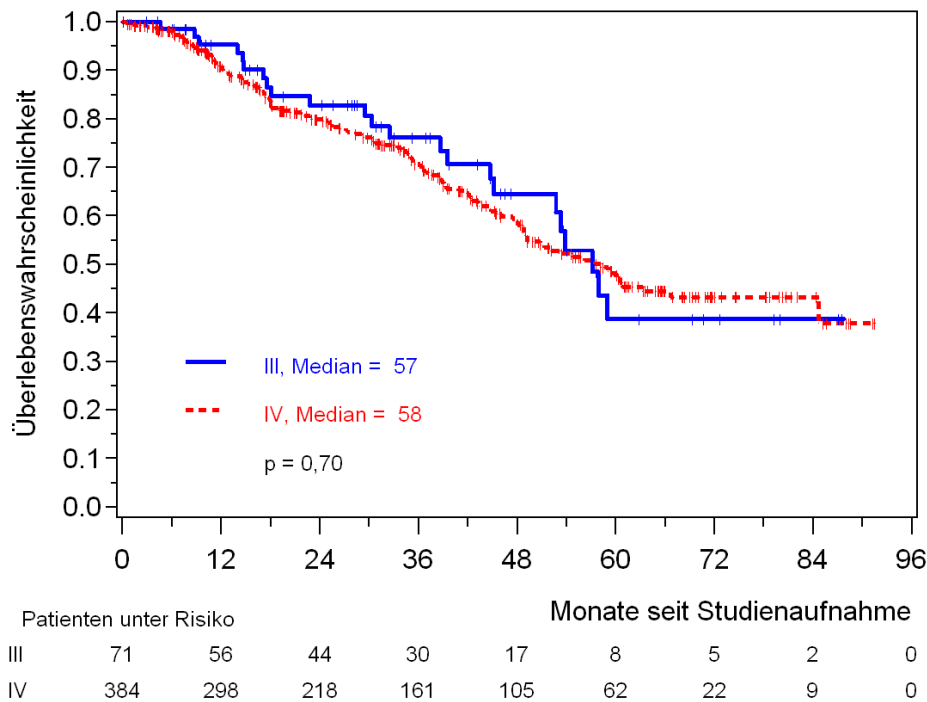


Abbildung 3.6: Gesamtüberleben nach dem Ann-Arbor-Stadium.

3.2.2 International Prognostic Index

Der IPI konnte bei 432 der 455 Patienten vollständig bestimmt werden. Davon waren 52% mindestens 60 Jahre alt, 9% hatten einen ECOG Performance Status 2 - 4, und je 32% hatten einen erhöhten LDH-Wert oder mehr als eine extranodale Manifestation (Tabelle 3.2). Alle Patienten waren im fortgeschrittenen Stadium und hatten damit mindestens einen Risikofaktor nach IPI.

Parameter	Wert	n	%
Alter	≥ 60 Jahre	226	52
Stadium	III oder IV	432	100
ECOG Performance Status	2 - 4	41	9
LDH	erhöht	138	32
Anzahl extranodaler Befälle	> 1	140	32
Anzahl Risikofaktoren	0	0	0
	1	99	23
	2	173	40
	3	119	28
	4	30	7
	5	11	3

Tabelle 3.2: Beschreibung der IPI-Risikofaktoren bei den 432 vollständig auswertbaren Patienten mit fortgeschrittenem MCL. n: absolute, %: relative Häufigkeit.

Die zahlenmäßig größte Gruppe bildeten mit 40% die Low Intermediate Risk-Patienten, gefolgt von den High Intermediate Risk- (28%), den Low Risk- (23%) und den High Risk-Patienten (9%, Tabelle 3.3). Die mediane Überlebenszeit war in der LR-Gruppe noch nicht erreicht, in der LIR-Gruppe 61 Monate, 45 Monate in der HIR- und 20 Monate in der HR-Gruppe (Abbildung 3.7). Mit $p < 0,0001$ beim Logrank-Test ($\text{Logrank-}\chi^2 = 38,82$)

zeigte der IPI einen signifikanten Effekt auf das Gesamtüberleben.

Risikogruppe	n	%	medianes OS	5-Jahres OS
Low Risk	99	23	nicht erreicht	59%
Low Intermediate Risk	173	40	61 Monate	52%
High Intermediate Risk	119	28	45 Monate	33%
High Risk	41	9	20 Monate	24%

Tabelle 3.3: Gesamtüberleben (OS) in den IPI-Risikogruppen bei 432 auswertbaren Patienten mit fortgeschrittenem MCL. n: absolute, %: relative Häufigkeit.

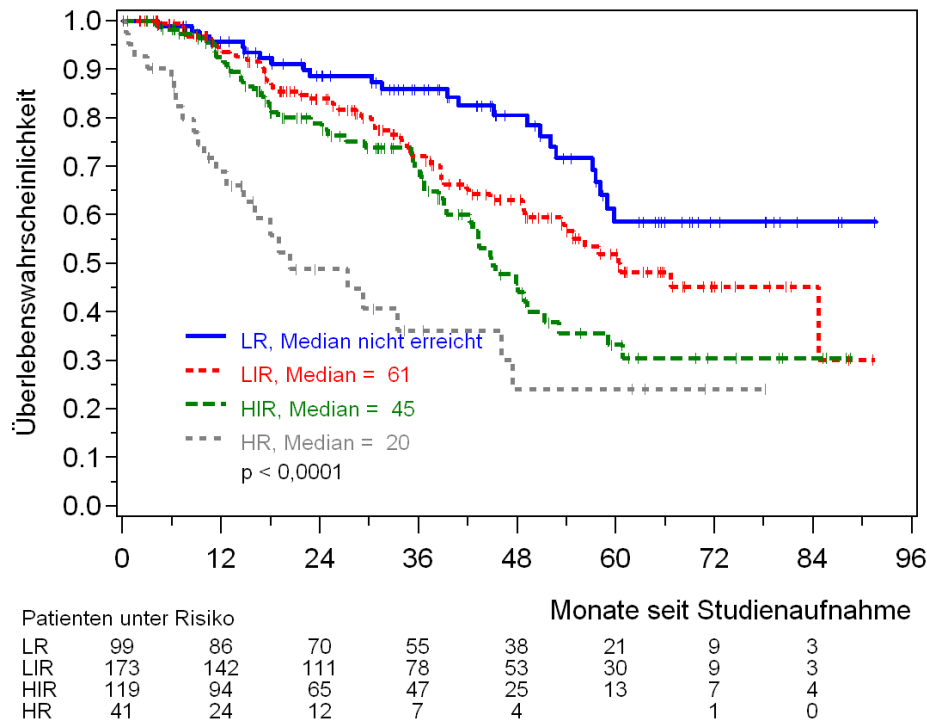


Abbildung 3.7: Gesamtüberleben der IPI-Risikogruppen im Kollektiv von 432 Patienten mit fortgeschrittenem MCL. LR: Low Risk, LIR: Low Intermediate Risk, HIR: High Intermediate Risk, HR: High Risk.

Im multiplen Cox-Modell der IPI-Risikofaktoren ohne Stadium zeigten Alter, LDH und ECOG Performance Status unabhängige prognostische Relevanz für das Gesamtüberleben, die Anzahl extranodaler Manifestationen hatte dagegen keine Relevanz (Tabelle 3.4).

Die HR-Gruppe war mit 9% der Patienten sehr klein. Die Überlebenskurven der LIR- und HIR-Gruppen verliefen in den ersten drei Jahren fast identisch und trennten sich erst danach auf. Die Beobachtungen jenseits der 48 Monate beruhten dabei auf weniger als einem Drittel der Patienten und die 95%-Konfidenzintervalle der 4-Jahres-Überlebensraten benachbarter Risikogruppen überlappten sich jeweils (LR: 69% - 88%, LIR: 54% - 71%, HIR: 34% - 57%, HR: 9% - 43%). Die Zusammenfassung der beiden mittleren Gruppen führte zu einer Gruppe mittleren Risikos von mehr als zwei Drittel der Patienten, ebenso die Zusammenfassung der beiden besten prognostischen Gruppen. Die Zusammenfassung der beiden schlechtesten Gruppen führte zwar zu ausgeglicheneren Patientenzahlen, aber

Risikofaktor	Vergleich	HazR (95%-KI)	p-Wert
Alter (Jahre)	≥ 60 vs. < 60	1,7 (1,2 - 2,3)	0,0018
LDH	> normal vs. ≤ normal	1,6 (1,2 - 2,3)	0,0055
ECOG	2 - 4 vs. 0 - 1	2,8 (1,8 - 4,5)	< 0,001
extranodale Befälle	> 1 vs. ≤ 1	1,2 (0,8 - 1,6)	0,33

Tabelle 3.4: Multiple Cox-Regression der IPI-Risikofaktoren auf das Gesamtüberleben im Datensatz von 432 Patienten mit fortgeschrittenem MCL. HazR: Hazard-Ratio, KI: Konfidenzintervall.

auch zu einer HR-Gruppe mit einem relativ guten medianen Gesamtüberleben von 43 Monaten.

3.2.3 Follicular Lymphoma International Prognostic Index

Bei 418 Patienten konnte der FLIPI vollständig bestimmt werden. Alle Patienten waren im Stadium III oder IV und hatten damit mindestens einen Risikofaktor nach FLIPI. 52% der Patienten waren mindestens 60 Jahre alt, 33% der Patienten hatten eine erhöhte LDH, 29% der Patienten einen erniedrigten Hb-Wert und 79% der Patienten mehr als vier befallene Lymphknotenregionen (Tabelle 3.5).

Parameter	Wert	n	%
Alter	≥ 60 Jahre	219	52
Stadium	III oder IV	418	100
LDH	erhöht	137	33
Hb	< 120g/l	120	29
Lymphknotenregionen	> 4	331	79
Anzahl Risikofaktoren	0	0	0
	1	26	6
	2	124	30
	3	148	35
	4	93	22
	5	27	6

Tabelle 3.5: Beschreibung der FLIPI-Risikofaktoren bei den 418 vollständig auswertbaren Patienten mit fortgeschrittenem MCL. n: absolute, %: relative Häufigkeit.

Nach der Anzahl Risikofaktoren hatten 6% der Patienten ein niedriges Risiko, 30% der Patienten mittleres Risiko und 64% der Patienten gehörten zur Hochrisikogruppe (Tabelle 3.6). Das mediane Gesamtüberleben war in LR- und IR-Gruppe nicht erreicht und betrug 48 Monate für die HR-Gruppe. Mit $p = 0,0009$ (Logrank- $\chi^2 = 14,14$) zeigte der FLIPI zwar einen signifikanten Einfluss auf das Gesamtüberleben, die Überlebenskurven der LR- und der IR-Gruppe waren jedoch nicht getrennt (Abbildung 3.8).

Risikogruppe	n	%	medianes OS	5-Jahres OS
Low Risk	26	6	nicht erreicht	61%
Intermediate Risk	124	30	nicht erreicht	57%
High Risk	268	64	48 Monate	38%

Tabelle 3.6: Gesamtüberleben (OS) in den FLIPI-Risikogruppen bei 418 auswertbaren Patienten mit fortgeschrittenem MCL. n: absolute, %: relative Häufigkeit.

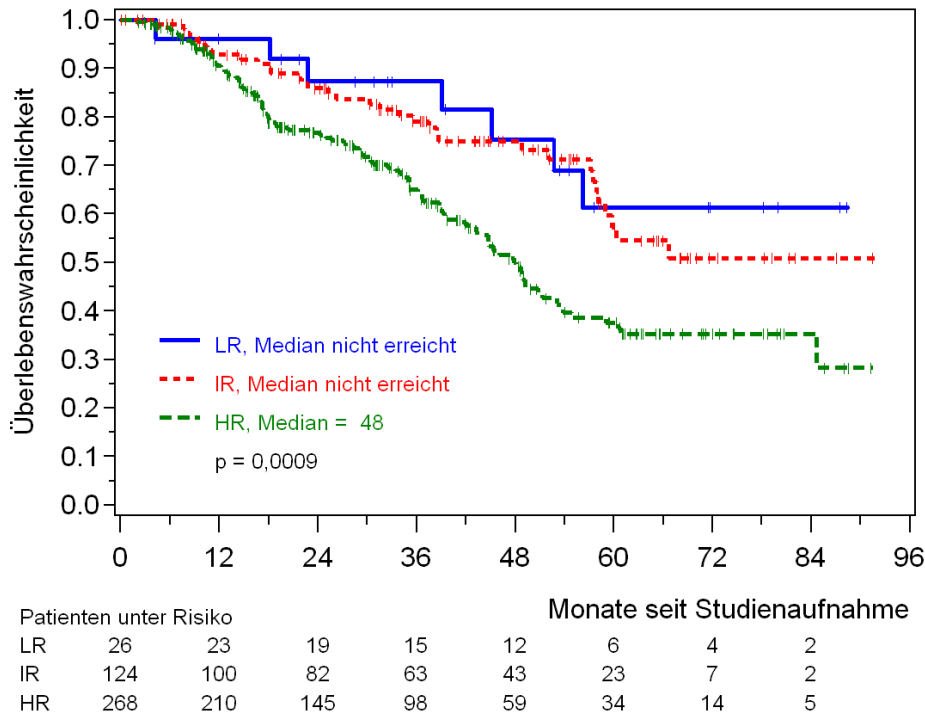


Abbildung 3.8: Gesamtüberleben der FLIPI-Risikogruppen im Kollektiv von 418 Patienten mit fortgeschrittenem MCL. LR: Low Risk, IR: Intermediate Risk, HR: High Risk.

Im multiplen Cox-Modell der FLIPI-Risikofaktoren außer dem Stadium zeigten Alter und LDH einen Einfluss auf das Gesamtüberleben, Hb und Anzahl befallener Lymphknotenregionen hatten dagegen keinen davon unabhängigen Einfluss (Tabelle 3.7).

Risikofaktor	Vergleich	HazR (95%-KI)	p-Wert
Alter (Jahre)	≥ 60 vs. < 60	1,8 (1,3 - 2,5)	0,00064
LDH	$> \text{normal}$ vs. $\leq \text{normal}$	1,8 (1,3 - 2,5)	0,0012
Hb	$< 120\text{g/l}$ vs. $\geq 120\text{g/l}$	1,1 (0,8 - 1,6)	0,50
Lymphknotenregionen	> 4 vs. ≤ 4	1,1 (0,7 - 1,6)	0,80

Tabelle 3.7: Multiple Cox-Regression der FLIPI-Risikofaktoren auf das Gesamtüberleben im Datensatz von 418 Patienten mit fortgeschrittenem MCL. HazR: Hazard-Ratio, KI: Konfidenzintervall.

Der FLIPI erwies sich für die Anwendung bei Patienten mit fortgeschrittenem Mantelzell-Lymphom als wenig praktikabel, da die Niedrigrisikogruppe sehr wenige Patienten enthielt, die Hochrisikogruppe dagegen fast zwei Drittel mit relativ gutem Outcome, und da in Bezug auf das Gesamtüberleben nicht klar drei Gruppen abgegrenzt werden konnten.

3.2.4 Vergleich und Bewertung

Die in Abschnitt 2.5.1 vorgestellten Maßzahlen der prognostischen Güte wurden für Stadium, IPI und FLIPI bestimmt und in Tabelle 3.8 vergleichend dargestellt. Das Stadium, das im Vergleich von Stadium III und Stadium IV keinen signifikanten Einfluss auf das

Gesamtüberleben zeigte, hatte auch ein Separationsmaß (SEP) nahe 1, einen Konkordanzindex c nahe 0,5 und den größten integrierten Brier Score.

Prognosemodell	Logrank- χ^2	df	p -Wert	SEP	c	IBS
Ann-Arbor-Stadium	0,15	1	0,70	1,02	0,5127	0,1898
IPI	38,82	3	$1,9 \cdot 10^{-8}$	1,41	0,6359	0,1747
IPI LR+LIR, HIR, HR	35,36	2	$2,1 \cdot 10^{-8}$	1,40	0,6149	0,1799
IPI LR, LIR+HIR, HR	35,10	2	$2,4 \cdot 10^{-8}$	1,31	0,6170	0,1753
IPI LR, LIR, HIR+HR	24,22	2	$5,5 \cdot 10^{-6}$	1,41	0,6178	0,1778
FLIPI	14,14	2	0,00085	1,36	0,5772	0,1763

Tabelle 3.8: Maßzahlen der prognostischen Güte von Stadium, IPI und FLIPI in Bezug auf das Gesamtüberleben im Kollektiv von Patienten mit fortgeschrittenem MCL. df: Anzahl Freiheitsgrade, IBS: integrierter Brier-Score.

Der IPI hatte im Vergleich zum FLIPI eine höhere Signifikanz, ein besseres Separationsmaß und einen kleineren integrierten Brier Score. Die deutlich schlechteren Ergebnisse des FLIPI im Vergleich zum IPI könnten widerspiegeln, dass das Mantelzell-Lymphom den aggressiven Lymphomen im Verlauf näher steht als dem follikulären Lymphom.

In den Publikationen zur Wertigkeit des IPI beim Mantelzell-Lymphom sind oft die ursprünglich vier Risikogruppen zu drei zusammengefasst worden. Tabelle 3.8 enthält daher auch die Maßzahlen der prognostischen Güte bei den drei sinnvollen Varianten des IPI mit drei Risikogruppen. Bei der Zusammenfassung der LR- und LIR-Gruppe bzw. der LIR- und HIR-Gruppe entstanden mit 63% bzw. 67% der Patienten ungünstig große Risikogruppen. Zahlenmäßig am günstigsten war die Zusammenfassung der HIR- und der HR-Gruppe (37% der Patienten). Dann aber waren die absoluten Unterschiede zwischen den Überlebenskurven weniger groß, was auch den um zwei Größenordnungen schlechteren p -Wert erklärte. Das Separationsmaß SEP war bei der Einteilung in vier Risikogruppen und bei der Zusammenfassung der jeweils äußeren Risikogruppen vergleichbar, bei der Zusammenfassung der mittleren beiden Gruppen deutlich schlechter. Das ist daher zu verstehen, dass die relativ große mittlere Gruppe im Wesentlichen den mittleren Verlauf der Gesamtkurve widerspiegelt und die beiden zahlenmäßig kleineren LR- bzw. HR-Gruppen ein kleineres Gewicht erhielten. Der Konkordanzindex war für die Einteilung in vier Risikogruppen besser als bei allen Modellen mit drei Risikogruppen, was vor allem an der geringeren Anzahl von Bindungen liegen mag. Die Zusammenfassung der HIR- und der HR-Gruppe zeigte hierbei den größten c -Index.

Zusammenfassend kann man feststellen, dass das Stadium nach Ann Arbor keine prognostische Relevanz bei Patienten mit fortgeschrittenem Mantelzell-Lymphom zeigt, und weder IPI noch FLIPI gute prognostische Indizes sind, wobei der IPI im Vergleich zum FLIPI im Bezug auf die Verteilung der Patienten zu den Risikogruppen, der Auftrennung der Überlebenskurven, sowie der prognostischen Relevanz der einzelnen Risikofaktoren überlegen ist. Diese Ergebnisse stehen im Gegensatz zu denen aus [9], bei denen auf eine Überlegenheit des FLIPI gegenüber dem IPI bei Patienten mit Mantelzell-Lymphom geschlossen wurde. In Übereinstimmung mit [9] erschien die Entwicklung einen neuen

Prognosemodells speziell für Patienten mit Mantelzell-Lymphom als erforderlich.

3.3 Entwicklung eines neuen prognostischen Index

3.3.1 Potenzielle Prognosefaktoren

Insgesamt 22 klinische Parameter wurden im Rahmen der Studien unmittelbar vor Therapiebeginn erhoben und kamen als potenzielle Prognosefaktoren in Betracht. Neben Alter, Geschlecht und Allgemeinzustand waren dies Tumorcharakteristika und Laborwerte. Zu den Tumorcharakteristika zählten das klinische Stadium nach Ann Arbor, das Vorhandensein von B-Symptomatik, Milzbefall oder Knochenmarksbefall, die Anzahl befallener Lymphknotenregionen und extranodaler Manifestationen, sowie die Größe des maximalen Lymphknotendurchmessers. Berücksichtigte Laborwerte waren Leukozyten-, Lymphozyten-, Granulozyten-, Monozyten- und Thrombozytenzahlen, LDH, Hb, Albumin, β_2 -Mikroglobulin, und die Immunglobulinwerte IgG, IgA und IgM.

Die Deskription der genannten Parameter im untersuchten Patientenkollektiv enthält Abschnitt 3.1.2. Im Folgenden sind für die einzelnen potenziellen Prognosefaktoren der univariate Einfluss auf das Gesamtüberleben (s. Tabelle 3.9), der univariate Anteil fehlender Werte, Besonderheiten der Verteilung metrischer Parameter (s. auch Abbildungen 3.1, 3.2) sowie mögliche Korrelationen der Parameter untereinander (s. auch Abbildung 3.3) zusammengefasst. Diese Informationen waren entscheidend für die endgültige Auswahl der potenziellen Prognosefaktoren und damit des Datensatzes der Modellentwicklung.

Alter bei Studienaufnahme

Das Alter der Patienten bei Studienaufnahme zeigte univariat einen signifikanten Einfluss auf das Gesamtüberleben. Bei Studienaufnahme um 10 Jahre ältere Patienten hatten ein geschätztes relatives Risiko von 1,38 (95%-Konfidenzintervall, KI, 1,15 - 1,65, $p = 0,0004$). Das Alter war vollständig dokumentiert.

Geschlecht

Das Geschlecht hatte univariat keinen Einfluss auf das Gesamtüberleben. Das relative Risiko von Männern gegenüber Frauen war 1,12 (95%-KI 0,77 - 1,63, $p = 0,55$). Auch das Geschlecht war vollständig dokumentiert.

Allgemeinzustand

Symptomatische Patienten mit ECOG Performance Status 1 hatten ein relatives Risiko von 1,61 (95%-KI 1,10 - 2,35, $p = 0,0137$) und bettlägerige oder nicht arbeitsfähige

Parameter	Vergleich	n	HazR (95%-KI)	p-Wert
Alter	+ 10 Jahre	455	1,38 (1,15 - 1,65)	0,0004
Geschlecht	männlich	455	1,12 (0,77 - 1,63)	0,55
ECOG	1 vs. 0	452	1,61 (1,10 - 2,35)	0,0137
	2 - 4 vs. 0		4,80 (2,84 - 8,12)	< 0,0001
	3 - 4 vs. 0 - 1		3,48 (2,22 - 5,43)	< 0,0001
Ann-Arbor-Stadium	IV vs. III	455	1,09 (0,70 - 1,70)	0,70
B-Symptome	vorhanden	451	1,60 (1,17 - 2,19)	0,0031
Milzbefall	vorhanden	420	1,40 (1,01 - 1,95)	0,0443
Knochenmarkbefall extranodale Befälle	vorhanden	453	1,04 (0,70 - 1,54)	0,85
	1 vs. 0	445	1,25 (0,72 - 2,16)	0,42
	> 1 vs. 0		1,57 (0,89 - 2,77)	0,12
	> 1 vs. 0 - 1		1,30 (0,94 - 1,80)	0,11
LK-Regionen	+ 1	432	1,03 (0,97 - 1,08)	0,34
Tumorgroße	+ 1cm	387	1,06 (1,01 - 1,10)	0,0155
LDH-Quotient*	+ 1	443	1,79 (1,38 - 2,31)	< 0,0001
	10-fach		6,04 (2,49 - 14,66)	0,0001
Hb	+ 10g/l	451	0,91 (0,85 - 0,98)	0,0092
wwHbMale			1,07 (0,91 - 1,26)	0,43
Leukozyten	+ 1.000/ μ l	451	1,001 (0,999 - 1,003)	0,40**
	10-fach		2,25 (1,59 - 3,18)	< 0,0001
Lymphozyten	+ 1.000/ μ l	422	1,004 (1,001 - 1,006)	0,0092
	10-fach		1,97 (1,40 - 2,77)	0,0001
Granulozyten	+ 1.000/ μ l	412	1,14 (1,08 - 1,19)	< 0,0001
	10-fach		4,69 (2,06 - 10,71)	0,0002
Monozyten	+ 1.000/ μ l	408	1,17 (1,05 - 1,31)	0,0036
	10-fach		2,61 (1,65 - 4,12)	< 0,0001
Thrombozyten	+ 10 ⁵ / μ l	453	1,01 (0,88 - 1,16)	0,85
Albumin-Quotient*	+ 1	268	0,23 (0,05 - 1,08)	0,0627
β_2 -Mikroglobulin-Quotient*	+ 1	285	1,26 (1,09 - 1,45)	0,0018
	2-fach		1,53 (1,22 - 1,92)	0,0003
IgG	+ 10g/l	289	1,07 (0,80 - 1,43)	0,65
	10-fach		1,05 (0,39 - 2,85)	0,92
IgA	+ 1g/l	286	1,03 (0,99 - 1,06)	0,13
	10-fach		1,72 (0,93 - 3,20)	0,0850
IgM	+ 1g/l	287	1,05 (0,93 - 1,20)	0,42
	10-fach		1,41 (0,83 - 2,40)	0,20

Tabelle 3.9: Ergebnis der univariaten Cox-Regression. HazR: Hazard-Ratio, KI: Konfidenzintervall. n: Anzahl der Werte, wwHbMale: Wechselwirkung zwischen Hb und Geschlecht.

*Quotient zum oberen Normwert

**bei Ausschluss des größten Ausreißers HazR 1,004 (95%-KI 1,001 - 1,006), $p = 0,0015$.

Patienten (ECOG Performance Status 2 - 4) von 4,80 (95%-KI 2,84 - 8,12, $p < 0,0001$) gegenüber asymptomatischen Patienten mit ECOG Performance Status 0. Patienten mit ECOG Performance Status 2 - 4 hatten gegenüber Patienten mit ECOG Performance Status 0 - 1 ein relatives Risiko von 3,48 (95%-KI 2,22 - 5,43, $p < 0,0001$). Der ECOG Performance Status war bei über 99% der Patienten dokumentiert.

Ann-Arbor-Stadium

Patienten im Stadium IV hatten im Vergleich zu Patienten im Stadium III kein erhöhtes Risiko (relatives Risiko 1,09, 95%-KI 0,70 - 1,70, $p = 0,70$). Das Stadium war vollständig dokumentiert.

B-Symptome

Das Vorhandensein von B-Symptomen war univariat mit einem relativen Risiko von 1,60 (95%-KI 1,17 - 2,19, $p = 0,0031$) verbunden. Angaben zur B-Symptomatik gab es von über 99% der Patienten.

Milzbefall

Patienten, bei denen die Milz durch das Mantelzell-Lymphom befallen war, hatten ein relatives Risiko von 1,40 (95%-KI 1,01 - 1,95, $p = 0,0443$) zu sterben. Angaben zum Milzbefall waren bei 92% der Patienten dokumentiert.

Knochenmarksbefall

Patienten mit Knochenmarksbefall hatten kein erhöhtes Risiko gegenüber Patienten ohne Knochenmarksbefall. Das relative Risiko wurde auf 1,04 (95%-KI 0,70 - 1,54, $p = 0,85$) geschätzt. Das Knochenmark war bei über 99% der Patienten untersucht worden. Knochenmarksbefall und Stadium wiesen eine hohe Übereinstimmung auf, da ein Stadium IV in den meisten Fällen bereits durch Knochenmarksbefall bedingt war. Alle 360 Patienten mit Knochenmarksbefall waren im Stadium IV und von den 93 Patienten ohne Knochenmarksbefall hatten nur 24 Patienten einen anderen extranodalen Befall, der ein Stadium IV zu Folge hatte.

Extranodaler Befall

Die Anzahl extranodaler Befälle hatte keinen Einfluss auf das Gesamtüberleben. Patienten mit einem extranodalen Befall hatten gegenüber Patienten ohne Extranodalbefall ein relatives Risiko von 1,25 (95%-KI 0,72 - 2,16, $p = 0,42$), und Patienten mit mehr als einem Extranodalbefall von 1,57 (95%-KI 0,89 - 2,77, $p = 0,12$). Das relative Risiko von Patienten mit mehr als einem gegenüber Patienten mit maximal einem extranodalen Befall

war 1,30 (95%-KI 0,94 - 1,80, $p = 0,11$). Angaben zum extranodalen Befall gab es von 98% der Patienten. Auch die Anzahl extranodaler Manifestationen war mit dem Stadium korreliert. Alle 53 Patienten ohne Extranodalbefall waren im Stadium III und von den 392 Patienten mit mindestens einem extranodalem Befall waren nur 16 im Stadium III, hatten also einen lokalisierten extranodalen Befall meist des Gastrointestinaltraktes.

Befallene Lymphknotenregionen

Die Anzahl befallener Lymphknotenregionen hatte mit einem relativen Risiko von 1,03 (95%-KI 0,97 - 1,08, $p = 0,34$) bei einer zusätzlichen Lymphknotenregion keinen Einfluss auf das Gesamtüberleben. Die Anzahl befallener Lymphknotenregionen war bei 95% der Patienten bestimmt worden.

Maximale Lymphknotengröße

Der maximale Durchmesser des größten Lymphknotens hatte einen signifikanten Einfluss auf das Gesamtüberleben. Das relative Risiko pro Zentimeter war 1,06 (95%-KI 1,01 - 1,10, $p = 0,0155$). Eine Angabe der Lymphknotengröße fehlte von 15% der Patienten. Die maximale Lymphknotengröße hatte zwar eine rechtsschiefe Verteilung, durch Logarithmierung wurde die Verteilung aber deutlich linksschief.

Differentialblutbild

Die Leukozytenzahl zeigte zunächst keinen Einfluss auf das Gesamtüberleben. Eine Erhöhung um $1000/\mu\text{l}$ war mit einem relativen Risiko von 1,001 (95%-KI 0,999 - 1,003, $p = 0,40$) verbunden. Bei Ausschluss des größten Ausreißers mit 764.000 Leukozyten/ μl zeigte sich jedoch ein signifikanter Einfluss mit relativem Risiko 1,004, 95%-KI 1,001 - 1,006, und $p = 0,0015$. Die Verteilung der Leukozytenzahlen war sehr stark rechtsschief (s. Abbildung 3.1), eine Logarithmierung führt hier zu einer wesentlich symmetrischeren Verteilung (s. Abbildung 3.2). Die logarithmierten Werte zeigten einen signifikanten Einfluss auf das Gesamtüberleben. Eine 10-fache Erhöhung der Leukozytenzahl war mit einem relativen Risiko von 2,25 (95%-KI 1,59 - 3,18, $p < 0,0001$) verbunden.

Bei einer um $1000/\mu\text{l}$ höheren Lymphozytenzahl war das relative Risiko 1,004 (95%-KI 1,001 - 1,006, $p = 0,0092$). Auch die Lymphozytenzahl war stark rechtsschief verteilt, und eine Logarithmierung führte auch hier zu einer deutlich symmetrischeren Verteilung. Das univariate relative Risiko für eine 10-fache Lymphozytenzahl wurde auf 1,97 (95%-KI 1,40 - 2,77, $p = 0,0001$) geschätzt. Mit einem Korrelationskoeffizient nach Pearson von $r = 0,83$ waren Leukozyten- und Lymphozytenzahlen stark korreliert, eine Leukozytose ging meist mit einer Lymphozytose einher. Mit 451 Werten war die Leukozytenzahl in über 99% der Fälle dokumentiert, die Lymphozytenzahl in 93%.

Auch erhöhte Granulozyten- und Monozytenzahlen waren mit einem höheren Risiko verbunden. Das geschätzte relative Risiko einer um $1.000/\mu\text{l}$ höheren Granulozytenzahl war 1,14 (95%-KI 1,08 - 1,19, $p < 0,0001$), das einer um $1.000/\mu\text{l}$ höheren Monozytenzahl 1,17 (95%-KI 1,05 - 1,31, $p = 0,0036$). Beide Parameter waren rechtsschief verteilt, und auch in logarithmierter Form mit Einfluss auf das Gesamtüberleben (relatives Risiko 4,69, 95%-KI 2,06 - 10,71, $p = 0,0002$, für eine 10-fache Granulozytenzahl und 2,61, 95%-KI 1,65 - 4,12, $p < 0,0001$, für eine 10-fache Monozytenzahl). Beide Parameter wiesen ebenfalls eine gewisse Korrelation mit der Leukozytenzahl auf ($r = 0,50$ bzw. $r = 0,53$) und waren mit 91% und 90% nicht so häufig dokumentiert.

Die Thrombozytenzahl hatte mit einem relativen Risiko von 1,01 für um $1.000/\mu\text{l}$ höhere Werte (95%-KI 0,88 - 1,16, $p = 0,85$) keinen Einfluss auf das Gesamtüberleben. Die Thrombozytenzahl war in über 99% der Patienten dokumentiert. Die Verteilung war bis auf wenige Ausreißer relativ symmetrisch, eine Logarithmierung bewirkte keine wesentlich symmetrischere Verteilung. Mit $r = 0,01$ waren Thrombozyten- und Leukozytenzahlen nicht korreliert.

LDH

Da für die Messung der Enzymaktivität der Laktat-Dehydrogenase (LDH) im Serum keine standardisierte Methode existiert und die Normalwerte der Labors eine erhebliche Schwankung aufweisen, wurde für die Analyse des LDH-Wertes der Quotient zum oberen Normalwert verwendet. Eine Erhöhung des LDH-Wertes um den oberen Normalwert war mit einem relativen Risiko von 1,79 (95%-KI 1,38 - 2,31, $p < 0,0001$) verbunden. LDH und oberer Normalwert waren bei über 97% der Patienten dokumentiert. Der LDH-Quotient zeigte eine rechtsschiefe Verteilung, die durch Logarithmierung deutlich symmetrischer gemacht werden konnte. Ein 10-facher LDH-Wert (bei gleichem oberen Normalwert) hatte dann ein relatives Risiko von 6,04 (95%-KI 2,49 - 14,66, $p = 0,0001$) zur Folge.

Hämoglobin

Ein höherer Blut-Hämoglobinwert führte zu einem niedrigeren Risiko (relatives Risiko 0,91 für einen um 10g/dl höheren Wert, 95%-KI 0,85 - 0,98, $p = 0,0092$). Diese Analyse war unabhängig vom Geschlecht. In dem Modell mit Berücksichtigung des Geschlechtes war die Wechselwirkung von Hämoglobinwert und Geschlecht nicht signifikant (relatives Risiko 1,07, 95%-KI 0,91 - 1,26, $p = 0,43$). Für die multiple Regression wurde daher der Hämoglobinwert unabhängig vom Geschlecht berücksichtigt. Der Hämoglobinwert war bei über 99% der Patienten dokumentiert.

Albumin

Der Albuminwert war bei 321 Patienten dokumentiert und lag im Median bei 43g/l (20g/l - 85g/l). Ein oberer Normwert für Albumin war in 272 Patienten dokumentiert und die Werte lagen zwischen 39g/l und 87g/l. Auf Grund der stark variierenden Normwerte erschien der absolute Albuminwert als nicht vergleichbar zwischen Patienten unterschiedlicher Zentren. Daher wurde bei 268 Patienten (59%) der Quotient von Albuminwert zum oberen Normwert berechnet. Der Albumin-Quotient lag hierbei im Median bei 80% (36% - 126%). In der univariaten Analyse zeigte Albumin als Quotient zum oberen Normalwert keinen signifikanten Einfluss auf das Gesamtüberleben. Das relative Risiko eines um den oberen Normalwert erhöhten Albuminwertes war 0,23 (95%-KI 0,05 - 1,08, $p = 0,0627$).

β_2 -Mikroglobulin

Da für die Bestimmung der Konzentration von β_2 -Mikroglobulin sehr unterschiedliche Normwerte angegeben wurden, musste auch hier wie bei der LDH-Aktivität der Quotient zum oberen Normalwert als potenzieller Prognosefaktor eingebracht werden. Eine Erhöhung des β_2 -Mikroglobulin-Wertes um den oberen Normalwert erhöhte das Risiko signifikant (relatives Risiko 1,26, 95%-KI 1,09 - 1,45, $p = 0,0018$). Auch die Verteilung der β_2 -Mikroglobulin-Werte war stark rechtsschief und war nach Logarithmierung symmetrischer. Ein doppelter β_2 -Mikroglobulin-Wert war dann mit einem relativen Risiko von 1,53, 95%-KI 1,22 - 1,92, $p = 0,0003$, verbunden. Der β_2 -Mikroglobulin-Quotient fehlte bei 37 % der Patienten.

Immunglobuline

IgG, IgA und IgM zeigten keinen Einfluss auf das Gesamtüberleben. Auch nach Logarithmierung der in allen drei Fällen rechtsschiefen Verteilungen konnte kein Effekt gezeigt werden. Immunglobulinwerte fehlten bei über 30% der Patienten.

3.3.2 Regressionsmodell

Datensatz zur Modellentwicklung

Die Auswahl des Datensatzes für die multiple Regression zur Identifizierung unabhängiger Prognosefaktoren erfolgte nach Analyse und Bewertung der Anzahl vollständiger Datensätze für die potenziellen Prognosefaktoren. Im Gesamtdatensatz von 455 Patienten variierte der Anteil fehlender Werte einzelner Parameter zwischen 0% und 41% (s. Tabelle 3.1). Die Angaben zu Alter, Geschlecht und Stadium waren vollständig. Der Anteil fehlender Angaben zu Knochenmarksbefall, ECOG Performance Status, Thrombozytenzahl, Hämoglobin, Leukozytenzahl, B-Symptomatik, Anzahl extranodaler Befälle, LDH-Quotient oder Anzahl befallener Lymphknotenregionen lag zwischen 0% und 5%. Anga-

ben zu Lymphozyten-, Granulozyten- und Monozytenzahlen, Milzbefall oder zur maximalen Tumorgröße fehlten in 7% bis 15% der Fälle. Albumin-, β_2 -Mikroglobulin- oder Immunglobulinwerte fehlten in über 30% der Fälle. Zu den 12 Parametern mit maximal 5% fehlenden Werten gab es 409 vollständige Datensätze (90% des Gesamtdatensatzes). Zu den 17 Parametern mit maximal 15% fehlenden Werten waren es 288 vollständige Datensätze (63% des Gesamtdatensatzes). Vollständige Datensätze für alle Variablen außer den Immunglobulinen gab es nur 159 (35%) und für alle Variablen 111 (24%). Für die multiple Cox-Regression zur Identifizierung unabhängiger Prognosefaktoren wurden die 409 vollständigen Datensätze zu den 12 potenziellen Prognosefaktoren mit einem Anteil fehlender Werte von maximal 5% verwendet.

Von diesen 12 potenziellen Prognosefaktoren gingen alle bis auf den Knochenmarksbefall in die multiple Regression ein. Der Knochenmarksbefall wurde auf Grund der großen Übereinstimmung mit Stadium IV nicht verwendet (5% Stadium IV ohne Knochenmarksbefall). Milzbefall, Tumorgröße, Albumin und β_2 -Mikroglobulin gingen trotz univariater Signifikanz, die Immunglobuline bei fehlender univariater Signifikanz wegen des hohen Anteils fehlender Werte nicht in das multiple Modell ein. Lymphozyten-, Granulozyten- und Monozytenzahl wurden bei hoher Korrelation mit der Leukozytenzahl wegen eines größeren Anteils fehlender Werte nicht berücksichtigt. Tabelle 3.10 fasst zusammen, welche Parameter als potenzielle Prognosefaktoren in das multiple Regressionsmodell eingebracht wurden.

Parameter	Wert	verwendet	Codierung	Begründung
Alter bei Studienaufnahme	Jahre	ja	kontinuierlich	
Geschlecht	m, w	ja	binär	
ECOG	0, 1, 2 - 4	ja	3 Kategorien	
Ann-Arbor-Stadium	III, IV	ja	binär	
B-Symptomatik	A, B	ja	binär	
Milzbefall	ja, nein	nein		8% fehlend
Anzahl extranodaler Befälle	≤ 1 , > 1	ja	binär	
Knochenmarksbefall	ja, nein	nein		Korrelation mit Stadium
LK-Regionen	Anzahl	ja	metrisch	
Tumorgröße	cm	nein		15% fehlend
LDH-Quotient		ja	kontinuierlich, logarithmiert	
Hb	g/l	ja	kontinuierlich	
Leukozyten	Zahl/ μ l	ja	kontinuierlich, logarithmiert	
Lymphozyten	Zahl/ μ l	nein		7% fehlend, Korrelation mit Leukozytenzahl
Granulozyten	Zahl/ μ l	nein		9% fehlend, Korrelation mit Leukozytenzahl
Monozyten	Zahl/ μ l	nein		10% fehlend, Korrelation mit Leukozytenzahl
Thrombozyten	Zahl/ μ l	ja	kontinuierlich	
Albumin-Quotient		nein		41% fehlend
β_2 -Mikroglobulin-Quotient		nein		37% fehlend
IgG	g/l	nein		37% fehlend
IgA	g/l	nein		36% fehlend
IgM	g/l	nein		37% fehlend

Tabelle 3.10: Potenzielle prognostische Faktoren für multiple Regression. Bei den für das multiple Modell nicht verwendeten Parametern ist die Begründung für die Nicht-Berücksichtigung angegeben.

Das Alter in Jahren ging kontinuierlich in das Modell ein, der ECOG Performance Status in der Klassifizierung „arbeitsunfähig oder bettlägerig“ (Grad 2 - 4) vs. „eingeschränkt“ (Grad 1) vs. „asymptomatisch“ (Grad 0), die Anzahl extranodaler Befälle auf Grund mangelnder univariater Signifikanz nur binär (> 1 vs. ≤ 1 , wie beim IPI [29]), die Anzahl befallener Lymphknotenregionen metrisch und die Leukozytenzahl sowie der LDH-Quotient zum oberen Normalwert metrisch und logarithmiert ein.

In dem Datensatz der 409 complete cases waren im Beobachtungszeitraum 146 Todesfälle beobachtet worden. Der ECOG Performance Status hatte zwei Freiheitsgrade, während alle übrigen potenziellen Prognosefaktoren, die für die multiple Regression berücksichtigt wurden, jeweils einen Freiheitsgrad aufwiesen. Damit errechnete sich die Zahl von Ereignissen pro Variable (EPV) für die multiple Cox-Regression zu $146/12 = 12,17$.

Multiple Cox-Regression

Im Cox-Modell zeigten nach Variablenselektion von den elf potenziellen Prognosefaktoren vier Parameter unabhängige Signifikanz. Ein höheres Alter, ein schlechterer Allgemeinzustand, ein höherer LDH-Wert sowie höhere Leukozytenzahlen waren mit einem kürzeren Gesamtüberleben assoziiert. Das relative Risiko (Hazard-Ratio) war 1,42 für um je 10 Jahre ältere Patienten, 2,01 für bettlägerige oder nicht arbeitsfähige Patienten, 3,92 für Patienten mit 10-fachem LDH-Wert und 2,56 für Patienten mit 10-facher Leukozytenzahl (Tabelle 3.11). Vorwärts- wie Rückwärts-Variablenselektion auf dem 5%-Niveau wie auf dem 10%-Niveau führten konsistent zu demselben Modell.

Parameter	Vergleich	β	SE	HazR (95%-KI)	p-Wert
Alter	+ 10 Jahre	0,3535	0,09604	1,42 (1,18 - 1,72)	0,0002
ECOG	2 - 4 vs. 0 - 1	0,6978	0,2663	2,01 (1,19 - 3,39)	0,0088
LDH	10-fach	1,367	0,4962	3,92 (1,48 - 10,37)	0,0059
Leukozyten	10-fach	0,9393	0,2220	2,56 (1,66 - 3,95)	< 0,0001

Tabelle 3.11: Unabhängige signifikante Einflussgrößen auf das Gesamtüberleben. β : Regressionskoeffizient, SE: Standardfehler von β , HazR: Hazard-Ratio, KI: Konfidenzintervall.

Mit den Regressionskoeffizienten aus Tabelle 3.11 errechnet sich der Prognosescore aus den Parametern Alter, ECOG Performance Status, LDH-Quotient und Leukozytenzahl zu

$$\begin{aligned}
 \text{Prognosescore} = & 0,03535 \cdot \text{Alter}(\text{Jahre}) \\
 & + 0,6978(\text{falls ECOG} \geq 2) \\
 & + 1,367 \cdot \log_{10}(\text{LDH-Quotient}) \\
 & + 0,9393 \cdot \log_{10}(\text{Leukozyten}/\mu\text{l}). \tag{3.1}
 \end{aligned}$$

3.3.3 Prognostischer Index (MIPI)

Nach Identifizierung der unabhängigen Prognosefaktoren erfolgte die Definition von Risikogruppen für den neuen prognostischen Index durch Kategorisierung des Prognosescores (Formel 3.1). Im Datensatz der Modellentwicklung lag der Wert des Prognosescores im Median bei 5,78 (Range 4,31 – 9,18) und die Verteilung des Prognosescores war unimodal (Abbildung 3.9). Aus der Verteilung ergab sich also kein Hinweis auf eine möglicherweise biologisch begründete Anzahl von Risikogruppen oder auf die Lage potenzieller Cutpoints. Als Grenzwerte zur Definition von Risikogruppen kamen daher alle Werte zwischen 5,15 und 6,65, also nahe dem 10%-Quantil (5,16) und dem 90%-Quantil (6,63), in Schritten von 0,05, prinzipiell in Betracht.

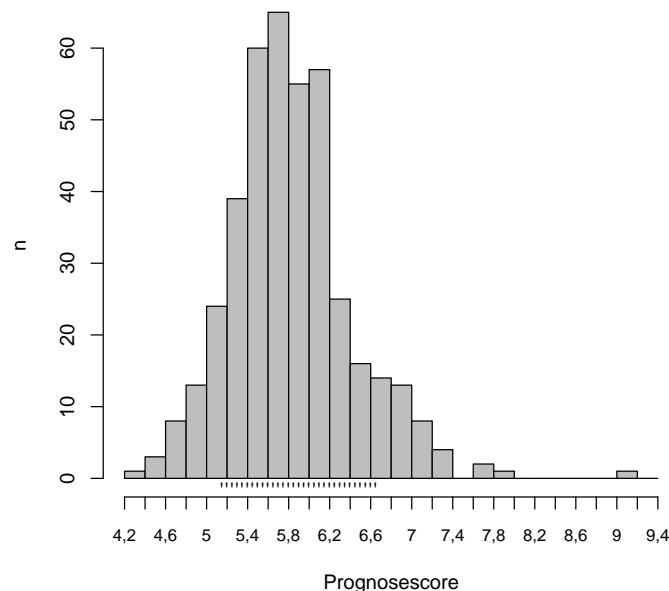


Abbildung 3.9: Histogramm der Verteilung des Prognosescores in den 409 zur Modellentwicklung verwendeten Daten von Patienten mit fortgeschrittenem MCL. Die vertikalen Striche unter dem Histogramm markieren die potenziellen Cutpoints zur Kategorisierung des Prognosescores.

Zunächst wurden durch Dichotomisierung des Prognosescores zwei Risikogruppen gebildet. Tabelle 3.12 und Abbildung 3.10, links, zeigen die Werte der Logrank-Statistik ($df = 1$) für die verschiedenen Cutpoints. Alle Cutpoints definierten in Bezug auf das Gesamtüberleben signifikant unterschiedliche Risikogruppen mit p -Werten < 0.001 unkorrigiert und < 0.028 nach Bonferroni-Korrektur (61 Tests). Der Cutpoint mit der höchsten Logrank-Statistik war 6,20 ($\chi^2 = 53,06$), gefolgt von 6,25 ($\chi^2 = 52,86$) und 6,10 ($\chi^2 = 50,83$). Zu diesen Cutpoints errechneten sich unkorrigierte p -Werte von $3,2 \cdot 10^{-13}$, $3,6 \cdot 10^{-13}$ und $1,0 \cdot 10^{-12}$ und Bonferroni-korrigierte Werte von $2,0 \cdot 10^{-11}$, $2,2 \cdot 10^{-11}$

und $6,2 \cdot 10^{-11}$.

Cutpoint	n	%	χ^2	Cutpoint	n	%	χ^2
6,20	84	21	53,06	5,70	229	56	30,66
6,25	77	19	52,86	6,00	141	34	28,18
6,10	108	26	50,83	5,90	160	39	28,04
6,15	97	24	45,73	5,95	150	37	26,68
6,60	43	11	44,02	5,65	249	61	26,68
6,35	65	16	43,91	5,50	290	71	25,21
6,30	69	17	42,19	5,60	261	64	23,23
6,55	47	11	41,96	5,45	308	75	22,78
6,65	39	10	37,75	5,55	274	67	21,26
6,40	59	14	37,61	5,30	339	83	19,69
6,50	52	13	36,69	5,35	330	81	19,22
6,45	54	13	36,51	5,40	321	78	18,49
6,05	123	30	36,43	5,25	353	86	17,29
5,80	196	48	35,15	5,20	360	88	15,74
5,75	214	52	33,36	5,15	371	91	12,31
5,85	174	43	32,75				

Tabelle 3.12: Logrank- χ^2 -Statistik für das Gesamtüberleben bei je zwei Risikogruppen (1 Freiheitsgrad) nach Dichotomisierung des Prognosescores mit verschiedenen Cutpoints. n, %: absolute, relative Häufigkeit von Patienten mit einem Prognosescore größer oder gleich dem Cutpoint.

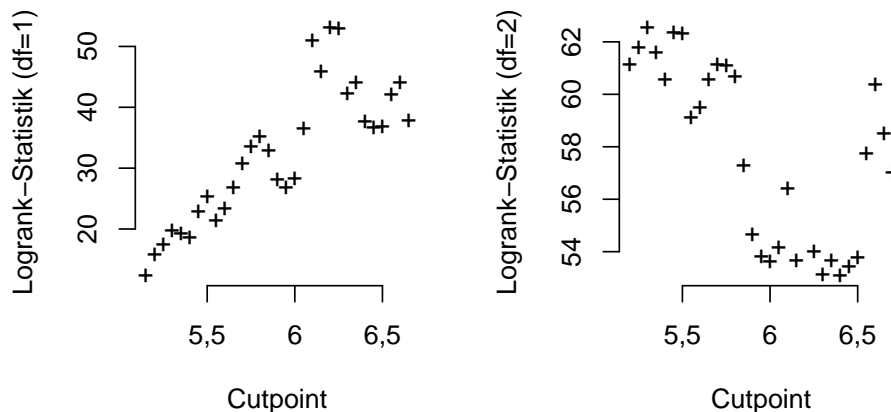


Abbildung 3.10: Logrank-Statistik für potenzielle Cutpoints. Links: erster Cutpoint. Rechts: zweiter Cutpoint mit erstem Cutpoint 6,2. df: Anzahl Freiheitsgrade.

Mit festem Cutpoint 6,20 zeigen Tabelle 3.13 und Abbildung 3.10, rechts, die Logrank-Statistik für das Gesamtüberleben bei Unterteilung des Prognosescores in drei Gruppen ($df = 2$). Auch hier definierten alle Cutpoints zusammen mit dem ersten Cutpoint 6,20 im Gesamtüberleben signifikant unterschiedliche Risikogruppen mit unkorrigierten p -Werten kleiner als 10^{-11} und Bonferroni-korrigierten p -Werten kleiner als 10^{-9} (61 Tests). Der beste Cutpoint war zunächst 5,30 mit Teststatistik $\chi^2 = 62,54$. Diese Wahl hätte aber zu einer Gruppe mittleren Risikos von 255 Patienten (62%) geführt. Unter den Cutpoints, die 204 Patienten oder weniger als mittleren Risikos klassifizieren würden, war 5,70 der

optimale (Logrank- $\chi^2 = 61,11$), gefolgt von 5,75 ($\chi^2 = 61,06$, Abbildung 3.10). Zu diesen Cutpoints errechneten sich p -Werte von $5,4 \cdot 10^{-14}$ und $5,5 \cdot 10^{-14}$ unkorrigiert sowie $3,3 \cdot 10^{-12}$ und $3,4 \cdot 10^{-12}$ nach Bonferroni-Korrektur (61 Tests).

Cutpoint	LR	%	IR	%	HR	χ^2	Cutpoint	LR	%	IR	%	HR	χ^2
5,30	70	17	255	62	84	62,54	6,55	325	79	37	9	47	57,73
5,45	101	25	224	55	84	62,35	5,85	235	57	90	22	84	57,28
5,50	119	29	206	50	84	62,31	6,70	325	79	50	12	34	56,99
5,25	56	14	269	66	84	61,76	6,10	301	74	24	6	84	56,40
5,35	79	19	246	60	84	61,57	5,90	249	61	76	19	84	54,63
5,20	49	12	276	67	84	61,11	6,05	286	70	39	10	84	54,13
5,70	180	44	145	35	84	61,11	6,25	325	79	7	2	77	53,97
5,75	195	48	130	32	84	61,06	5,95	259	63	66	16	84	53,79
5,80	213	52	112	27	84	60,67	6,50	325	79	32	8	52	53,77
5,40	88	22	237	58	84	60,55	6,35	325	79	19	5	65	53,63
5,65	160	39	165	40	84	60,52	6,15	312	76	13	3	84	53,62
6,60	325	79	41	10	43	60,35	6,00	268	66	57	14	84	53,60
5,60	148	36	177	43	84	59,48	6,45	325	79	30	7	54	53,43
5,55	135	33	190	46	84	59,08	6,30	325	79	15	4	69	53,11
6,65	325	79	45	11	39	58,46	6,40	325	79	25	6	59	53,08

Tabelle 3.13: Patientenverteilung und Logrank- χ^2 -Statistik für das Gesamtüberleben bei je drei Risikogruppen (2 Freiheitsgrade) mit erstem Cutpoint 6,2 und unterschiedlichem zweiten Cutpoint für den Prognosescore (PS). LR: Low Risk, $PS < \text{Minimum}(\text{Cutpoint}, 6,2)$, IR: Intermediate Risk, $\text{Minimum}(\text{Cutpoint}, 6,2) \leq PS < \text{Maximum}(\text{Cutpoint}, 6,2)$, HR: High Risk, $\text{Maximum}(\text{Cutpoint}, 6,2) \leq PS$.

Zur Kontrolle wurden auch alle Paare von Cutpoints auf die signifikante Auftrennung von drei Risikogruppen in Bezug auf das Gesamtüberleben untersucht. Das Paar (6,2;5,7) zeigte hierbei auch die höchste Signifikanz unter allen Paaren, die Risikogruppen mit mindestens 20% und höchstens 50% der Patienten definierten. Somit erzielten sowohl sequenzielle als auch gleichzeitige Auswahl von zwei Cutpoints dasselbe Ergebnis.

Nach dem neuen prognostischen Index (MCL International Prognostic Index, MIPI) wurden Patienten mit einem Prognosescore (Formel 3.1) kleiner als 5,7 zur Gruppe niedrigen Risikos (Low Risk, LR), Patienten mit einem Score von mindestens 5,7 aber kleiner als 6,2 zur Gruppe mittleren Risikos (Intermediate Risk, IR) und Patienten mit einem Score von 6,2 oder höher zur Gruppe höchsten Risikos (High Risk, HR) zusammengefasst. Tabelle 3.14 und Abbildung 3.11 zeigen das Gesamtüberleben der Patienten in den drei Risikogruppen.

Risikogruppe	n	%	Medianes OS	2-Jahres OS (95%-KI)	5-Jahres OS (95%-KI)
Low Risk	180	44	nicht erreicht	90% (86% - 95%)	60% (51% - 72%)
Intermediate Risk	145	35	51 mo	83% (77% - 90%)	42% (32% - 55%)
High Risk	84	21	29 mo	54% (43% - 67%)	18% (9% - 35%)

Tabelle 3.14: Patientenverteilung und Gesamtüberleben (OS) nach dem neu entwickelten prognostischen Index auf dem Datensatz der Modellentwicklung mit 409 Patienten. n: absolute, %: relative Häufigkeit.

So waren 44% der Patienten in der Gruppe niedrigen Risikos, in der das mediane Gesamtüberleben noch nicht erreicht war, 35% der Patienten waren in der Gruppe mittleren Risikos mit einem medianen Gesamtüberleben von 51 Monaten und 21% wurden als Hoch-

risikopatienten klassifiziert, bei einem medianen Gesamtüberleben von 29 Monaten. Die Auftrennung war mit einer Logrank-Statistik von $\chi^2 = 61, p < 0,0001$, hochsignifikant, allerdings ist das die überoptimistische Schätzung am Datensatz der Modellentwicklung.

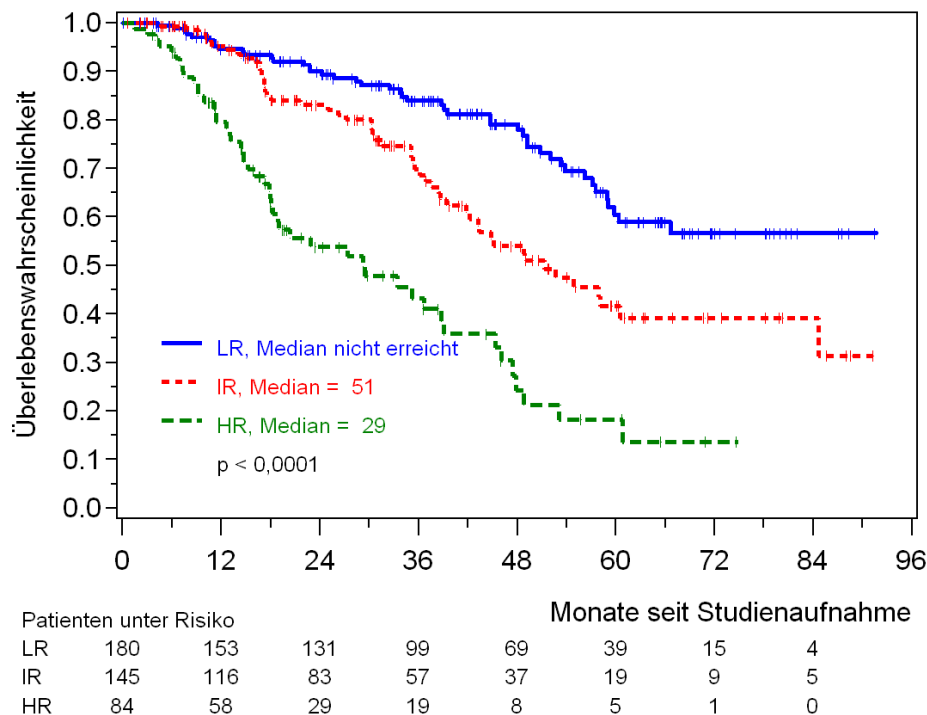


Abbildung 3.11: Gesamtüberleben der Risikogruppen nach dem neu entwickelten prognostischen Index (MIPI) auf dem Datensatz der Modellentwicklung mit 409 Patienten.

3.3.4 Vereinfachter prognostischer Index

In der Diskussion mit Klinikern erwies sich die Akzeptanz eines prognostischen Index, der auf einem kontinuierlichen Prognosescore und kontinuierlichen Prognosefaktoren beruht, als sehr begrenzt. Als einfachste und einzig praktikable Lösung erschien den Klinikern wie beim IPI und dem FLIPI nur bereits dichotomisierte Prognosefaktoren in das Modell aufzunehmen. Dem verständlichen Argument der schnellen und einfachen Anwendung am Krankenbett ohne Taschenrechner und Internet mussten jedoch biometrische Argumente gegen eine Dichotomisierung wie ein massiver Powerverlust sowie eine wenig realistische Modellierung der Effekte entgegen werden. Aus diesem Dilemma zwischen methodischer Validität und praktischer Anwendbarkeit heraus definierte ich einen vereinfachten prognostischen Index. Ziel war hierbei nicht die Entwicklung eines weiteren prognostischen Index, sondern eine Vereinfachung der Gruppenzuteilung, die größtmögliche Übereinstimmung mit dem MIPI erzielte. Die Übereinstimmung wurde mit dem Spearmanschen Rangkorrelationskoeffizient ρ und Cohen's gewichtetem κ quantifiziert.

Standard-Cutpoints

Zunächst untersuchte ich, inwiefern mit der Dichotomisierung der metrischen Prognosefaktoren unter Verwendung von Standard-Grenzwerten eine Vereinfachung des MIPI erreicht werden konnte. Für das Alter wird bei Patienten mit Non-Hodgkin-Lymphom in der Regel eine Altersgrenze von 60 Jahren gezogen, bis zu der Patienten mit aggressiveren Therapien, z.B. Stammzelltransplantation, behandelt werden können. Für die Leukozytenzahl gilt der Wert $10.000/\mu\text{l}$ als eine obere Norm. Die stark laborabhängigen Normalwerte für die LDH wurden bei jedem Patienten neben dem LDH-Wert dokumentiert.

Zählte man als Risikofaktoren ein Alter von 60 Jahren oder mehr, ein ECOG Performance Status 2 - 4, einen LDH-Wert über dem oberen Normwert und eine Leukozytenzahl größer als $10.000/\mu\text{l}$, so ergab sich im Gesamtkollektiv eine Einteilung von 98, 166, 100, 33 und 12 Patienten mit 0, 1, 2, 3 und 4 Risikofaktoren. Unter Berücksichtigung dieser Fallzahlen war es notwendig, Patienten mit 2 oder mehr Risikofaktoren zu einer Hochrisikogruppe zusammenzufassen. Patienten mit keinem Risikofaktor galten demnach als Low Risk, und Patienten mit einem Risikofaktor als Intermediate Risk. Mit dieser Einteilung ergab sich eine Verteilung der Patienten zu LR, IR und HR von 24%, 41% und 35%. Tabelle 3.15 zeigt den Vergleich dieser Gruppenzuordnung mit dem MIPI. Der Spearmansche Korrelationskoeffizient der Anzahl Risikofaktoren mit dem Prognosescore betrug $\rho = 0,83$ und der gewichtete Kappa-Koeffizient beim Vergleich mit dem MIPI $\kappa = 0,56$.

Risikogruppe	MIPI LR	MIPI IR	MIPI HR
LR	97	1	0
IR	75	84	7
HR	8	60	77

Tabelle 3.15: Dichotomisierung der Prognosefaktoren an Standard-Cutpoints: Vergleich der Patientenzuteilung mit dem MIPI.

Abbildung 3.12 zeigt das Gesamtüberleben in den so definierten Risikogruppen. Sowohl die Übereinstimmung mit dem MIPI als auch die Separation der Überlebenskurven waren nach Dichotomisierung der Prognosefaktoren mit den Standard-Cutpoints nicht ausreichend gut. Daher untersuchte ich eine differenziertere Kategorisierung der Prognosefaktoren.

Explorative Kategorisierung der Prognosefaktoren

Zur Vereinfachung der Zuteilung zu Risikogruppen wurde explorativ eine Kategorisierung der Prognosefaktoren versucht, die die ursprüngliche Gruppenzuteilung mit der Kategorisierung des Prognosescores besser reproduzieren sollte als nach Dichotomisierung der Prognosefaktoren. Dazu wurde für Alter, LDH und Leukozytenzahl eine Einteilung in jeweils vier Gruppen vorgeschlagen, wobei der mittlere Schwellenwert jeweils der Standard-Cutpoint war. Für den ECOG Performance Status wurde die Einteilung 0 - 1 und 2 - 4 beibehalten.

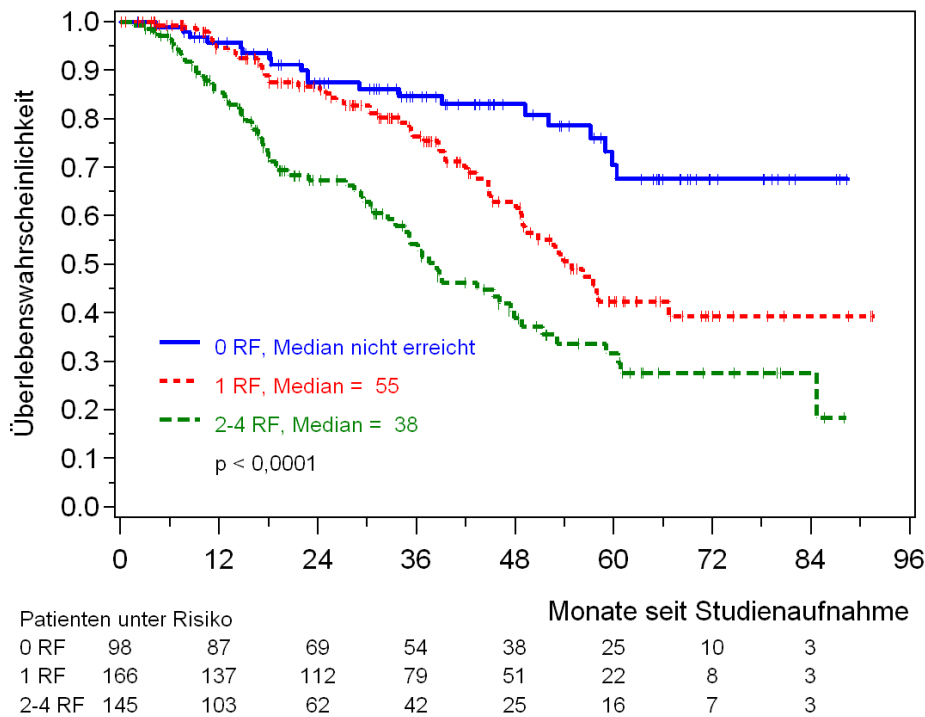


Abbildung 3.12: Gesamtüberleben nach Anzahl Risikofaktoren (RF) gemäß Standard-Cutpoints (Alter ≥ 60 , ECOG 2 - 4, LDH erhöht, Leukozyten $\geq 10.000 /\mu\text{l}$).

Für das Alter wurden die Grenzwerte 50, 60 und 70 bestimmt. Für den LDH-Quotient und die Leukozytenzahlen wurden wegen der Schiefe der Verteilungen asymmetrische Grenzwerte gewählt, und zwar $2/3$ des oberen Normalwertes, der obere Normalwert und $3/2$ des oberen Normalwertes. Auf Grund der Fallzahlen waren die Hälfte des oberen Normalwertes bzw. der doppelte obere Normalwert nicht geeignet. Für die Prognosefaktoren wurden dann nach folgendem Schema Risikopunkte vergeben:

Prognosefaktor	0 Punkte	1 Punkt	2 Punkte	3 Punkte
Alter (Jahre)	< 50	≥ 50 und < 60	≥ 60 und < 70	≥ 70
ECOG	0 oder 1	2 bis 4*	2 bis 4**	
LDH (% der oberen Norm)	< 67	≥ 67 und < 100	≥ 100 und < 150	≥ 150
Leukozytenzahl $/\mu\text{l}$	< 6.700	≥ 6.700 und < 10.000	≥ 10.000 und < 15.000	≥ 15.000

Tabelle 3.16: Risikopunkte für Werte der Prognosefaktoren.

* einfache, ** doppelte Gewichtung des ECOG Performance Status.

Im multiplen Cox-Modell mit den vier so kategorisierten Prognosefaktoren waren die geschätzten Regressionsparameter bei einfacher Gewichtung des ECOG Performance Status 0,3872, 0,7294, 0,2191 und 0,3608 für Alter, ECOG Performance Status, LDH und Leukozytenzahl. Der Regressionskoeffizient für den ECOG Performance Status war also deutlich höher, während die übrigen Parameter vergleichbar waren. Bei doppelter Gewichtung des ECOG Performance Status (2 Punkte für ECOG Performance Status 2 bis 4) ergaben sich ähnlichere Regressionsparameter 0,3872, 0,3647, 0,2191 und 0,3608 (Tabelle 3.17).

Variable	β	SE	p-Wert
Alter	0,3872	0,0965	0,0001
ECOG*	0,7294	0,2571	0,0045
ECOG**	0,3647	0,1285	0,0045
LDH	0,2191	0,0929	0,0183
Leukozyten	0,3608	0,0773	< 0,0001

Tabelle 3.17: Multiples Cox-Modell mit den kategorisierten Prognosefaktoren. β : Regressionskoeffizient, SE: Standardfehler von β .

* einfache, ** doppelte Gewichtung des ECOG Performance Status.

Die Summenwerte der Risikopunkte lagen bei doppelter Gewichtung des ECOG Performance Status zwischen 0 und 11, bei einfacher Gewichtung zwischen 0 und 10. Der Korrelationskoeffizient nach Spearman mit dem kontinuierlichen Prognosescore berechnete sich zu 0,91 bzw. 0,93 bei einfacher bzw. doppelter Gewichtung des ECOG Performance Status. Unter Berücksichtigung der minimalen und maximalen Fallzahlen waren jeweils zwei Varianten der Einteilungen in je drei Gruppen sinnvoll (Tabelle 3.18).

ECOG	Variante	Gruppe	Risikopunkte	Patienten	%
einfach gewichtet	Variante 1	Low Risk	0 - 2	87	21
		Intermediate Risk	3 - 4	176	43
		High Risk	5 - 10	146	36
einfach gewichtet	Variante 2	Low Risk	0 - 3	178	44
		Intermediate Risk	4 - 5	154	38
		High Risk	5 - 10	77	19
doppelt gewichtet	Variante 1	Low Risk	0 - 2	86	21
		Intermediate Risk	3 - 4	172	42
		High Risk	5 - 11	151	37
doppelt gewichtet	Variante 2	Low Risk	0 - 3	176	43
		Intermediate Risk	4 - 5	154	38
		High Risk	6 - 11	79	19

Tabelle 3.18: Vergleich der Varianten des vereinfachten prognostischen Index: Verteilung der Patienten zu Risikogruppen nach der Summe an Risikopunkten zu den kategorisierten Prognosefaktoren.

Variante 1 zeigte durchweg eine schlechtere Übereinstimmung mit der Einteilung nach dem kontinuierlichen Prognosescore als Variante 2 (Tabelle 3.19). Die doppelte Gewichtung des ECOG Performance Status mit 2 Risikopunkten für einen ECOG Performance Status von 2 - 4 zeigte ausgewogenere Regressionskoeffizienten im multiplen Cox-Modell (Tabelle 3.16) sowie eine bessere Übereinstimmung mit dem MIPI. Die Einteilung nach Variante 2 mit doppelter Gewichtung erschien daher als gute Approximation der Einteilung mit dem kontinuierlichen Prognosescore unter Kategorisierung der Prognosefaktoren und damit als Alternative zum Einsatz in der klinischen Praxis. Diese Zuordnung zu Risikogruppen war auch der Einteilung mit Dichotomisierung der Prognosefaktoren nach Standard-Cutpoints in allen Parametern überlegen. Abbildung 3.13 zeigt das Gesamtüberleben der drei Risikogruppen eingeteilt nach dem so vereinfachten prognostischen Index.

ECOG	Variante	Gruppe	MIPI LR	MIPI IR	MIPI HR	gewichtetes κ
einfach gewichtet	Variante 1	LR	87	0	0	0,54
		IR	91	80	5	
		HR	2	65	79	
einfach gewichtet	Variante 2	LR	153	25	0	0,77
		IR	27	112	15	
		HR	0	8	69	
doppelt gewichtet	Variante 1	LR	86	0	0	0,55
		IR	92	78	2	
		HR	2	67	82	
doppelt gewichtet	Variante 2	LR	153	23	0	0,79
		IR	27	114	13	
		HR	0	8	71	

Tabelle 3.19: Varianten des vereinfachten prognostischer Index: Vergleich mit dem MIPI.

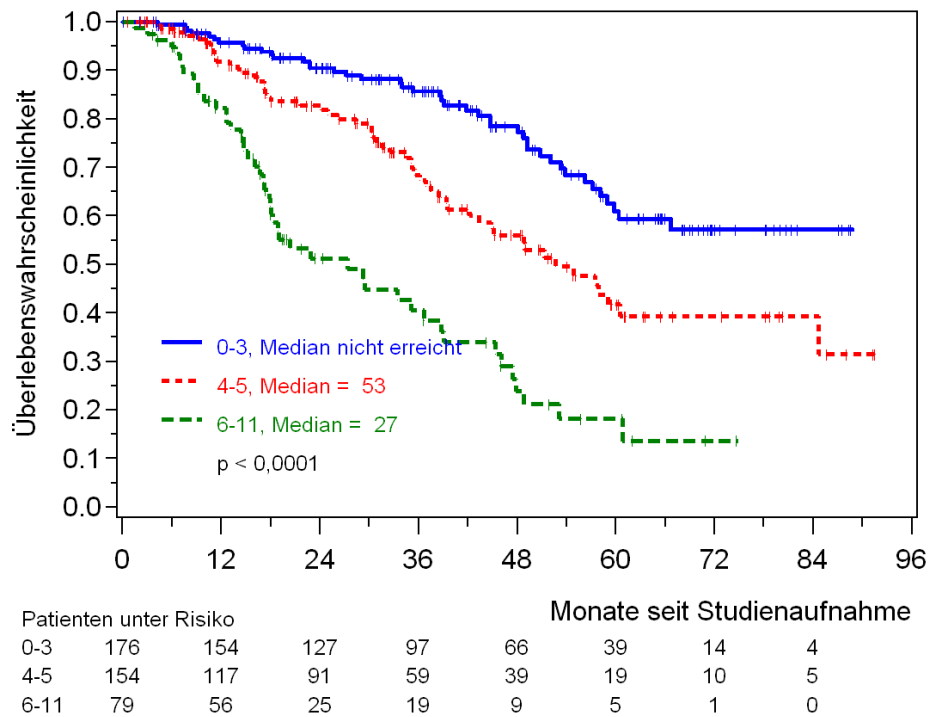


Abbildung 3.13: Gesamtüberleben nach der Summe von Risikopunkten aus Tabelle 3.16 bei doppelter Gewichtung des ECOG Performance Status (vereinfachter prognostischer Index).

3.4 Schätzung der Modellgüte

Nach der Modellentwicklung erfolgte die Schätzung der Modellgüte sowohl für das Regressionsmodell als auch für den prognostischen Index. Zunächst wurden die Maßzahlen der Modellgüte an dem zur Modellentwicklung verwendeten Datensatz überoptimistisch geschätzt. Die Gütemaße wurden daraufhin mit dem Bootstrap-Verfahren in Analogie zu [55] für eine mögliche Überanpassung an die vorliegende Stichprobe korrigiert (interne Validierung nach dem „regular bootstrap“).

3.4.1 Regressionsmodell

Die Güte des durch Variablenselektion ausgewählten Prognosemodells wurde an Hand dreier Maßzahlen bewertet, der Likelihood-Ratio (LikR) Statistik des Modells ohne Kovariaten im Vergleich zu der des Modells mit Kovariaten, der Steigung der Kalibrierungsgeraden (calibration slope) und des Konkordanzindex c nach Harrell. Das Separationsmaß SEP und der Brier Score konnten auf dieser Stufe nicht herangezogen werden, da noch keine Risikogruppen definiert waren. Ausgehend von 12 potenziellen Prädiktoren und 146 Ereignissen bei 409 Patienten lag die Zahl der Ereignisse pro Variable (EPV) bei 12,17. Dieser Wert lag über dem auch in [55] empfohlenen Wert von 10 EPV.

Die Likelihood-Ratio- χ^2 -Statistik reduzierte sich durch Bootstrap-Korrektur von 55,8 auf 46,6, die Steigung der Kalibrierungsgeraden von 1 auf 0,91 und der Konkordanzindex von 0,69 auf 0,68 (Tabelle 3.20). Im Vergleich zu den optimistischen Werten war die Korrektur für alle drei Maßzahlen relativ gering. Der Konkordanzindex war mit 0,68 relativ klein, was vermutlich an dem relativ hohen Anteil zensierter Überlebenszeiten liegt, deren Paare jeweils mit 0,5 eingehen.

Method	LikR- χ^2	slope	c
optimistisch	55,842	1	0,6867
bootstrap-korrigiert	46,624	0,907	0,6791

Tabelle 3.20: Optimistische und Bootstrap-korrigierte Schätzung der prognostischen Güte des Cox-Modells nach Variablenselektion. LikR: Likelihood-Ratio des Modells mit vs. ohne Kovariaten, slope: Steigung der Kalibrierungsgeraden.

3.4.2 Prognostischer Index

Für die Klassifikationsmodelle wurden die Logrank-Statistik, der Konkordanzindex c sowie das Separationsmaß SEP als Maßzahlen für die Modellgüte berechnet. Tabelle 3.21 zeigt zunächst für den MIPI die überoptimistisch geschätzten Maßzahlen und die entsprechenden Werte nach Bootstrap-Korrektur. Mit dem Bootstrap-Verfahren wurde hier die zweistufige Auswahl optimaler Cutpoints, unter Berücksichtigung der minimalen und

maximalen Gruppengröße, ausgehend von dem Cox-Modell mit den vier Kovariaten Alter, ECOG Performance Status, LDH und Leukozytenzahl validiert. Danach reduzierte sich der Wert der Logrank-Statistik durch die Bootstrap-Korrektur von 61 auf 48, der Konkordanzindex sank von 0,67 auf 0,65 und das Separationsmaß SEP von 1,59 auf 1,49. Auch nach Bootstrap-Korrektur war der Wert der Logrank-Statistik mit $p = 5 \cdot 10^{-11}$ hochsignifikant.

Modell	Schätzung	Logrank- χ^2	c	SEP
MIPI	optimistisch	61,11	0,6668	1,59
	bootstrap-korrigiert	47,64	0,6536	1,49
Vereinfachter Index Standard-Cutpoints	optimistisch	63,15	0,6716	1,60
	optimistisch	34,77	0,6393	1,51

Tabelle 3.21: Optimistische und bootstrap-korrigierte Schätzung der Modellgüte nach Einteilung in 3 Risikogruppen durch zweistufige Auswahl von Cutpoints für den Prognosescore (MIPI), sowie optimistische Schätzung für den vereinfachten Index und die Einteilung mit Standard-Cutpoints für Alter, LDH und Leukozytenzahl.

Darüber hinaus zeigt Tabelle 3.21 die überoptimistischen Werte für den vereinfachten Index. Logrank-Statistik, Konkordanzindex und Separationsmaß waren den optimistischen Werten des Original-MIPI vergleichbar. Für die vereinfachte Risikogruppen-Zuteilung mit dichotomisierten Prognosefaktoren an Standard-Cutpoints waren die Maßzahlen mit $\chi^2 = 35$ für die Logrank-Statistik, Konkordanzindex 0,64 und Separationsmaß 1,51, selbst überoptimistisch geschätzt, deutlich schlechter als für die Einteilung nach dem MIPI. In beiden Fällen konnte keine Bootstrap-Korrektur durchgeführt werden, da die Cutpoints a priori feststanden.

3.5 Überprüfung der Modellannahmen

3.5.1 Proportional-Hazards-Annahme

Mit den skalierten Schoenfeldresiduen wurde eine eventuelle Zeitabhängigkeit der Regressionschätzer für die vier Kovariaten Alter, ECOG Performance Status, LDH und Leukozytenzahl im endgültigen Cox-Modell untersucht. Die Residuen zeigten keinen klaren Trend und waren in den wesentlichen Zeitbereichen relativ konstant (Abbildung 3.14). Für den ECOG Performance Status zeigte sich bei den frühen Zeiten bis 20 Monate ein Abfallen und für die Leukozytenzahl im selben Zeitraum ein Ansteigen des Regressionskoeffizienten. Beide Veränderungen waren im statistischen Test einzeln grenzwertig signifikant ($p = 0,0478$ und $p = 0,0546$, Tabelle 3.22), der globale Test für die Abweichung von der Annahme proportionaler Risiken fiel mit $p = 0,0957$ dagegen nicht signifikant aus. Somit musste die Annahme proportionaler Risiken für das endgültige Prognosemodell nicht verworfen werden.

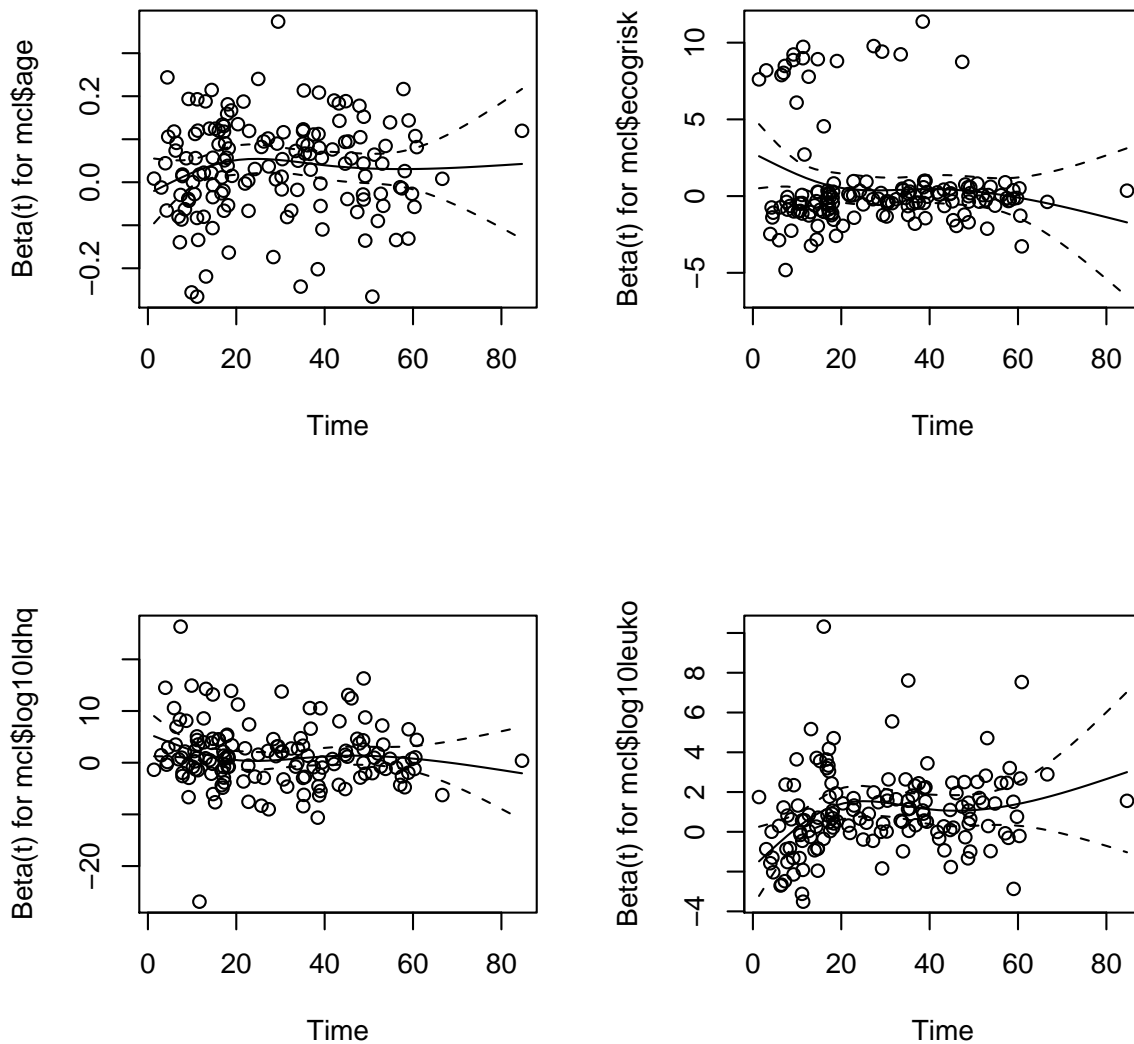


Abbildung 3.14: Skalierte Schoenfeldresiduen für das Cox-Modell mit den vier Risikofaktoren Alter (links oben), ECOG Performance Status (rechts oben), LDH-Quotient (links unten) und Leukozytenzahl (rechts unten) mit LOWESS-Glättung zur Überprüfung der Proportional-Hazards-Annahme.

Parameter	Rho	χ^2	p -Wert
Alter	0,0374	0,187	0,6652
ECOG > 1	-0,1642	3,916	0,0478
log-LDH-Quotient	-0,0913	1,269	0,2599
log-Leukozyten	0,2104	3,694	0,0546
GLOBAL		7,865	0,0957

Tabelle 3.22: Test auf Proportional Hazards nach Therneau et al. [60] im Cox-Modell mit den vier Risikofaktoren Alter, ECOG Performance Status, LDH und Leukozyten.

3.5.2 Linearität

Wie die Martingalresiduen des Cox-Modells ohne Kovariaten zeigten, war die Linearitätsannahme für Alter, logarithmierter LDH-Quotient und logarithmierte Leukozytenzahl im wesentlichen Wertebereich erfüllt (Abbildung 3.15). Sowohl für das Alter als auch für die Leukozytenzahl zeigte sich jedoch eine gewisse Sättigung in den Bereichen sehr hoher Werte. Die Linearität war für LDH-Quotient und die Leukozytenzahl bei den logarithmierten Werten klarer ausgeprägt, so dass auch nach Bewertung der Martingalresiduen eine Logarithmierung bevorzugt werden sollte.

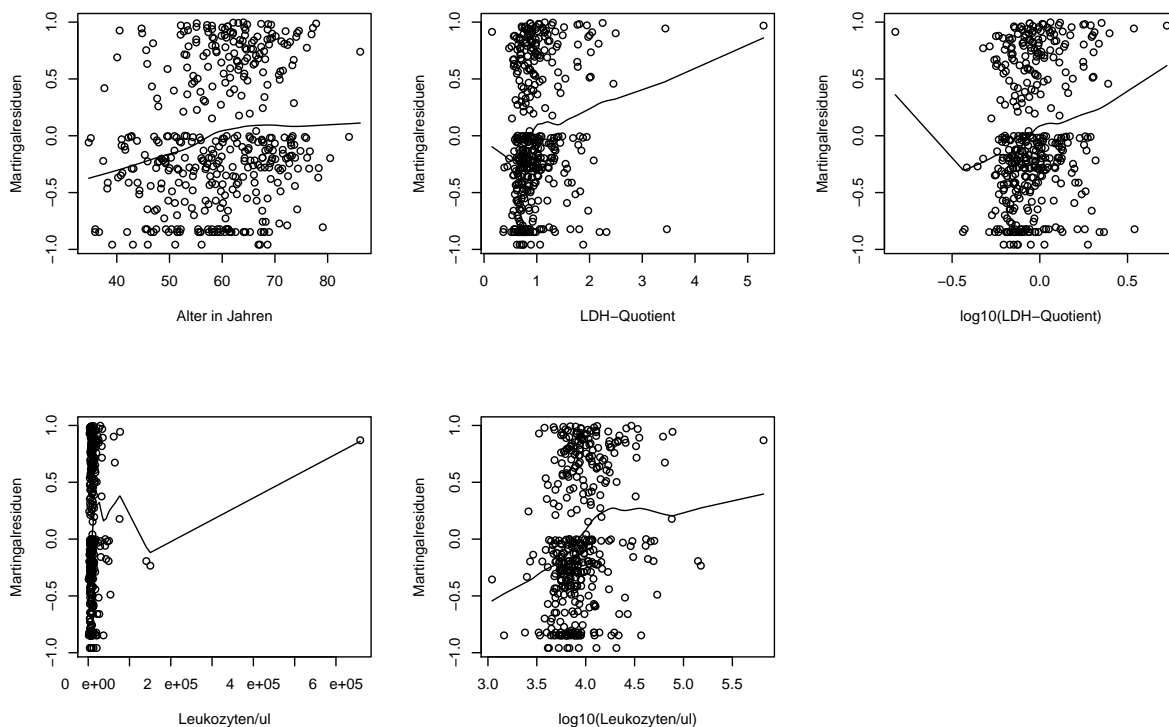


Abbildung 3.15: Martingalresiduen mit LOWESS-Glättung zur Überprüfung der Linearitätsannahme im Cox-Modell bei den metrischen Kovariaten Alter, LDH (roh und logarithmiert) und Leukozytenzahl (roh und logarithmiert).

3.5.3 Extrembeobachtungen

Abbildung 3.16 zeigt die Score-Residuen, die zur Beurteilung von Beobachtungen mit starkem Effekt auf die Regressionsschätzer des endgültigen Cox-Modells mit den vier Kovariaten Alter, ECOG Performance Status, LDH und Leukozytenzahl berechnet wurden. Bei den metrischen Parametern zeigte sich die typische Schmetterlingsform, die darauf hinweist, dass das Weglassen von Beobachtungen am Rand des Wertebereichs der Kovariaten die Regressionsparameter stärker beeinflusst als von Beobachtungen mit mittleren

Kovariatenwerten. Die meisten Ausreißer zeigten negative Score-Residuen, was bedeutet, dass deren Weglassen im Cox-Modell zu betragsmäßig höheren Schätzern für die Regressionsparameter führte. Damit schienen die Regressionsparameter unter Einbeziehung auch dieser Extrembeobachtungen eher konservativ geschätzt worden sein. Extrembeobachtungen wurden auf Datenfehler überprüft, es musste aber kein Wert korrigiert werden.

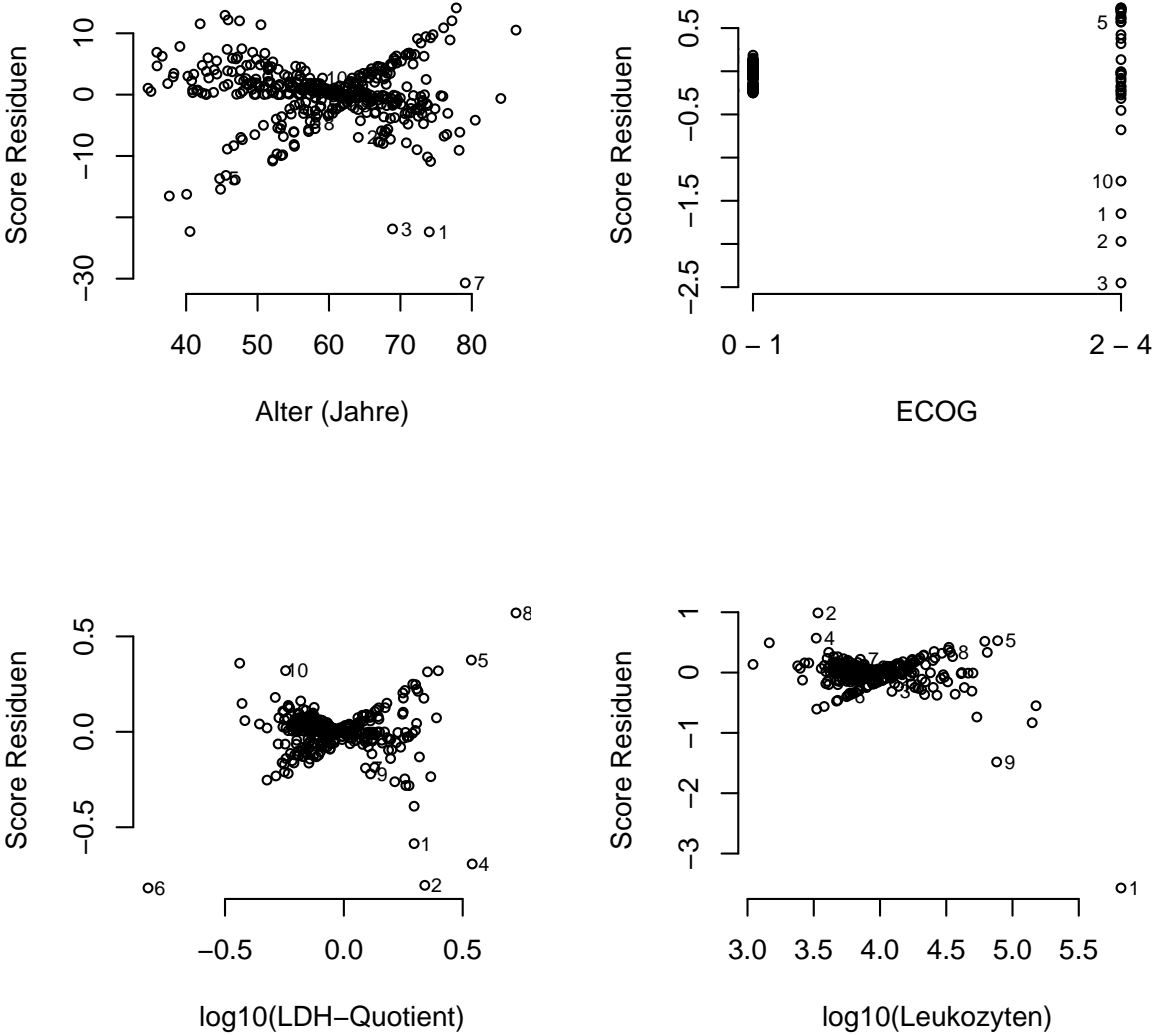


Abbildung 3.16: Score-Residuen zur Identifizierung von Individuen mit extremem Einfluss auf die Regressionskoeffizienten. Die mit Nummern versehenen Extrembeobachtungen sind im Text beschrieben.

Einige Patienten zeigten gleichzeitig bezüglich mehrerer Kovariaten extreme Score-Residuen. Ein Patient war bei allen Kovariaten unter den vier kleinsten Residuen-Werten (Nr. 1 in Abbildung 3.16). Dieser Patient hatte eine relativ kurze Überlebenszeit bei hohem Alter, schlechtem Allgemeinzustand, deutlich erhöhter LDH, und hoher Leukozytenzahl.

Ein weiterer Patient zeigte extreme Score-Residuen bezüglich ECOG Performance Status, LDH und Leukozytenzahl (Nr. 2). Dieser Patient hatte eine relativ lange zensierte Überlebenszeit bei schlechtem Allgemeinzustand, deutlich erhöhter LDH und einer niedrigen Leukozytenzahl. Drei Patienten hatten bei zwei Kovariaten extreme Score-Residuen. Ein Patient hatte bei schlechtem Allgemeinzustand und einem Alter von 68 Jahren eine relativ lange zensierte Überlebenszeit (Nr. 3). Ein Patient hatte bei stark erhöhter LDH und niedriger Leukozytenzahl eine relativ lange zensierte Überlebenszeit (Nr. 4). Ein dritter Patient hatte bei stark erhöhter LDH und hohen Leukozytenzahlen eine sehr kurze Überlebenszeit (Nr. 5). Schließlich fielen noch fünf Beobachtungen mit extremen Score-Residuen bezüglich einer Kovariate auf. Ein Patient hatte bei sehr kleiner LDH auch eine sehr kurze Überlebenszeit (Nr. 6), ein weiterer bei einem Alter von 79 Jahren eine relativ lange zensierte Überlebenszeit (Nr. 7), und ein dritter bei stark erhöhter LDH eine sehr kurze Überlebenszeit (Nr. 8). Ein Patient mit einer relativ langen Überlebenszeit hatte auch eine hohe Leukozytenzahl (Nr. 9). Ein Patient mit schlechtem Allgemeinzustand hatte dennoch eine lange zensierte Überlebenszeit (Nr. 10).

3.6 Weitere explorative Analysen

3.6.1 Effekt der Auswahl vollständiger Datensätze

Gesamtüberleben

Die Überlebenskurven der Patienten mit vollständigen (complete cases, CC) und mit unvollständigen Datensätzen (NCC) verliefen sehr ähnlich. Das mediane Gesamtüberleben war 57 Monate für die CC und 63 Monate für die NCC (Abbildung 3.17). Es konnte kein signifikanter Unterschied im Gesamtüberleben beider Gruppen entdeckt werden ($p = 0,82$, Logrank-Test). Somit unterschieden sich die Patienten mit vollständigen Daten von denen mit unvollständigen Daten in ihrer Prognose nicht und es gab keinen Hinweis für einen dahin gehenden Selektionsbias.

Univariate Analyse potenzieller Prognosefaktoren

Regressionskoeffizienten und Signifikanz der potenziellen Prognosefaktoren waren in der univariaten Analyse auf dem Datensatz der 409 complete cases (Tabelle 3.23) ähnlich wie auf dem Gesamtdatensatz von 455 Patienten (Tabelle 3.9). Somit führte die Auswahl vollständiger Datensätze nicht zu einer wesentlichen Veränderung der univariaten Relevanz potenzieller Prognosefaktoren, so dass es auch hier keinen Hinweis auf einen Selektionsbias gab.

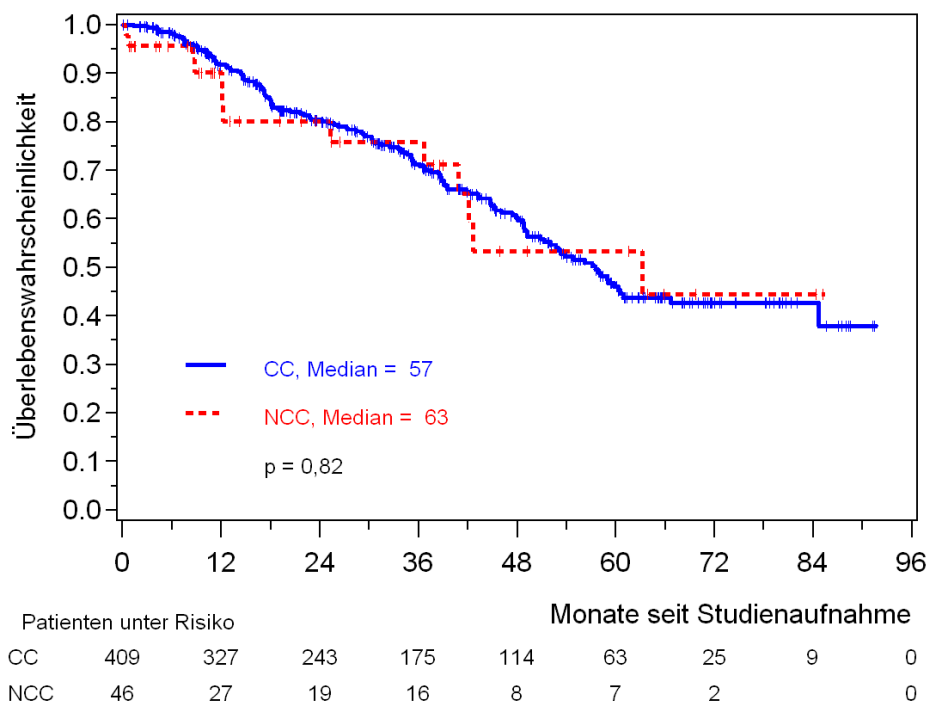


Abbildung 3.17: Vergleich des Gesamtüberlebens der 409 bezüglich der 11 ausgewählten potenziellen Prognosefaktoren vollständigen Datensätzen (complete cases, CC) mit den nicht vollständigen Datensätzen (NCC).

Parameter	Vergleich	n	HazR (95%-KI)	p-Wert
Alter	+ 10 Jahre	409	1,47 (1,23 - 1,77)	< 0,0001
Geschlecht	männlich	409	1,08 (0,73 - 1,60)	0,71
ECOG	3 - 4 vs. 0 - 1	409	3,15 (1,94 - 5,12)	< 0,0001
Ann-Arbor-Stadium	IV vs. III	409	1,13 (0,72 - 1,79)	0,59
B-Symptome	vorhanden	409	1,55 (1,12 - 2,15)	0,0079
Milzbefall	vorhanden	380	1,49 (1,06 - 2,09)	0,0228
Knochenmarkbefall	vorhanden	409	1,14 (0,76 - 1,73)	0,52
extranodale Befälle	> 1 vs. 0 - 1	409	1,38 (0,98 - 1,93)	0,0640
LK-Regionen	+ 1	409	1,03 (0,97 - 1,09)	0,33
Tumorgröße	+ 1cm	359	1,06 (1,02 - 1,11)	0,0081
LDH-Quotient*	+ 1	409	1,74 (1,33 - 2,27)	< 0,0001
	10-fach		6,03 (2,44 - 14,90)	0,0001
Hb	+ 10g/l	409	0,91 (0,85 - 0,98)	0,0090
Leukozyten	+ 1.000/ μ l	409	1,004 (1,001 - 1,006)	0,0029
	10-fach	409	3,56 (2,25 - 5,63)	< 0,0001
Lymphozyten	+ 1.000/ μ l	385	1,004 (1,001 - 1,007)	0,0085
	10-fach		2,02 (1,42 - 2,88)	0,0001
Thrombozyten	+ 10 ⁵ / μ l	409	1,00 (0,86 - 1,15)	0,9556
Albumin-Quotient*	+ 1	246	0,26 (0,05 - 1,29)	0,0983
β_2 -Mikroglobulin-Quotient*	+ 1	259	1,22 (1,04 - 1,42)	0,0123
	2-fach		1,44 (1,14 - 1,82)	0,0025

Tabelle 3.23: Ergebnis der univariaten Cox-Regression der wichtigsten potenziellen Prognosefaktoren auf dem reduzierten Datensatz der Modellentwicklung von 409 Patienten. HazR: Hazard-Ratio, KI: Konfidenzintervall. n: Anzahl der Werte.

*Quotient zum oberen Normalwert.

Milzbefall

Milzbefall zeigte mit einer Hazard-Ratio von 1,05 (95%-KI 0,72 - 1,53, $p = 0,80$) keine von den vier identifizierten Prognosefaktoren Alter, ECOG Performance Status, LDH und Leukozytenzahl unabhängige prognostische Relevanz. Regressionskoeffizienten und Signifikanz der vier unabhängigen Prognosefaktoren waren bei den 380 Patienten mit Angaben zum Milzbefall ähnlich wie bei den vollständigen Datensätzen.

Knochenmarksbefall

Mit einer Hazard-Ratio von 0,89 (95%-KI 0,58 - 1,35, $p = 0,57$) zeigte der Knochenmarksbefall keine zusätzliche prognostische Relevanz. Im vollständigen Datensatz ($n = 409$) waren auch die Angaben zum Knochenmarksbefall vollständig.

Maximale Tumorgröße

Die maximale Tumorgröße zeigte keine unabhängige prognostische Relevanz bei einer Hazard-Ratio von 1,04 pro cm (95%-KI 0,997 - 1,09, $p = 0,0665$). Regressionskoeffizienten und Signifikanz der vier unabhängigen Prognosefaktoren waren bei den 359 Patienten mit Angabe zur maximalen Tumorgröße ähnlich wie bei den vollständigen Datensätzen.

Lymphozytenzahl

Die Lymphozytenzahl zeigte keine unabhängige prognostische Relevanz bei einer Hazard-Ratio von 0,61 für eine 10-fache Erhöhung (95%-KI 0,32 - 1,18, $p = 0,14$). Auch bei den 385 Patienten mit Angabe zur Lymphozytenzahl waren Regressionskoeffizienten und Signifikanz der vier unabhängigen Prognosefaktoren ähnlich wie bei den vollständigen Datensätzen.

Granulozytenzahl

Die Granulozytenzahl zeigte keine unabhängige prognostische Relevanz bei einer Hazard-Ratio von 2,26 für eine 10-fache Erhöhung (95%-KI 0,89 - 5,73, $p = 0,0846$). Bei den 378 Patienten mit Angabe zur Granulozytenzahl waren Regressionskoeffizienten und Signifikanz der vier unabhängigen Prognosefaktoren ähnlich wie bei den vollständigen Datensätzen.

Monozytenzahl

Die Monozytenzahl zeigte keine unabhängige prognostische Relevanz bei einer Hazard-Ratio von 1,73 für eine 10-fache Erhöhung (95%-KI 0,97 - 3,07, $p = 0,0614$). Auch bei den 372 Patienten mit Angabe zur Monozytenzahl waren Regressionskoeffizienten und

Signifikanz der vier unabhängigen Prognosefaktoren ähnlich wie bei den vollständigen Datensätzen.

Albumin

Bei den 246 vollständigen Datensätzen zu den vier Prognosefaktoren und dem Albumin-Quotienten zum oberen Normalwert zeigten ECOG Performance Status und LDH keine unabhängige prognostische Relevanz mehr und deutlich kleinere Regressionskoeffizienten. Regressionskoeffizienten und Signifikanz der vier Prognosefaktoren änderten sich dann nicht bei Hinzunahme des Albumin-Quotienten zum oberen Normalwert. Albumin zeigte hierbei mit einer Hazard-Ratio von 0,48 für eine Erhöhung um den oberen Normalwert (95%-KI 0,085 - 2,65, $p = 0,40$) keine zusätzliche prognostische Relevanz.

β_2 -Mikroglobulin

Bei den 259 vollständigen Datensätzen zu den vier Prognosefaktoren und dem Quotienten von β_2 -Mikroglobulin zum oberen Normalwert zeigten die vier Prognosefaktoren ähnliche Regressionskoeffizienten wie in den 409 vollständigen Datensätzen, mit $p = 0,0866$ verlor allerdings die LDH ihre Signifikanz. Bei Hinzunahme des β_2 -Mikroglobulin-Quotienten zum oberen Normalwert änderten sich dann Regressionskoeffizienten und Signifikanz der vier Prognosefaktoren nicht und β_2 -Mikroglobulin zeigte keine unabhängige prognostische Relevanz (Hazard-Ratio 1,004 für eine Verdoppelung des Wertes, 95%-KI 0,74 - 1,36, $p = 0,98$).

3.6.2 Wechselwirkungen

Zu den vier unabhängigen Prognosefaktoren des endgültigen Cox-Regressionsmodells wurden die sechs paarweisen Wechselwirkungen auf ihre Signifikanz im multiplen Cox-Modell mit den vier Haupteffekten überprüft. Nach der Rückwärts-Selektion ausgehend von dem Modell mit allen 6 Wechselwirkungstermen und den vier Haupteffekten erwies sich als einzige die Wechselwirkung zwischen Alter und ECOG Performance Status als signifikant (Tabelle 3.24)

Parameter	Vergleich	β	SE	HazR (95%-KI)	p-Wert
Alter	+ 10 Jahre	0,4228	0,1015	1,53 (1,25 - 1,86)	< 0,0001
ECOG	2 - 4 vs. 0 - 1	5,056	1,864	157 (4,06 - 6059)	0,0067
LDH	10-fach	1,389	0,4879	4,01 (1,54 - 10,44)	0,0044
Leukozyten	10-fach	1,047	0,2366	2,85 (1,79 - 4,53)	< 0,0001
Alter*ECOG	10-fach	-0,6965	0,3019	0,50 (0,28 - 0,90)	0,0210

Tabelle 3.24: Cox-Regressionsmodell mit den vier Haupteffekten Alter, ECOG Performance Status, LDH und Leukozyten sowie der nach Rückwärts-Selektion auf dem 5%-Niveau signifikanten Wechselwirkung zwischen Alter und ECOG Performance Status (Alter*ECOG). β : Regressionskoeffizient, SE: Standardfehler von β , HazR: Hazard-Ratio, KI: Konfidenzintervall.

Mit dem Wert $-0,01044$ der Kovarianz zwischen dem Alter und dem Wechselwirkungsterm Alter mit ECOG Performance Status errechnet sich die Hazard-Ratio für um zehn Jahre ältere Patienten im Stratum der Patienten mit ECOG Performance Status 0 - 1 zu 1,53, 95%-KI 1,25 - 1,86, und im Stratum der Patienten mit ECOG Performance Status 2 - 4 zu 0,76, 95%-KI 0,44 - 1,33, d.h. im Stratum der arbeitsunfähigen oder bettlägerigen Patienten zeigte sich kein signifikanter Effekt des Alters auf das Gesamtüberleben mehr. Die Likelihood-Ratio- χ^2 -Statistik erhöhte sich von 55,84 auf 60,75 durch Hinzunahme der Wechselwirkung, der Konkordanzindex war mit $c = 0,6847$ jedoch nicht besser als ohne Wechselwirkung $c = 0,6867$.

Auf dem 5%-Niveau war der Wechselwirkungsterm zwischen Alter und ECOG Performance Status signifikant und biologisch könnte ein Fehlen eines Alterseffektes bei sehr schlechtem ECOG Performance Status durchaus plausibel sein. Dennoch erschien mir die Höhe des p -Wertes in Anbetracht der Anzahl durchgeführter statistischer Tests und im Vergleich zur Signifikanz der Haupteffekte nicht klein genug, um den Wechselwirkungsterm in das endgültige Modell aufzunehmen. Zudem befanden sich im Stratum der arbeitsunfähigen oder bettlägerigen Patienten auf Grund der Einschlusskriterien der Studien nur 42 Patienten, die endgültige Beurteilung der Wechselwirkung sollte sich daher erst auf einem größeren Patientenkollektiv mit mehr arbeitsunfähigen oder bettlägerigen Patienten bestätigen.

3.6.3 Alter und Allgemeinzustand

Modell ohne Alter und ECOG Performance Status

Nach Rückwärts-Selektion auf dem 5%-Niveau unter Ausschluss von Alter und ECOG Performance Status erwiesen sich wieder LDH und Leukozytenzahl als unabhängige Prognosefaktoren, und kein weiterer Parameter zeigte davon unabhängige prognostische Relevanz (Tabelle 3.25). Regressionskoeffizienten und Signifikanz waren dem Gesamtmodell mit Alter und ECOG Performance Status (Tabelle 3.11) vergleichbar.

Parameter	Vergleich	β	SE	HazR (95%-KI)	p -Wert
LDH	10-fach	1,585	0,4808	4,88 (1,90 - 12,52)	0,0010
Leukozytenzahl	10-fach	1,107	0,2289	3,03 (1,93 - 4,74)	< 0,0001

Tabelle 3.25: Unabhängig signifikante Einflussgrößen auf das Gesamtüberleben nach Rückwärtsselektion mit Signifikanzniveau 5% unter Ausschluss von Alter und ECOG Performance Status. β : Regressionskoeffizient, SE: Standardfehler von β , HazR: Hazard-Ratio, KI: Konfidenzintervall.

Patienten unter 65 Jahre

Auch in der Untergruppe der 278 Patienten, die jünger als 65 Jahre waren, zeigte der neue prognostische Index eine gute Auftrennung der Überlebenskurven für das Gesamtüberle-

ben mit Median nicht erreicht (LR), 45 Monate (IR) und 27 Monate (HR, Abbildung 3.18).

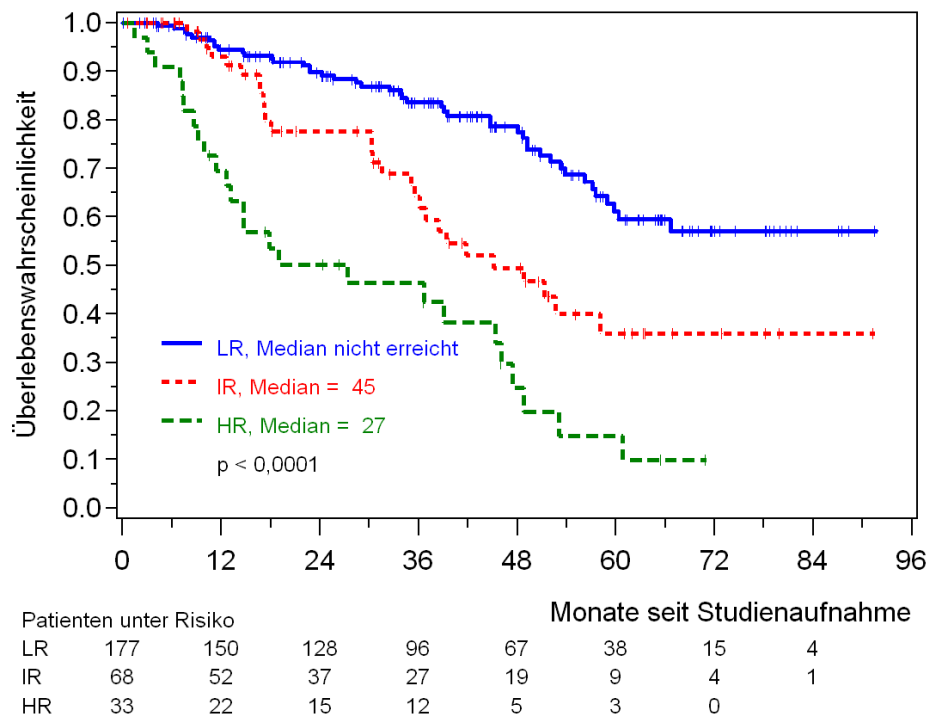


Abbildung 3.18: Gesamtüberleben der Risikogruppen nach dem neu entwickelten prognostischen Index (MIPI) bei Patienten unter 65 Jahren.

Parameter	Vergleich	β	SE	HazR (95%-KI)	p -Wert
Alter	+ 10 Jahre	0,5835	0,1676	1,79 (1,29 - 2,49)	0,0005
ECOG	2 - 4 vs. 0 - 1	0,8779	0,3076	2,41 (1,32 - 4,40)	0,0043
LDH	10-fach	1,123	0,5619	3,07 (1,02 - 9,24)	0,0457
Leukozyten	10-fach	1,294	0,3061	3,65 (2,00 - 6,65)	< 0,0001

Tabelle 3.26: Prognostische Relevanz der vier Prognosefaktoren im multiplen Cox-Regressionsmodell nach Rückwärts-Selektion für die 278 Patienten unter 65 Jahren. β : Regressionskoeffizient, SE: Standardfehler von β , HazR: Hazard-Ratio, KI: Konfidenzintervall.

Die Rückwärts-Selektion aller potenzieller Prognosefaktoren auf dem 5%-Niveau auf dem Datensatz der unter 65-jährigen ergab dieselben vier unabhängigen Prognosefaktoren (Tabelle 3.26). Insbesondere zeigte das Alter auch in dieser Untergruppe einen signifikanten Einfluss auf das Gesamtüberleben.

3.6.4 Zeit bis zum Therapieversagen

Nach dem neuen prognostischen Index trennten sich auch die Kaplan-Meier-Kurven für die Zeit bis zum Therapieversagen auf mit einer medianen TTF von 27 Monaten für die Low Risk-, 18 Monaten für die Intermediate Risk- und 11 Monaten für die High Risk-Gruppe (Abbildung 3.19).

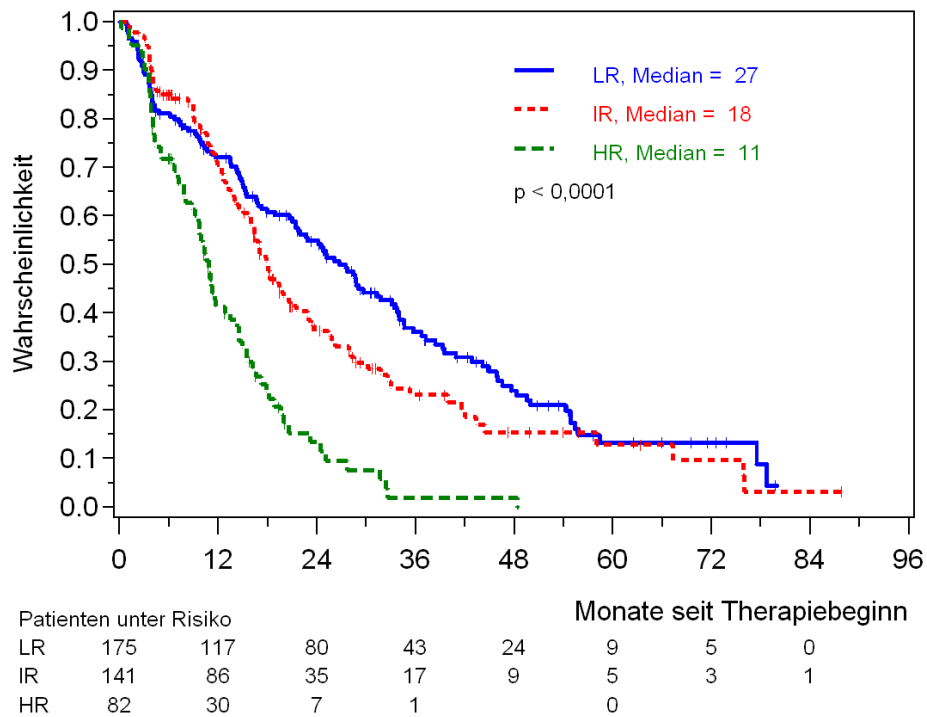


Abbildung 3.19: Zeit bis zum Therapieversagen in den Risikogruppen nach dem neu entwickelten prognostischen Index auf dem Datensatz der Modellentwicklung mit 409 Patienten.

Von den unabhängigen Prognosefaktoren für das Gesamtüberleben zeigten ECOG Performance Status, LDH und Leukozytenzahl signifikanten Einfluss auf die Zeit bis zum Therapieversagen (Tabelle 3.27). Das Alter zeigte dagegen weder im multiplen noch im univariaten Cox-Regressionsmodell einen Einfluss auf die Zeit bis zum Therapieversagen.

Parameter	Vergleich	HazR (95%-KI)	p-Wert
Alter	+ 10 Jahre	1,10 (0,98 - 1,24)	0,12
ECOG	2 - 4 vs. 0 - 1	1,82 (1,18 - 2,79)	0,0062
LDH	10-fach	1,45 (1,18 - 1,78)	0,0004
Leukozytenzahl	10-fach	1,76 (1,21 - 2,58)	0,0035

Tabelle 3.27: Ergebnis der multiplen Cox-Regression der vier Prognosefaktoren des neuen prognostischen Index auf die Zeit bis zum Therapieversagen. HazR: Hazard-Ratio, KI: Konfidenzintervall.

Kapitel 4

Diskussion

4.1 Kurzzusammenfassung der Ergebnisse

Im Rahmen der vorliegenden Arbeit wurde ein neuer prognostischer Index für Patienten mit fortgeschrittenem Mantelzell-Lymphom (Mantle Cell Lymphoma International Prognostic Index, MIPI) entwickelt, der auf Basis der vier unabhängigen Prognosefaktoren Alter, Allgemeinzustand, LDH und Leukozytenzahl Patienten in drei Risikogruppen mit unterschiedlichem Gesamtüberleben einteilt. Die Ergebnisse beruhen auf dem größten bisher publizierten Kollektiv von Patienten mit Mantelzell-Lymphom. Die Daten stammen aus randomisierten Therapiestudien mit zentraler Referenzpathologie und standardisierter Behandlung und weisen daher, wie von Altman [62] für Prognosestudien gefordert, große Homogenität in Bezug auf Diagnose, Dokumentation und Therapie auf. Die interne Validierung mit dem Bootstrap-Verfahren bestätigte die Stabilität des prognostischen Modells und die Trennung der Überlebenskurven der drei Risikogruppen.

Der MIPI ist der erste prognostische Index für Patienten mit Mantelzell-Lymphom. In der klinischen Praxis wurde bislang meist der International Prognostic Index (IPI) verwendet, der auf Daten von Patienten mit aggressiven Lymphomen beruht, zu denen das Mantelzell-Lymphom nicht zählte. Die Überprüfung der prognostischen Wertigkeit des IPI bei Patienten mit fortgeschrittenem Mantelzell-Lymphom ergab eine ungünstige Verteilung der Patienten auf die Risikogruppen mit einer sehr kleinen Hochrisikogruppe und zwei großen Gruppen mittleren Risikos, deren Überlebenskurven schlecht getrennt verliefen. Ein unabhängiger Prognosefaktor des IPI, die Anzahl extranodaler Befälle, wies im vorliegenden Kollektiv keine prognostische Relevanz auf. Auch der für Patienten mit follikulärem Lymphom entwickelte Follicular Lymphoma International Prognostic Index (FLIPI) wies in Anwendung auf das Kollektiv von Patienten mit Mantelzell-Lymphom mit 63% Hochrisikopatienten eine ungünstige Verteilung der Patienten auf die Risikogruppen und eine mangelnde Trennung der Überlebenskurven von Patienten niedrigen und mittleren Risikos auf. Zwei unabhängige Prognosefaktoren des FLIPI, die Anzahl befall-

ner Lymphknotenregionen und der Hämoglobinwert, wiesen keine oder keine unabhängige prognostische Relevanz in dem Patientenkollektiv fortgeschrittenen Mantelzell-Lymphoms auf.

Von den potenziellen Prognosefaktoren zeigten Alter, Allgemeinzustand, B-Symptome, Milzbefall, maximale Lymphknotengröße, LDH, Hämoglobin, Leukozyten-, Lymphozyten-, Granulozyten- und Monozytenzahl, sowie β_2 -Mikroglobulin univariaten Einfluss auf das Gesamtüberleben in dem Kollektiv von Patienten mit fortgeschrittenem Mantelzell-Lymphom. Dagegen hatten Geschlecht, Stadium, Knochenmarksbefall, Anzahl extranodaler Befälle oder befallener Lymphknotenregionen, Thrombozytenzahl, Albumin, sowie die Immunglobuline IgM, IgG und IgA keinen univariaten Einfluss auf das Gesamtüberleben.

4.2 Vergleich mit publizierten Ergebnissen

4.2.1 Ergebnisse zum IPI

Im Vergleich zur Original-Publikation des IPI [29] war in dem vorliegenden Kollektiv sowohl die Niedrigrisiko- als auch die Hochrisikogruppe kleiner (35% vs. 23% und 16% vs. 9%). Alle in Tabelle 1.7 dargestellten Publikationen zum Mantelzell-Lymphom überprüften die prognostische Relevanz des IPI. Auch in den genannten Publikationen umfasste die Gruppe niedrigen Risikos meist um die 20% der Patienten oder weniger [14], [15], [16], [18], [17], [9], [21], [11]. Hierbei spiegelte sich vermutlich der insgesamt höhere Anteil an Patienten in fortgeschrittenen Stadien III oder IV und das höhere mediane Alter beim Mantelzell-Lymphom im Vergleich zu den diffus-großzelligen Lymphomen wider. Der Anteil Patienten hohen Risikos nach IPI war in den meisten der genannten Publikationen größer als in dem vorliegenden Kollektiv, der Anteil variierte zwischen 10% und 29%. Eine Erklärung könnten hier die Einschlusskriterien der Studien bieten, welche nur ausreichend therapiefähige Patienten zuließen. Insgesamt schien der IPI jedoch auch in den anderen Kollektiven die Patienten mit Mantelzell-Lymphom im Vergleich zu den aggressiven Lymphomen in [29] eher in die mittleren Risikogruppen zu klassifizieren.

Wie in dem vorliegenden Patientenkollektiv zeigte der IPI in fast allen genannten Publikationen einen signifikanten Einfluss auf das Gesamtüberleben. Eine Auftrennung von vier Überlebenskurven schien jedoch wie in der vorliegenden Validierung meist nicht möglich zu sein. Meist wurden bereits a priori die Risikogruppen zu drei [17], [19] oder zwei [14], [8], [18], [21], [22] zusammengefasst. Die Beobachtung, dass die beiden mittleren Risikogruppen ein ähnliches Gesamtüberleben zeigten, machten die Autoren explizit auch in [17] bei 94 Patienten, in [20] bei 65 Patienten und vermutlich auch in [19] bei 68 Patienten.

Im multiplen Cox-Modell mit den Risikofaktoren des IPI bestätigte sich im vorliegenden Datensatz die unabhängige prognostische Relevanz von Alter, ECOG Performance

Status und LDH, während die Anzahl extranodaler Manifestationen nicht signifikant war und das Stadium nicht berücksichtigt werden konnte. Alter und LDH waren auch in den meisten untersuchten Publikationen sowohl univariat als auch multivariat signifikant (s. Abschnitt 1.2.2). Die Anzahl extranodaler Manifestationen war in wenigen univariaten Analysen signifikant, und in keiner multivariaten. Damit scheinen die neuen Ergebnisse im Wesentlichen die bereits auf kleineren Datensätzen gewonnenen Erkenntnisse zu bestätigen und zu erweitern.

4.2.2 Ergebnisse zum FLIPI

Mit 6% bzw. 64% der Patienten niedrigen bzw. hohen Risikos nach dem FLIPI zeigte sich im vorliegenden Kollektiv eine noch extremere Patientenverteilung als bei der einzigen publizierten Untersuchung [9] zum FLIPI beim Mantelzell-Lymphom mit 23% bzw. 53%. In der Original-Publikation des FLIPI [30] war bei Patienten mit follikulärem Lymphom die Patientenverteilung mit 36% LR, 37% IR und 27% HR wesentlich homogener. Die Unterschiede zwischen meinen Ergebnissen und denen aus [9] beruhen vermutlich wieder auf den Einschlusskriterien der Studien, vor allem da es sich bei [9] um eine bevölkerungsbezogene Studie handelt. Meine Ergebnisse bestätigen jedoch die bereits in [9] gemachte Beobachtung einer kleineren Niedrigrisikogruppe und vor allem einer großen Hochrisikogruppe.

Wie in der vorliegenden Analyse zeigte der FLIPI in [9] einen signifikanten Einfluss auf das Gesamtüberleben. Jedoch erschien wie in meiner Analyse die Trennung der Gruppen niedrigen und mittleren Risikos auch in [9] nicht überzeugend. In univariaten Analysen zeigten in Übereinstimmung mit meinen Ergebnissen die Risikofaktoren Hämoglobin und ECOG Performance Status einen Einfluss aufs Gesamtüberleben, im Gegensatz zu meinen Ergebnissen auch die Anzahl befallener Lymphknotenregionen. Die multiple Cox-Regression in [9] war schwer zu interpretieren, da sie neben den prognostischen Indizes IPI und FLIPI auch deren definierende und teilweise übereinstimmende Risikofaktoren enthielt, und damit mehrfache Korrelationen der Parameter zu erwarten waren.

Ganz im Gegensatz zu [9] erwies sich in meinen Analysen der FLIPI in Bezug auf die Patientenverteilung und die Trennung der Überlebenskurven nicht als dem IPI überlegen. Darüber hinaus zeigten die Analysen der Einzelfaktoren einerseits widersprüchliche Ergebnisse zur Bedeutung der Anzahl befallener Lymphknotenregionen, deren Abgrenzung für den FLIPI schwierig und schlecht reproduzierbar sein kann. Zum anderen bestätigten die Ergebnisse die prognostische Relevanz des ECOG Performance Status, der in die Entwicklung des FLIPI nicht miteinbezogen werden konnte.

4.2.3 Prognosefaktoren

In den univariaten Analysen zeigten die folgenden potenziellen Prognosefaktoren keinen Einfluss auf das Gesamtüberleben: Geschlecht, Stadium (III vs. IV), Knochenmarksbefall, Anzahl extranodaler Befälle, Anzahl befallener Lymphknotenregionen, Thrombozytenzahl und Immunglobuline. Die Ergebnisse zu Prognosefaktoren beim Mantelzell-Lymphom in der Literatur wurden bereits in Tabelle 1.7 sowie in den Abschnitten 1.2.2, 1.2.3, und 1.2.4 zusammengefasst. Hierbei zeigte das Geschlecht in keiner Analyse einen signifikanten Einfluss [14], [8], [17], [18], [20], [21]. Das Stadium wurde in allen Publikationen bis auf [21] in der Kategorisierung I+II vs. III+IV untersucht. In [21] zeigte sich für Patienten mit Stadium IV kein signifikanter Überlebensnachteil im Vergleich zu den Patienten in den Stadien I, II oder III. Das Vorhandensein von Knochenmarksbefall zeigte in fünf von acht Analysen einen Einfluss auf das Gesamtüberleben [14], [17], [19], [22], [11]. Die Anzahl extranodaler Befälle war in drei [14], [18], [11] von zehn Analysen signifikant, davon bei zwei in der Unterscheidung kein extranodaler Befall vs. mindestens einer. Die Anzahl befallener Lymphknotenregionen wurde nur in [9] im Rahmen der Validierung des FLIPI untersucht, und zeigte dort einen signifikanten Einfluss. Die Thrombozytenzahl kam in nur drei Publikationen als potenzieller Prognosefaktor in Betracht und war nur bei [8] univariat signifikant. Von den Immunglobulinen wurde nur in [8] das IgG auf seine prognostische Relevanz überprüft und zeigte dort keinen Einfluss. Bis auf den Knochenmarksbefall zeigten alle univariat nicht signifikanten Parameter auch in der Literatur meist keine univariate prognostische Relevanz für das Gesamtüberleben. Da aber alle Patienten mit Knochenmarksbefall im Stadium IV sind, spiegelt dieses Ergebnis möglicherweise den Effekt eines fortgeschrittenen Stadiums wider, zumal sehr wenige Patienten mit fortgeschrittenem Stadium im Stadium III waren.

In der vorliegenden Arbeit zeigten dagegen die folgenden Parameter univariat signifikanten Einfluss auf das Gesamtüberleben: Alter, Allgemeinzustand, B-Symptomatik, Milzbefall, Tumorgröße, Leukozyten-, Lymphozyten-, Granulozyten- und Monozytenzahl, Hb, LDH, Albumin und β_2 -Mikroglobulin. Auch in der Literatur zeigten Alter, Allgemeinzustand, LDH und Hb relativ übereinstimmend eine prognostische Bedeutung in univariaten Analysen. B-Symptomatik war in vier von acht Analysen relevant, Milzbefall in drei von sechs Analysen, die Tumorgröße (Bulk > 10cm) und Albumin in keiner von je zwei, und β_2 -Mikroglobulin in zwei von zwei (s. Tabelle 1.7). Leukozyten-, Lymphozytenzahl oder der Befall des peripheren Blutes wurden in einigen Analysen als potenzielle prognostische Faktoren berücksichtigt [8], [17], [16], [10], [21], [22] und zeigten bis auf den Befall des peripheren Blutes in [21], [22] univariate prognostische Relevanz.

In 9 von 13 der untersuchten Publikationen zu Prognosefaktoren beim Mantelzell-Lymphom wurde zusätzlich zur univariaten Cox-Regression für das Gesamtüberleben eine multiple Regression durchgeführt. Die Ergebnisse hierzu sind schwer vergleichbar, da

unterschiedliche Parameter in die Regression eingebracht wurden und manchmal nicht dargestellt wurde, welche Parameter berücksichtigt wurden, oder ob und wie eine Variableselektion durchgeführt wurde. Darüber hinaus waren bei allen Analysen mit deutlich unter 10 Events pro Variable die Power zu gering, um solide Schlüsse aus der multiplen Regression zu ziehen [32]. Fünf der 9 Publikationen schlossen neben den Einzelfaktoren des IPI auch den IPI selbst als potenziellen Prognosefaktor mit in die Regression ein [8], [9], [18], [20], [21]. Diese Ergebnisse sind ebenfalls schwer zu interpretieren, da der IPI als Kovariate sicher von den Einzelfaktoren abhängig ist und Korrelationen zwischen den Kovariaten in multiplen Regressionen zu schlecht kontrollierbaren Ergebnissen führen können.

In den übrigen vier Publikationen erwies sich wie in meiner Analyse das Alter auch in der multiplen Regression als unabhängiger Prognosefaktor. Bei den beiden Modellen, die keine histopathologischen Parameter berücksichtigten [16], [17] blieb daneben der Befall des peripheren Blutes als unabhängiger Prognosefaktor bestehen. Über die Ergebnisse der univariaten Analysen hinaus könnte diese Beobachtung die Identifizierung des neuen unabhängigen Prognosefaktors Leukozytenzahl beim Mantelzell-Lymphom stützen. Zwar geht ein Befall des peripheren Blutes nicht zwingend, wohl aber häufig mit einer Leukozytose einher. Die Interpretation der Bedeutung des Befalls der peripheren Blutes wird aber durch die fehlende Beschreibung der Diagnose eines solchen Befalls erschwert. Bei [16] bedeutete eine Lymphozytenzahl von über $4000/\mu\text{l}$ ebenfalls einen Befall des peripheren Blutes, und die Leukozytenzahl war im vorliegenden Datensatz stark mit der Lymphozytenzahl korreliert. In [17] dagegen erwiesen sich Leukozytenzahl und Lymphozytenzahl in Anwesenheit der übrigen Parameter als keine unabhängigen Prognosefaktoren, was auch an der Korrelation mit dem Vorhandensein von Befall des peripheren Blutes liegen kann. Mit der Leukozytenzahl enthält der neue prognostische Index damit einen leichter und eindeutiger bestimmbar Parameter als die Lymphozytenzahl oder den Befall des peripheren Blutes. Darüber hinaus erwiesen sich in Übereinstimmung mit meinen Ergebnissen in [16] der Allgemeinzustand nach dem ECOG Performance Status und in [17] der LDH-Wert als unabhängiger Prognosefaktor.

4.3 Mögliche Limitationen

4.3.1 Patientenkollektiv und Verallgemeinerbarkeit

Gemäß den Ein- und Ausschlusskriterien der betrachteten Studien umfasste das zur Analyse verwendete Patientenkollektiv nur Patienten mit Mantelzell-Lymphom in fortgeschrittenem Stadium III oder IV, bei denen eine potenziell kurative Strahlentherapie nicht durchführbar war. Darüber hinaus mussten die Patienten in einem ausreichend guten Allgemeinzustand sein (ECOG Performance Status maximal 2, außer lymphombedingt) und

es durften keine Begleiterkrankungen oder andere Kontraindikationen gegen die gemäß Protokoll durchzuführenden Therapien bestehen. Diese Therapien waren jedoch im Wesentlichen Standard-Therapien, die im Rahmen der Studien in Wirksamkeit und Verträglichkeit verglichen wurden.

Im Vergleich zu den vier bevölkerungsbasierten Studien zu Patienten mit Mantelzell-Lymphom [8], [9], [10], [11] war das mediane Alter der Patienten im vorliegenden Patientenkollektiv fünf bis acht Jahre niedriger. Alle Patienten waren im Stadium III oder IV, im Vergleich zu 80% bis 86% der Patienten in den vier genannten Publikationen. Mit 9% der Patienten gab es im Studienkollektiv deutlich weniger arbeitsunfähige oder bettlägerige Patienten, mit einem ECOG Performance Status von 2 oder mehr, als in den bevölkerungsbasierten Studien mit 20% bis 28%. Damit enthielt das Kollektiv einerseits nicht Patienten mit einer möglicherweise guten Prognose auf Grund eines frühen Stadiums, und andererseits auch nicht Patienten mit einer möglicherweise sehr schlechten Prognose auf Grund mangelnder Therapierbarkeit.

Daher dürfen die Ergebnisse dieser Arbeit nur auf Patienten verallgemeinert werden, die diesem Risikoprofil entsprechen. Dies galt jedoch im Wesentlichen auch schon für den IPI und den FLIPI, denn auch dort wurden keine bevölkerungsbezogenen Daten sondern Daten aus klinischen Institutionen bzw. Studien verwendet. Auch die Anwendungen der Ergebnisse werden sich auf therapiefähige Patienten entweder in Kliniken oder in klinischen Studien konzentrieren. Die vorliegenden Ergebnisse gelten insbesondere auch nur für Patienten mit fortgeschrittenem Stadium III oder IV. Beim Mantelzell-Lymphom ist der Anteil von Patienten in limitierten Stadien I oder II jedoch vergleichsweise gering, und es gab keine vergleichbare Studie, die diese Patientengruppe untersuchte. Darüber hinaus sind für diese Patienten lokalisierte Therapiestrategien vorgesehen im Gegensatz zu den systemischen Therapien beim fortgeschrittenen Lymphom.

Das zur Analyse verwendete Kollektiv definierte sich vor allem dadurch, dass bei den Patienten die protokollgerechten Therapien durchführbar sein mussten. Im Gegenzug konnte zur Entwicklung des Prognosemodells ein relativ homogen behandeltes und dokumentiertes Patientenkollektiv verwendet werden, ein zum Beispiel von Altman [62] gefordertes wichtiges Qualitätskriterium an eine Prognosestudie.

Der neue prognostische Index trennte auch in der Untergruppe der unter 65-jährigen Patienten die Überlebenskurven der Risikogruppen auf. Darüber hinaus ergab die Rückwärts-Selektion in der Cox-Regression in dieser Untergruppe dieselben unabhängigen Prognosefaktoren. Im Gegensatz zum IPI war es in der vorliegenden Analyse daher nicht erforderlich, einen altersadjustierten prognostischen Index für die jüngeren Patienten zu entwickeln.

4.3.2 Methodik

Auswahl der Fälle und Variablen zur Modellentwicklung

Bei der Entwicklung des neuen prognostischen Index mussten einige potenziellen Prognosefaktoren weggelassen werden, da sonst die Auswahl der vollständigen dokumentierten Patienten zu einem deutlich reduzierten Datensatz geführt hätte. Dazu ist zum einen zu sagen, dass die wesentlichen in der Literatur dargestellten potenziellen Prognosefaktoren fast vollständig dokumentiert waren (Alter, Geschlecht, Stadium, Knochenmarksbefall, ECOG Performance Status, B-Symptomatik, extranodale Befälle, befallene Lymphknotenregionen, LDH, Hämoglobin, Leukozytenzahl und Thrombozytenzahl). Die Angaben des Differentialblutbildes Lymphozyten-, Granulozyten- und Monozytenzahl waren mit um die 90% etwas schlechter dokumentiert. Jedoch bestand erwartungsgemäß eine hohe Korrelation insbesondere der Lymphozytenzahl mit der Leukozytenzahl. Albumin und β_2 -Mikroglobulin sowie die Immunglobuline waren mit über 30% fehlenden Werten am schlechtesten dokumentiert. Albumin zeigte jedoch bereits univariat keinen signifikanten Einfluss auf das Gesamtüberleben. Die Immunglobuline stellten in der univariaten Analyse und in der Literatur keine wesentlichen Prognosefaktoren beim Mantelzell-Lymphom dar.

Milzbefall, Knochenmarksbefall, Albumin und β_2 -Mikroglobulin zeigten jeweils in einem Regressionsmodell mit den vier identifizierten Prognosefaktoren keine zusätzliche prognostische Relevanz. Das gleiche gilt für die maximale Tumorgröße, Lymphozytenzahl, Granulozytenzahl und Monozytenzahl, wobei hier auf Grund des Signifikanzniveaus ein kleinerer Effekt möglicherweise übersehen wurde. Im Hinblick auf die Natur des Lymphoms steht die Lymphozytenzahl im peripheren Blut möglicherweise in einem direkteren kausalen Zusammenhang zum Fortschreiten der Erkrankung als die Leukozytenzahl. Dennoch wurde das Prognosemodell mit der Leukozytenzahl entwickelt, da sie leichter und eindeutiger zur Verfügung steht und weitgehend als Parameter der leukämischen Generalisation gilt.

Mit der Auswahl potenzieller Prognosefaktoren konnte auch eine im Rahmen der Fallzahl akzeptable [32] Zahl von ca. 12 Ereignissen pro Variable erreicht werden. Durch die Auswahl der sehr gut dokumentierten Parameter wurde auch sichergestellt, dass in das prognostische Modell nur Parameter aufgenommen wurden, die in der Routineversorgung leicht verfügbar sind, was die praktische Anwendbarkeit des Modells erleichtert. Der reduzierte Datensatz auf die vollständigen Fälle umfasste knapp 90% der Patienten. Die potenziellen Prognosefaktoren zeigten auch auf dem reduzierten Datensatz eine ähnliche univariate prognostische Relevanz wie auf dem Gesamtdatensatz. Darüber hinaus zeigten die Patienten des Kollektivs der Modellentwicklung kein signifikant unterschiedliches Gesamtüberleben von den nicht zur Modellentwicklung verwendeten Patienten.

Statistische Methoden

Bei der Modellentwicklung kam mit der Cox-Regression ein etabliertes Verfahren zur Anwendung. Die Voraussetzungen des Cox-Modells wurden überprüft. Die Bildung von Risikogruppen erfolgte nach der „Minimum-p-Value“-Methode in Anwendung auf den Prognosescore des endgültigen Cox-Modells. Die „Minimum-p-Value“-Methode wird von statistischer Seite oft kritisiert [45], [44], da mit ihr ein hohes Risiko falsch positiver Ergebnisse einhergehen kann. Ich glaube, dass in meiner Anwendung dieses Problem minimiert ist. Ich suchte nach Grenzwerten zur Kategorisierung des Prognosescores, und nicht der Kovariaten. Dieses Vorgehen ist nach [44] durchaus sinnvoll. Der signifikante lineare Einfluss des Prognosescores auf das Gesamtüberleben als kontinuierlicher Parameter war nach der Cox-Regression bereits gegeben. Unter dieser Annahme muss prinzipiell jeder Cut-point, der eine relevante Patientenzahl abgrenzt, signifikant unterschiedliche Gruppen in Bezug auf das Gesamtüberleben definieren können. Als Auswahlkriterium für Schwellenwerte erschien mir die maximale statistische Trennung in Bezug auf das Gesamtüberleben als wichtigstes Ziel. Ich suchte daher nach Grenzwerten, die die Logrank-Statistik maximierten. Die Zahl der potenziellen Grenzwerte war im Vergleich zur erreichten Signifikanz relativ gering, so dass die konservative Bonferroni-Adjustierung keine Änderung der Ergebnisse nach sich zog. Schließlich bestätigte die Bootstrap-Validierung die Trennung der Überlebenskurven der drei Risikogruppen des neuen prognostischen Index.

Die interne Validierung mit dem Bootstrap-Verfahren war technisch leicht durchzuführen und bestätigte die signifikante Auftrennung der drei Überlebenskurven der MIPI-Risikogruppen unter Korrektur für die Überanpassung an die zur Modellentwicklung verwendeten Daten. Weitere Maßzahlen für die Güte des Prognosemodells zeigten auch nach Bootstrap-Korrektur akzeptable Werte. Wie bereits in der Literatur dargestellt [48], beleuchtete jede dieser Maßzahlen unterschiedliche Aspekte der Modellgüte und keine war daher aussagekräftig genug, allein zur Bewertung herangezogen zu werden. Die geforderte Überprüfung der Modellgüte gestaltete sich damit trotz des verwendeten Standard-Verfahrens der Cox-Regression wegen der zensierten Zielgröße als schwierig.

Mit dem zur Verfügung stehenden Datensatz war es nicht möglich, eine externe Validierung durchzuführen. Dazu waren die Studien zu homogen im Aufbau, es gab kein a-priori-Argument, mit dem man einen externen Datensatz hätte definieren können. Außerdem wollte ich bei der seltenen Entität die Fallzahl zur Modellentwicklung maximal erhalten. Mit der Bootstrap-Validierung führte ich jedoch ein internes Validierungsverfahren durch, das eine möglichst hohe Power aus den verfügbaren Daten zog [51]. Die externe Validierung ist an den Daten der laufenden Studien für Patienten mit Mantelzell-Lymphom des European MCL Network sowie an den Daten anderer Studiengruppen geplant.

4.3.3 Mögliche Confounder

Der wichtigste mögliche Confounder der vorliegenden Ergebnisse ist vermutlich die Therapie, die standardisiert gegeben wurde, jedoch im randomisierten Vergleich mehrere Möglichkeiten zuließ. Knapp ein Drittel der Patienten hatte seit 2000 im Rahmen der initialen Chemotherapie Rituximab erhalten. 80 Patienten unter 66 Jahren mit Ansprechen auf die Induktion waren autolog transplantiert worden. Rituximab hat in der Therapie der diffus-großzelligen und der folliculären Lymphome einen Durchbruch erzielt, der auch zu einer Verlängerung des Gesamtüberlebens geführt hat [63], [35]. Beim Mantelzell-Lymphom ist bislang keine Verbesserung des Gesamtüberlebens durch Rituximab in der Primärtherapie beobachtet worden [34]. Daher gehe ich davon aus, dass Rituximab direkt keinen Einfluss auf die Bedeutung von Prognosefaktoren für das Gesamtüberleben hat. Hierfür spricht auch, dass die Aufnahme von Rituximab-haltiger Chemotherapie als Kovariate zum gleichen prognostischen Modell führte wie ohne Berücksichtigung diese Kovariate. Jedoch ist die Beobachtungsdauer von Patienten, die mit Rituximab behandelt wurden, noch kurz. Eine Validierung des neuen prognostischen Index an einem homogen mit Rituximab behandelten Kollektiv von Patienten mit Mantelzell-Lymphom ist daher wichtig und im Rahmen der laufenden Studien des European MCL Network geplant.

Die autologe Stammzelltransplantation konnte bei Patienten mit Mantelzell-Lymphom einen signifikanten Einfluss auf die Remissionsdauer zeigen [37], der sich jedoch wiederum nicht auf das Gesamtüberleben übertrug. Die Transplantation konnte nicht ohne weiteres als mögliche Kovariate im Prognosemodell berücksichtigt werden, da es sich um einen zeitabhängigen Parameter handelt, der darüber hinaus nur für jüngere Patienten mit Ansprechen auf die Primärtherapie in Frage kam. Ich gehe aber ebenfalls davon aus, dass die Stammzelltransplantation auf Grund bislang fehlenden Einflusses auf das Gesamtüberleben keinen starken Confounder darstellt.

Die betrachteten Prognosefaktoren können durchaus Surrogatparameter für biologische Prognosefaktoren sein. Mit dem Ziel, einen klinischen prognostischen Index zu entwickeln, der auf leicht verfügbaren Parametern beruht, meine ich alle relevanten Größen in die Analyse einbezogen zu haben. In der Literatur wird für Patienten mit Mantelzell-Lymphom oft die Zellproliferation als wichtiger Prognosefaktor berichtet und diskutiert [14], [15], [8], [18], [11]. In [10] und [19] fordern die Autoren eine Überprüfung der prognostischen Relevanz biologischer Parameter auf größeren Datensätzen und in [14] und [15] die Entwicklung eines aus klinischen und biologischen Parametern kombinierten Prognosemodells. Bei knapp der Hälfte der Patienten des vorliegenden Kollektivs standen mir Proliferationsdaten, gemessen im Anteil Ki-67 positiver Zellen, zur Verfügung. Dennoch wollte ich die Proliferation nicht in das Prognosemodell aufnehmen, da sie nicht von genügend vielen Patienten verfügbar war und die Proliferationsbestimmung bislang nicht standardisiert ist.

4.4 Schlussfolgerungen und Ausblick

Im Rahmen der vorliegenden Arbeit entwickelte ich den ersten prognostischen Index speziell für Patienten mit fortgeschrittenem Mantelzell-Lymphom. Die Ergebnisse beruhen auf dem größten bisher publizierten Kollektiv von Patienten dieser Entität. Die Überprüfung der prognostischen Wertigkeit des für aggressive Lymphome entwickelten IPI, des für folliculäre Lymphome entwickelten FLIPI sowie des vom Morbus Hodgkin übernommenen Ann-Arbor-Stadiums beim Mantelzell-Lymphom unterstreicht die schlechte Übertragbarkeit der Ergebnisse von Prognosestudien einzelner Lymphomentitäten auf andere und damit die Notwendigkeit der Entwicklung getrennter prognostischer Indizes für die unterschiedlichen Lymphomentitäten.

Die vorliegenden Ergebnisse können verschiedene Anwendungsmöglichkeiten bieten. Mit Hilfe der identifizierten Prognosefaktoren können Patientenkollektive fortgeschrittenen Mantelzell-Lymphoms auf ihre Risikostruktur hin beurteilt und studienübergreifend verglichen werden. In Beobachtungsstudien können Effekte adjustiert für Prognosefaktoren geschätzt werden. In klinischen Studien kann für den neuen prognostischen Index oder einzelne Prognosefaktoren stratifiziert randomisiert werden, um Strukturgleichheit vor allem bei kleineren Patientenzahlen zu erreichen.

Der neue prognostische Index beruht bewusst auf einfach verfügbaren aber auch relativ unspezifischen klinischen Parametern. Mit dem technischen Fortschritt in der Diagnostik sind und werden in der Zukunft Prognosefaktoren bestimmbar, die spezifischer die Biologie des Mantelzell-Lymphoms widerspiegeln, wie zum Beispiel Genexpressionsdaten. Die vorliegenden Ergebnisse können als klinische Referenz dienen, um die Unabhängigkeit und Überlegenheit bei der Etablierung solcher biologischer Prognosefaktoren zu gewährleisten.

Schließlich könnten die Ergebnisse dieser Arbeit bei der Risiko-Nutzen-Abwägung für die therapeutische Entscheidung herangezogen werden, wenn auch nie als alleiniges Kriterium. Somit könnten die Ergebnisse einen Beitrag in Richtung einer risikoadaptierten, individualisierten Therapie für Patienten mit dieser schweren Erkrankung leisten.

Teile dieser Arbeit wurden vorgetragen auf

der gemeinsame Jahrestagung „Evaluation 2006“ der Deutschen Region der Internationalen Biometrischen Gesellschaft und des Deutschen Netzwerks Evidenz-basierte Medizin e.V. am 7. März 2006 in Bochum,

der 48. Jahrestagung der American Society of Hematology am 12. Dezember 2006 in Orlando, Florida, USA,

dem 113. Kongress der Deutschen Gesellschaft für Innere Medizin am 17. April 2007 in Wiesbaden (Young Investigators' Award),

der Gemeinsamen Jahrestagung der Deutschen, Österreichischen und Schweizerischen Gesellschaften für Hämatologie und Onkologie am 8. Oktober 2007 in Basel, Schweiz.

Teile dieser Arbeit wurden veröffentlicht in [64] und als Abstract in [65].

Literaturverzeichnis

- [1] Leonhardt, H. (1990). Histologie, Zytologie und Mikroanatomie des Menschen. Stuttgart, Thieme.
- [2] Harris, N. L., Jaffe, E. S., Stein, H., Banks, P. M., Chan, J. K., Cleary, M. L., Delsol, G., De Wolf-Peeters, C., Falini, B., Gatter, K. C., et al. (1994). A revised European-American classification of lymphoid neoplasms: a proposal from the International Lymphoma Study Group. *Blood* 84(5): 1361-92.
- [3] Lennert, K., Stein, H., Kaiserling, E. (1975). Cytological and functional criteria for the classification of malignant lymphomata. *Br J Cancer* 31 SUPPL 2: 29-43.
- [4] Tolksdorf, G., Stein, H., Lennert, K. (1980). Morphological and immunological definition of a malignant lymphoma derived from germinal-centre cells with cleaved nuclei (centrocytes). *Br J Cancer* 41(2): 168-82.
- [5] (1982). National Cancer Institute sponsored study of classifications of non-Hodgkin's lymphomas: summary and description of a working formulation for clinical usage. The Non-Hodgkin's Lymphoma Pathologic Classification Project. *Cancer* 49(10): 2112-35.
- [6] Harris, N. L., Jaffe, E. S., Diebold, J., Flandrin, G., Muller-Hermelink, H. K., Vardiman, J., Lister, T. A., Bloomfield, C. D. (1999). World Health Organization classification of neoplastic diseases of the hematopoietic and lymphoid tissues: report of the Clinical Advisory Committee meeting-Airlie House, Virginia, November 1997. *J Clin Oncol* 17(12): 3835-49.
- [7] (2000). International Classification of Diseases for Oncology. 3rd Edition, Geneva, World Health Organization.
- [8] Andersen, N. S., Jensen, M. K., de Nully Brown, P., Geisler, C. H. (2002). A Danish population-based analysis of 105 mantle cell lymphoma patients: incidences, clinical features, response, survival and prognostic factors. *Eur J Cancer* 38(3): 401-8.
- [9] Møller, M. B., Pedersen, N. T., Christensen, B. E. (2006). Mantle cell lymphoma: prognostic capacity of the Follicular Lymphoma International Prognostic Index. *Br J Haematol* 133(1): 43-9.

- [10] Argatoff, L. H., Connors, J. M., Klasa, R. J., Horsman, D. E., Gascoyne, R. D. (1997). Mantle cell lymphoma: a clinicopathologic study of 80 cases. *Blood* 89(6): 2067-78.
- [11] Velders, G. A., Kluin-Nelemans, J. C., De Boer, C. J., Hermans, J., Noordijk, E. M., Schuurin, E., Kramer, M. H., Van Deijk, W. A., Rahder, J. B., Kluin, P. M., Van Krieken, J. H. (1996). Mantle-cell lymphoma: a population-based clinical study. *J Clin Oncol* 14(4): 1269-74.
- [12] (2004). *Maligne Lymphome. Empfehlungen zur Diagnostik, Therapie und Nachsorge.* München, Tumorzentrum München.
- [13] (1997). A clinical evaluation of the International Lymphoma Study Group classification of non-Hodgkin's lymphoma. The Non-Hodgkin's Lymphoma Classification Project. *Blood* 89(11): 3909-18.
- [14] Tiemann, M., Schrader, C., Klapper, W., Dreyling, M. H., Campo, E., Norton, A., Berger, F., Kluin, P., Ott, G., Pileri, S., Pedrinis, E., Feller, A. C., Merz, H., Janssen, D., Hansmann, M. L., Krieken, H., Möller, P., Stein, H., Unterhalt, M., Hiddemann, W., Parwaresch, R. (2005). Histopathology, cell proliferation indices and clinical outcome in 304 patients with mantle cell lymphoma (MCL): a clinicopathological study from the European MCL Network. *Br J Haematol* 131(1): 29-38.
- [15] Rätty, R., Franssila, K., Joensuu, H., Teerenhovi, L., Elonen, E. (2002). Ki-67 expression level, histological subtype, and the International Prognostic Index as outcome predictors in mantle cell lymphoma. *Eur J Haematol* 69(1): 11-20.
- [16] Samaha, H., Dumontet, C., Ketterer, N., Moullet, I., Thieblemont, C., Bouafia, F., Callet-Bauchu, E., Felman, P., Berger, F., Salles, G., Coiffier, B. (1998). Mantle cell lymphoma: a retrospective study of 121 cases. *Leukemia* 12(8): 1281-7.
- [17] Oinonen, R., Franssila, K., Teerenhovi, L., Lappalainen, K., Elonen, E. (1998). Mantle cell lymphoma: clinical features, treatment and prognosis of 94 patients. *Eur J Cancer* 34(3): 329-36.
- [18] Schrader, C., Meusers, P., Brittinger, G., Teymoortash, A., Siebmann, J. U., Janssen, D., Parwaresch, R., Tiemann, M. (2004). Topoisomerase IIalpha expression in mantle cell lymphoma: a marker of cell proliferation and a prognostic factor for clinical outcome. *Leukemia* 18(7): 1200-6.
- [19] Weisenburger, D. D., Vose, J. M., Greiner, T. C., Lynch, J. C., Chan, W. C., Bierman, P. J., Dave, B. J., Sanger, W. G., Armitage, J. O. (2000). Mantle cell lymphoma. A clinicopathologic study of 68 cases from the Nebraska Lymphoma Study Group. *Am J Hematol* 64(3): 190-6.

- [20] Zucca, E., Roggero, E., Pinotti, G., Pedrinis, E., Cappella, C., Venco, A., Cavalli, F. (1995). Patterns of survival in mantle cell lymphoma. *Ann Oncol* 6(3): 257-62.
- [21] Bosch, F., Lopez-Guillermo, A., Campo, E., Ribera, J. M., Conde, E., Piris, M. A., Vallespi, T., Woessner, S., Montserrat, E. (1998). Mantle cell lymphoma: presenting features, response to therapy, and prognostic factors. *Cancer* 82(3): 567-75.
- [22] Decaudin, D., Bosq, J., Munck, J. N., Bayle, C., Koscielny, S., Boudjemaa, S., Benaceur, A., Venuat, A. M., Naccache, P., Bendahmane, B., Ribrag, V., Carde, P., Pico, J. L., Hayat, M. (1997). Mantle cell lymphomas: characteristics, natural history and prognostic factors of 45 cases. *Leuk Lymphoma* 26(5-6): 539-50.
- [23] Carbone, P. P., Kaplan, H. S., Musshoff, K., Smithers, D. W., Tubiana, M. (1971). Report of the Committee on Hodgkin's Disease Staging Classification. *Cancer Res* 31(11): 1860-1.
- [24] Hiddemann W., Dreyling M., Stein, H. (Hrsg., 2005) *Lymphome*. Thieme Verlag.
- [25] Musshoff, K., Schmidt-Vollmer, H. (1975). Proceedings: Prognosis of non-Hodgkin's lymphomas with special emphasis on the staging classification. *Z Krebsforsch Klin Onkol Cancer Res Clin Oncol* 83(4): 323-41.
- [26] Musshoff, K. (1977). Clinical staging classification of non-Hodgkin's lymphomas (author's transl). *Strahlentherapie* 153(4): 218-21.
- [27] Armitage, J. O. (2005). Staging non-Hodgkin lymphoma. *CA Cancer J Clin* 55(6): 368-76.
- [28] Armitage, J. O. (2005). Defining the stages of aggressive non-Hodgkin's lymphoma—a work in progress. *N Engl J Med* 352(12): 1250-2.
- [29] (1993). A predictive model for aggressive non-Hodgkin's lymphoma. The International Non-Hodgkin's Lymphoma Prognostic Factors Project. *N Engl J Med* 329(14): 987-94.
- [30] Solal-Celigny, P., Roy, P., Colombat, P., White, J., Armitage, J. O., Arranz-Saez, R., Au, W. Y., Bellei, M., Brice, P., Caballero, D., Coiffier, B., Conde-Garcia, E., Doyen, C., Federico, M., Fisher, R. I., Garcia-Conde, J. F., Guglielmi, C., Hagenbeek, A., Haioun, C., LeBlanc, M., Lister, A. T., Lopez-Guillermo, A., McLaughlin, P., Milpied, N., Morel, P., Mounier, N., Proctor, S. J., Rohatiner, A., Smith, P., Soubeyran, P., Tilly, H., Vitolo, U., Zinzani, P. L., Zucca, E., Montserrat, E. (2004). Follicular lymphoma international prognostic index. *Blood* 104(5): 1258-65.

- [31] Concato, J., Peduzzi, P., Holford, T. R., Feinstein, A. R. (1995). Importance of events per independent variable in proportional hazards analysis. I. Background, goals, and general strategy. *J Clin Epidemiol* 48(12): 1495-501.
- [32] Peduzzi, P., Concato, J., Feinstein, A. R., Holford, T. R. (1995). Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. *J Clin Epidemiol* 48(12): 1503-10.
- [33] Nickenig, C., Dreyling, M., Hoster, E., Pfreundschuh, M., Trümper, L., Reiser, M., Wandt, H., Lengfelder, E., Unterhalt, M., Hiddemann, W. (2006). Combined cyclophosphamide, vincristine, doxorubicin, and prednisone (CHOP) improves response rates but not survival and has lower hematologic toxicity compared with combined mitoxantrone, chlorambucil, and prednisone (MCP) in follicular and mantle cell lymphomas: results of a prospective randomized trial of the German Low-Grade Lymphoma Study Group. *Cancer* 107(5): 1014-22.
- [34] Lenz, G., Dreyling, M., Hoster, E., Wörmann, B., Dührsen, U., Metzner, B., Eimermacher, H., Neubauer, A., Wandt, H., Steinhauer, H., Martin, S., Heidemann, E., Aldaoud, A., Parwaresch, R., Hasford, J., Unterhalt, M., Hiddemann, W. (2005). Immunotherapy with rituximab and cyclophosphamide, doxorubicin, vincristine, and prednisone significantly improves response and time to treatment failure, but not long-term outcome in patients with previously untreated mantle cell lymphoma: results of a prospective randomized trial of the German Low Grade Lymphoma Study Group (GLSG). *J Clin Oncol* 23(9): 1984-92.
- [35] Hiddemann, W., Kneba, M., Dreyling, M., Schmitz, N., Lengfelder, E., Schmits, R., Reiser, M., Metzner, B., Harder, H., Hegewisch-Becker, S., Fischer, T., Kropff, M., Reis, H. E., Freund, M., Wörmann, B., Fuchs, R., Planker, M., Schimke, J., Eimermacher, H., Trümper, L., Aldaoud, A., Parwaresch, R., Unterhalt, M. (2005). Frontline therapy with rituximab added to the combination of cyclophosphamide, doxorubicin, vincristine, and prednisone (CHOP) significantly improves the outcome for patients with advanced-stage follicular lymphoma compared with therapy with CHOP alone: results of a prospective randomized study of the German Low-Grade Lymphoma Study Group. *Blood* 106(12): 3725-32.
- [36] Lenz, G., Dreyling, M., Schiegnitz, E., Forstpointner, R., Wandt, H., Freund, M., Hess, G., Trümper, L., Diehl, V., Kropff, M., Kneba, M., Schmitz, N., Metzner, B., Pfirrmann, M., Unterhalt, M., Hiddemann, W. (2004). Myeloablative radiochemotherapy followed by autologous stem cell transplantation in first remission prolongs progression-free survival in follicular lymphoma: results of a prospective, randomized trial of the German Low-Grade Lymphoma Study Group. *Blood* 104(9): 2667-74.

- [37] Dreyling, M., Lenz, G., Hoster, E., Van Hoof, A., Gisselbrecht, C., Schmits, R., Metzner, B., Trümper, L., Reiser, M., Steinbauer, H., Boiron, J. M., Boogaerts, M. A., Aldaoud, A., Silingardi, V., Kluin-Nelemans, H. C., Hasford, J., Parwaresch, R., Unterhalt, M., Hiddemann, W. (2005). Early consolidation by myeloablative radiochemotherapy followed by autologous stem cell transplantation in first remission significantly prolongs progression-free survival in mantle-cell lymphoma: results of a prospective randomized trial of the European MCL Network. *Blood* 105(7): 2677-84.
- [38] Cheson, B. D., Horning, S. J., Coiffier, B., Shipp, M. A., Fisher, R. I., Connors, J. M., Lister, T. A., Vose, J., Grillo-Lopez, A., Hagenbeek, A., Cabanillas, F., Klippel, D., Hiddemann, W., Castellino, R., Harris, N. L., Armitage, J. O., Carter, W., Hoppe, R., Canellos, G. P. (1999). Report of an international workshop to standardize response criteria for non-Hodgkin's lymphomas. NCI Sponsored International Working Group. *J Clin Oncol* 17(4): 1244.
- [39] Whitehead, J. (1983). *The design and analysis of sequential clinical trials*. Chichester, Ellis Horwood Ltd.
- [40] Oken, M. M., Creech, R. H., Tormey, D. C., Horton, J., Davis, T. E., McFadden, E. T., Carbone, P. P. (1982). Toxicity and response criteria of the Eastern Cooperative Oncology Group. *Am J Clin Oncol* 5(6): 649-55.
- [41] Korn, E. L. (1986). Censoring distributions as a measure of follow-up in survival analysis. *Stat Med* 5(3): 255-60.
- [42] Altman, D. G., Royston, P. (2006). The cost of dichotomising continuous variables. *BMJ* 332(7549): 1080.
- [43] Lausen B., Schumacher M. (1992). Maximally Selected Rank Statistics. *Biometrics* 48: 73-85.
- [44] Royston, P., Altman, D. G., Sauerbrei, W. (2006). Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med* 25(1): 127-41.
- [45] Altman, D. G., Lausen, B., Sauerbrei, W., Schumacher, M. (1994). Dangers of using „optimal“ cutpoints in the evaluation of prognostic factors. *J Natl Cancer Inst* 86(11): 829-35.
- [46] Hasford, J., Pffirmann, M., Hehlmann, R., Allan, N. C., Baccarani, M., Kluin-Nelemans, J. C., Alimena, G., Steegmann, J. L., Ansari, H. (1998). A new prognostic score for survival of patients with chronic myeloid leukemia treated with interferon alfa. Writing Committee for the Collaborative CML Prognostic Factors Project Group. *J Natl Cancer Inst* 90(11): 850-8.

- [47] Cohen, J. (1968). Weighted Kappa: Nominal Scale Agreement with Provision for Scaled Disagreement or Partial Credit. *Psychological Bulletin* 70(4): 213-220.
- [48] Graf, E., Schmoor, C., Sauerbrei, W., Schumacher, M. (1999). Assessment and comparison of prognostic classification schemes for survival data. *Stat Med* 18(17-18): 2529-45.
- [49] Efron, B. (1983). Estimating the error rate of a prediction rule: improvement on cross-validation. *J. Amer. Statist. Assoc.* 78(382): 316-331.
- [50] Harrell, F. E., Jr., Califf, R. M., Pryor, D. B., Lee, K. L., Rosati, R. A. (1982). Evaluating the yield of medical tests. *Jama* 247(18): 2543-6.
- [51] Harrell, F. E., Jr., Lee, K. L., Mark, D. B. (1996). Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 15(4): 361-87.
- [52] Hanley, J. A., McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143(1): 29-36.
- [53] Sauerbrei, W., Hubner, K., Schmoor, C., Schumacher, M. (1997). Validation of existing and development of new prognostic classification schemes in node negative breast cancer. German Breast Cancer Study Group. *Breast Cancer Res Treat* 42(2): 149-63.
- [54] Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review* 78: 1-3.
- [55] Steyerberg, E. W., Harrell, F. E., Jr., Borsboom, G. J., Eijkemans, M. J., Vergouwe, Y., Habbema, J. D. (2001). Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* 54(8): 774-81.
- [56] Efron, B., Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York, Chapman and Hall.
- [57] Efron, B., Tibshirani, R. J. (1997). Improvements on cross-validation: the .632+ bootstrap method. *J. Amer. Statist. Assoc.* 92(438): 548-560.
- [58] Altman, D. G., Royston, P. (2000). What do we mean by validating a prognostic model? *Stat Med* 19(4): 453-73.
- [59] Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model. *Biometrika* 69(1): 239-241.
- [60] Grambsch, P. M., Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 81(3): 515-526.

- [61] Therneau, T. M., Grambsch, P. M., Fleming, Th. R. (1990). Martingale-based residuals for survival models. *Biometrika* 77(1): 147-160.
- [62] Altman, D. G. (2001). Systematic reviews of evaluations of prognostic variables. *BMJ* 323(7306): 224-8.
- [63] Coiffier, B., Lepage, E., Briere, J., Herbrecht, R., Tilly, H., Bouabdallah, R., Morel, P., Van Den Neste, E., Salles, G., Gaulard, P., Reyes, F., Lederlin, P., Gisselbrecht, C. (2002). CHOP chemotherapy plus rituximab compared with CHOP alone in elderly patients with diffuse large-B-cell lymphoma. *N Engl J Med* 346(4): 235-42.
- [64] Hoster, E., Dreyling, M., Klapper, W., Gisselbrecht, C., van Hoof, A., Kluin-Nelemans, H. C., Pfreundschuh, M., Reiser, M., Metzner, B., Einsele, H., Peter, N., Jung, W., Wörmann, B., Ludwig, W. D., Dührsen, U., Eimermacher, H., Wandt, H., Hasford, J., Hiddemann, W., Unterhalt, M. (2008). A new prognostic index (MIPI) for patients with advanced-stage mantle cell lymphoma. *Blood* 111(2): 558-65.
- [65] Hoster, E., Unterhalt, M., Dreyling, Hasford, J., Hiddemann, W. (2007). Ein neuer prognostischer Index (MIPI) für Patienten mit fortgeschrittenem Mantelzell-Lymphom. *Dtsch. Med. Wochenschr.* 132, Nr. 31/32: 1655.