

Aus dem Institut für Urologie  
des Universitätsklinikums Eppendorf  
Institutsdirektor: Prof. Hartwig Hulan

**Validierung bestehender Erstbiopsie-Nomogramme zur Vorhersage  
eines Prostatakarzinoms**

**Dissertation**

zur Erlangung des Grades eines Doktors der Medizin  
dem Fachbereich Medizin der Universität Hamburg vorgelegt von

Julia Marie Hopp  
aus Buchholz in der Nordheide

Hamburg 2008

Angenommen vom Fachbereich Medizin  
der Universität Hamburg am: 04.11.2008

Veröffentlichung mit Genehmigung des Fachbereiches  
Medizin der Universität Hamburg

Prüfungsausschuss, der Vorsitzende: Prof. Dr. M. Graefen

Prüfungsausschuss: 2. Gutachter: Prof Dr. H. Heinzer

Prüfungsausschuss: 3. Gutachter: Prof. Dr. G. Sauter

Stellv. Mitglied: Prof. Dr. H. Huland

# INHALT

<b>1</b>	<b>EINLEITUNG</b>	<b>1</b>
1.1	EPIDEMIOLOGIE DES PROSTATAKARZINOMS	1
1.2	RISIKOFAKTOREN DES PROSTATAKARZINOMS	2
1.2.1	<i>Alter</i>	2
1.2.2	<i>DRU (Digitale rektale Untersuchung)</i>	3
1.2.3	<i>TRUS (Transrektaler Ultraschall)</i>	3
1.2.4	<i>Serummarker</i>	4
1.3	HISTOLOGIE DES PROSTATAKARZINOMS	7
1.4	KLINISCH-PATHOLOGISCHE EINTEILUNG DES PROSTATAKARZINOMS	9
1.5	THERAPEUTISCHE MÖGLICHKEITEN DES PROSTATAKARZINOMS	11
1.6	RISIKOVORHERSAGEMODELLE	16
1.6.1	<i>Kriterien</i>	16
1.6.2	<i>Limitationen der Risikovorhersagemodelle</i>	19
1.6.3	<i>Nomogramme</i>	20
1.6.4	<i>CART (Classification and Regression Tree)</i>	22
1.6.5	<i>Artifizielle neuronale Netzwerke (ANN)</i>	23
1.7	ARBEITSHYPOTHESE	25
<b>2</b>	<b>METHODEN</b>	<b>26</b>
2.1	ANALYSIERTE MODELLE	26
2.1.1	<i>Nomogramme</i>	26
2.1.2	<i>CART-Analyse (Classification and Regression Tree)</i>	26
2.1.3	<i>Artifzielles neuronales Netzwerk (ANN)</i>	27
2.2	PATIENTENPOPULATION	27
2.3	HISTOLOGISCHE UND KLINISCHE DATEN	28
2.4	STATISTISCHE ANALYSEN	28

2.4.1	<i>Regressionsanalyse: Externe Validierung</i>	28
2.4.2	<i>Kalibrierungskurven</i>	29
<b>3</b>	<b>ERGEBNISSE</b>	<b>30</b>
3.1	<b>DESKRIPTIVE ANALYSEN</b>	<b>30</b>
3.2	<b>DIREKTVERGLEICH DER EINZELNEN ERSTBIOPSIE- VORHERSAGEMODELLE</b>	<b>34</b>
<b>4</b>	<b>DISKUSSION</b>	<b>38</b>
<b>5</b>	<b>ZUSAMMENFASSUNG</b>	<b>47</b>
<b>6</b>	<b>LITERATURVERZEICHNIS</b>	<b>48</b>
<b>7</b>	<b>DANKSAGUNG</b>	<b>55</b>
<b>8</b>	<b>LEBENS LAUF</b>	<b>56</b>
<b>9</b>	<b>ERKLÄRUNG</b>	<b>57</b>

## Verzeichnis der Abbildungen

Abbildung 1: Molekularstruktur des PSA Moleküls aus Nurmikko et al. (2000) Clin. Chem. 46 (Nurmikko, Vaisanen et al. 2000)	5
Abbildung 2: Differenzierungsgrad nach Gleason	8
Abbildung 3: Gesamt-PSA freies Überleben sowie stratifiziert nach klinischen Charkateristika nach radikaler retropubischer Prostatektomie aus Chun et al. WJU 2006 (Chun, Graefen et al. 2006)	13
Abbildung 4: Prädiktive Genauigkeit unterschiedlicher PCa-Nomogramme aus Chun et al. WJU 2007 in press	17
Abbildung 5: Schematische Darstellung eines Nomogramms unter Verwendung prä-operativer Risikofaktoren zur Vorhersage eines Prostatakarzinoms in der Erstbiopsie aus Karakiewicz et al. JUrol 2005 (Karakiewicz, Benayoun et al. 2005) mit Kalibrierung	21
Abbildung 6: Schematische Darstellung einer CART-Analyse unter Verwendung prä-operativer Risikofaktoren zur Vorhersage eines extraprostatatischen PCa-Wachstums im endgültigen pathologischen Präparat mit Kalibrierung. Aus Karakiewicz et al: CanJUrol 2006 (Karakiewicz, Chun et al. 2006)	23
Abbildung 7: Schematische Darstellung eines artifiziellen neuronalen Netzwerks	24
Abbildung 8: Kalibrierungskurven der externen Validierung der etablierten Prostataerstbiopsie Vorhersagemodelle.	37

## **Verzeichnis der Tabellen**

Tabelle 1: TNM-Klassifikation	11
Tabelle 2: Grundlegende Charakteristika der 3980 Patienten, die einer Prostata- Erstbiopsie unterzogen wurden und die Studienkohorte bilden	31
Tabelle 3: Vergleich der Ergebnisse von externer Überprüfung und ursprünglich angegebenen Werten der Vorhersagemodelle	33
Tabelle 4: Direktvergleich der prädiktiven Genauigkeiten der einzelnen Erstbiopsie- Modelle mittels des Mantel-Haenszel-Test	35
Tabelle 5: Prädiktive Genauigkeiten der einzelnen Erstbiopsievorhersage Modelle in der externen Validierung (n=3980).	40

## Benutzte Abkürzungen und Fremdwörter

Abb.	Abbildung
AJCC	American Joint Committee on Cancer
ACT	Antichymotrypsin
ANN	Artifiziell neuronales Netzwerk
AUC	Area under the curve
Aufl.	Auflage
BPH	Benigne Prostatahyperplasie
bzw.	beziehungsweise
CART	Classification and Regression Tree
ca.	circa (ungefähr)
cm <sup>3</sup>	Kubik-Zentimeter
cPSA	Komplex-gebundenes PSA
cTNM	klinische Stadieneinteilung nach dem TNM-System
d.h.	das heißt
DRU	Digital rektale Untersuchung
et al.	et alii (und andere)
etc.	et cetera (und so weiter)
fPSA	freie Form des Prostata spezifischen Antigens
fPSA-I	intact PSA
fPSA-N	nicked PSA

%fPSA	Anteil an freiem PSA in Prozent
GnRH	Gonadotropin Releasing Hormone
Gy	Gray
HDR	high-dosis-rate
hK2	humanes Kalikrein 2
LDR	low-dosis-rate
ml	Milliliter
µl	Mikroliter
n	Anzahl
ng	Nanogramm
PCa	Prostatakarzinom
PSA	Prostata-spezifisches Antigen
pTNM	pathologische Stadieneinteilung nach dem TNM-System
ROC	Receiver operating characteristics Analyse
TGF	Tumor growth factor
tPSA	Total- oder Gesamt-PSA
TNM	Tumor, Lymphknoten, (Fern-)Metastasen
TRUS	Transrektaler Ultraschall

WHO

World Health Organisation

# **1 Einleitung**

## **1.1 Epidemiologie des Prostatakarzinoms**

Das Prostatakarzinom (PCa) ist der häufigste urologische Tumor und rangiert nach dem Bronchialkarzinom an zweiter Stelle bei den zum Tode führenden Krebserkrankungen des Mannes (Quelle: Statistisches Bundesamt, 2006). Nach den Angaben des Statistischen Bundesamtes verstarben 2006 in Deutschland 11 577 Männer an den Folgen einer bösartigen Neubildung der Prostata. Gleichzeitig liegt das PCa mit 3,0% aller Todesfälle an sechster Stelle aller Todesursachen bei Männern. Diese epidemiologische Entwicklung ist keine rein europäische. Auch in den USA ist das PCa der am häufigsten diagnostizierte Tumor bei Männern, die Inzidenz lag im Jahre 2004 bei 150 Neuerkrankungen je 100 000 Männer (2005; 2005, 28.02.; Ärztezeitung 2005, 28.02.). Als Folge dieser Entwicklung ist anzunehmen, dass das PCa aufgrund der steigenden Lebenserwartung in Zukunft einen noch stärkeren Einfluß auf die Gesundheit der männlichen Bevölkerung haben wird.

Seit Ende der 80er Jahre ist in Deutschland ein steiler Anstieg der Neuerkrankungsraten (Inzidenz) zu verzeichnen. Dieser Anstieg ist größtenteils auf den Einsatz neuer diagnostischer Methoden wie die Einführung des Prostata spezifischen Antigens (PSA), einem Serummarker, zurückzuführen (Noldus, Graefen et al. 2000). PCa werden durch den Einsatz der regelmäßigen Serumdiagnostik häufiger in prognostisch günstigeren, potentiell heilbaren Frühstadien entdeckt. Dies sind Tumore, die auf die Prostata beschränkt sind (Chun, Graefen et al. 2006; Graefen, Walz et al. 2006). Daneben ist auch zu erwähnen, dass, gemäß Untersuchungen der WHO, mit Einführung des PSA als Serummarker, die Mortalitätsraten für das PCa in den letzten zehn Jahren um 3,6% gesunken sind (Labrie, Candas et al. 2004). Es hat sich jedoch in Autopsiestudien gezeigt, dass die Prävalenz des PCa erheblich höher angesiedelt ist als die Inzidenz - die Rate der jährlich registrierten Neuerkrankungen. So

genannte „Autopsiekarzinome“ konnten einer Studie zufolge bei 42% aller Männer in der 5. Lebensdekade und 64% aller Männer in der 7. Lebensdekade festgestellt werden (Sakr, Grignon et al. 1994). Obwohl die Prävalenz (Anteil der Bevölkerung, der zu einer bestimmten Zeiteinheit erkrankt ist oder es jemals war) des PCa sehr hoch ist, entwickelt nur ein geringer Anteil der Männer (9%) ein „signifikantes“, also klinisch relevantes PCa. Und von diesen 9% versterben lediglich 3% an den direkten Folgen ihres Tumors (Sakr, Grignon et al. 1994). Während für die hohen Prävalenzwerte des PCa geographische Übereinstimmungen festgestellt werden konnten, unterscheiden sich die Inzidenzraten des Karzinoms je nach geographischer Lage deutlich. So wurden Unterschiede zwischen 2 Neuerkrankungen/100.000/Jahr in China und 140 Neuerkrankungen/100.000/Jahr in den USA beobachtet (Postma and Schroder 2005). Verantwortlich für diese Diskrepanzen sind epidemiologischen Studien zufolge vor allem unterschiedliche Ernährungsgewohnheiten, welche den Testosteronspiegel beeinflussen, Umweltfaktoren und allgemeine ethnische Einflüsse. Zu bedenken ist außerdem die Tatsache, dass in weiten Teilen der USA bei Männern PSA – Screeningprogramme durchgeführt werden und daraus logischerweise höhere Neuentdeckungsraten folgen. Dagegen existiert in Europa kein PSA-Screening (Jemal, Tiwari et al. 2004). Gleichzeitig besteht auch in Deutschland der Anspruch, das Auftreten eines PCa möglichst früh zu entdecken, da dann die Heilungschancen für den Patienten am besten sind (Zincke, Oesterling et al. 1994; Pound, Partin et al. 1997; Chun, Graefen et al. 2006).

## **1.2 Risikofaktoren des Prostatakarzinoms**

### **1.2.1 Alter**

Das Alter ist einer der wichtigsten Risikofaktoren für die Ausbildung eines PCa. Vor dem 40. Lebensjahr tritt es nur in Ausnahmefällen auf, doch mit jeder weiteren Lebensdekade ist ein

logarithmischer Anstieg seiner Prävalenz zu beobachten, dessen Maximum zwischen dem 70. und 80. Lebensjahr liegt.

### **1.2.2 DRU (Digitale rektale Untersuchung)**

Die digitale rektale Untersuchung stellt wohl die simpelste Form der Prostatauntersuchung dar. Sie hilft dem Untersucher, einen ersten Eindruck von der Größe und der Beschaffenheit der Prostata zu gewinnen. Oberflächen- und Konsistenzveränderungen der Prostata lassen hierbei das Vorhandensein eines Karzinoms vermuten. Bevor der PSA-Test in die Klinik eingeführt wurde, war die DRU die einzige Untersuchung zur Früherkennung des PCas und konnte in maximal 50% ein Karzinom nachweisen, hier vor allem bereits lokal-fortgeschrittene Stadien (Jewett 1956). Der positive prädiktive Wert einer DRU schwankt sehr und hängt von verschiedenen Faktoren ab wie z.B. der Erfahrung auf Seiten des Untersuchers oder der untersuchten Population (z.B. Zuweiserpopulation vs Screeningpopulation). Die Methode versagt bei Karzinomen, die sich in Regionen befinden, die digital nicht zu erreichen sind und ist auch nur limitiert reproduzierbar. So scheint diese Form der Untersuchung allein eher eine grobe Orientierung zu bieten und muss dringend durch andere Parameter wie die Bestimmung des PSA-Wertes und eine auffällige Anamnese ergänzt werden; hierdurch stellt sie dann aber die Grundlage der Früherkennung des PCa dar (Catalona, Richie et al. 1994). Bei auffälligen Befunden dieser Parameter findet eine transrektale ultraschall-gesteuerte Stanzbiopsie der Prostata statt, die alleinig die Diagnose eines PCa ausschließen oder sichern kann.

### **1.2.3 TRUS (Transrektaler Ultraschall)**

Der transrektale Ultraschall bildet zusammen mit der DRU und der Bestimmung des PSA-Wertes die Basis in der Früherkennung eines PCa. Der TRUS ist eine weit verbreitete Methode, um nähere Informationen zur lokalen Ausdehnung der Prostata bzw. eines

vorhandenen PCa zu bekommen. Als typischer Befund der Sonographie bei Vorhandensein eines PCa wird dabei eine hypoechogene Raumforderung in der peripheren Zone der Prostata angesehen. Außerdem kann der Ultraschall zum Beispiel bei der Klärung der Frage helfen, ob bei vorliegendem Karzinom bereits eine Kapselinfiltration stattgefunden hat. In der Sensitivität scheint der TRUS der alleinigen DRU überlegen zu sein, kombiniert man jedoch die DRU mit der PSA-Bestimmung, weist der TRUS eine niedrigere Sensitivität auf. Mittels dieser Kombination aus PSA-Wert und DRU werden dann bis zu 40% aller Tumore detektiert, die durch den TRUS nicht erkannt werden (Shinohara 1989). Zusammenfassend ist der TRUS als Untersuchungsmethode oft ungenau und es werden teilweise bis zu 50% Kapselinfiltration sowie kleine Karzinomherde ( $<0,5\text{cm}^3$ ) übersehen. Es ist auch möglich, dass anderweitige Veränderungen der Prostata als Karzinombefall interpretiert werden, deren Ursache hingegen in einer Infektion oder ähnlichem besteht.

Nichtsdestotrotz ist die Ultraschalluntersuchung eine preiswerte, wenig aufwendige und den Patienten minimal belastende und zu jedem Zeitpunkt wiederholbare Untersuchung, die jedoch nicht als Früherkennungsmethode der ersten Wahl anzusehen ist, sondern in Kombination mit DRU und PSA-Bestimmung zu erfolgen hat. Ein weiterer Anwendungsbereich des TRUS ist im Verlauf der Diagnostik die Steuerung der Gewebeentnahme aus der Prostata bei der Stanzbiopsie (Kuefer 2004).

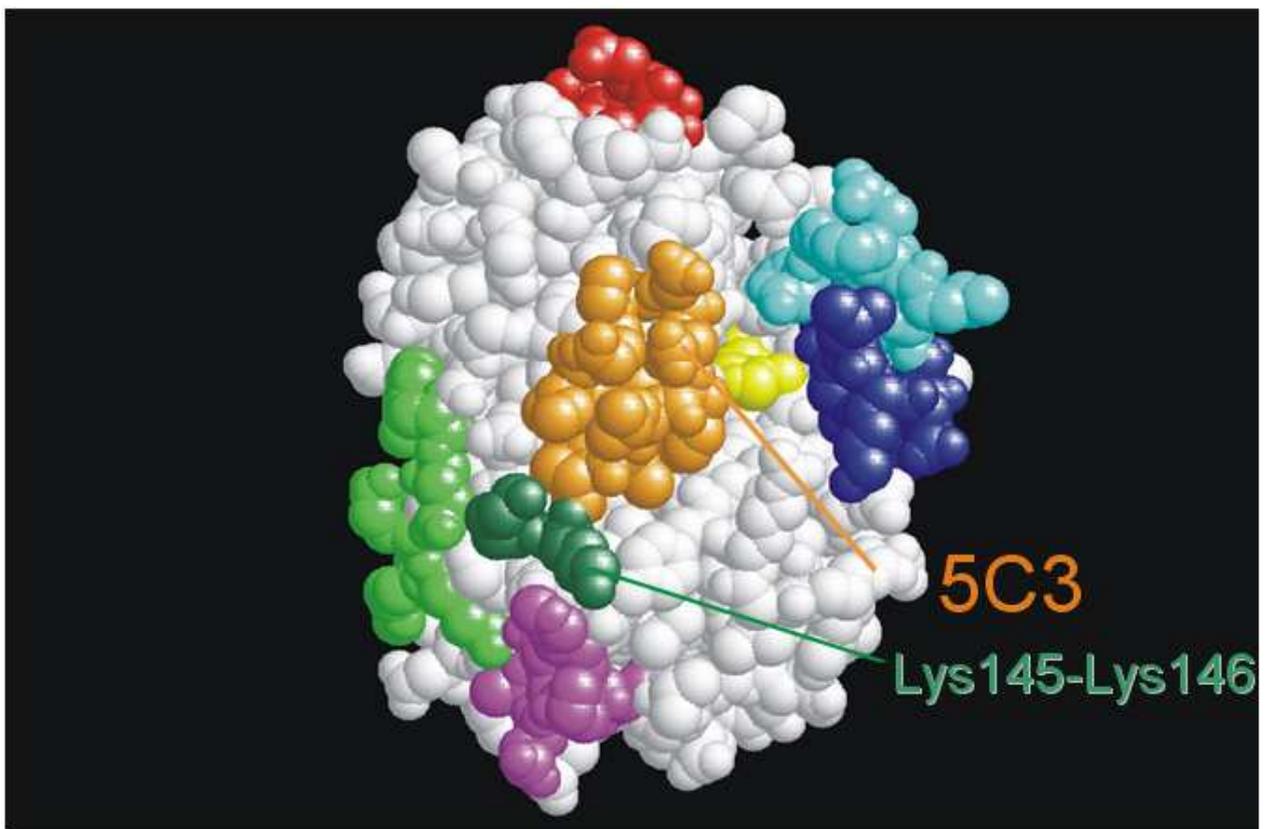
#### **1.2.4 Serummarker**

##### **Prostata-spezifisches Antigen (PSA) und seine Subformen**

PSA ist ein Glykoprotein, welches aus 237 Aminosäuren besteht. Beim Mann wird es zum größten Teil in den epithelialen Zellen der Prostata-drüsen gebildet. PSA besitzt die enzymatischen Eigenschaften einer neutralen Serinprotease. Die Hauptfunktion des PSA besteht in der Spaltung von Proteinen der Samenblase zur Verflüssigung des Sekrets. 1979 gelang es Forschern erstmals PSA im Prostatagewebe nachzuweisen, etwa 10 Jahre später

wurde es dann in die Klinik eingeführt (Wang, Valenzuela et al. 1979; Lilja 2000) und löste die saure Phosphatase als bisherigen Marker des PCa ab.

Ein Teil des PSA gelangt aus den Drüsengängen der Prostata heraus in die Blutbahn, wo es zum größten Teil (ca. 80%) an zwei Proteine im Komplex (cPSA) gebunden ist: 1. Alpha-1-Antichymotrypsin (ACT) und 2. Alpha-2-Makroglobulin. Der überwiegende Anteil des cPSA ist hierbei an ACT gebunden und macht ca. 70% aus. Etwa 20% des PSA im Blut können in der Regel als freie Form (fPSA) gemessen werden. Beide Formen zusammen, cPSA und fPSA, ergeben dann das Gesamt- oder Total-PSA (tPSA).



**Abbildung 1: Molekularstruktur des PSA Moleküls aus Nurmikko et al. (2000) Clin. Chem. 46 (Nurmikko, Vaisanen et al. 2000)**

Stenman et al. zeigten 1991 als erste, dass Männer mit einem PCa einen größeren Anteil an ACT gebundenem cPSA aufwiesen als Männer mit einer PSA-Erhöhung aufgrund einer benignen Prostatahyperplasie (Stenman, Leinonen et al. 1991). Der Anteil an freiem PSA ist

bei ihnen also niedriger. Das fPSA ist enzymatisch inaktiv und kann seinerseits nochmals in molekulare Unterformen unterteilt werden.

1.) *fPSA-N (nicked PSA)*: stellt Fraktionen dar, die mittels interner proteolytischer Spaltung der Molekülkette enzymatisch inaktiviert werden; fPSA-N scheint mit dem Vorliegen einer Benigen Prostatahyperplasie (BPH) assoziiert zu sein (Mikolajczyk, Millar et al. 2000).

2.) *fPSA-I (intact PSA)* : Subformen mit intakter Molekülkette, zum einen Proformen des PSA (proPSA) und reifes, inaktives PSA (Noldus, Chen et al. 1997). Nur intakte Isoformen konnten in PCa-Zellen detektiert werden und gelten als tumorspezifisch (Vaisanen, Lovgren et al. 1999).

Mittels des Quotienten  $fPSA / tPSA$  kann der prozentuale Anteil des freien PSA bestimmt werden (%fPSA), der seinerseits neben dem Gesamt-PSA mittlerweile als akzeptierter Serummarker gilt. Je kleiner der Quotient ausfällt, desto wahrscheinlicher liegt ein malignes Geschehen vor.

Der normale unauffällige tPSA-Wert liegt zwischen 0,0 und 4,0 ng/ml (Hudson, Bahnson et al. 1989). Ausgehend von internationalen Studien hat sich in Europa dementsprechend ein Cut-off-Wert von 4 ng/ml für das tPSA im Blut von Patienten etabliert (Kleer and Oesterling 1993). Ab diesem Wert wird der Patient zur möglichst frühzeitigen Entdeckung eines PCa einer Prostatastanzbiopsie unterzogen. Es muss jedoch beachtet werden, dass ca. 30% der PCa-Patienten einen tPSA-Wert  $< 4$  ng/ml aufweisen und somit trügerischerweise als gesund beurteilt werden. Auch falschpositive Werte werden bei bis zu 75% der Patienten ermittelt, die in der Folge dann einer unnötigen Biopsie unterzogen werden. Sollte der tPSA-Wert eines Patienten höher bestimmt werden, muss dies eben nicht zwingend auf ein PCa hinweisen. Das Serum-PSA steigt mit zunehmendem Alter des Patienten unabhängig von einem Malignom. Dies ist mit der gleichzeitig stattfindenden Größenzunahme der Prostata zu erklären, denn es

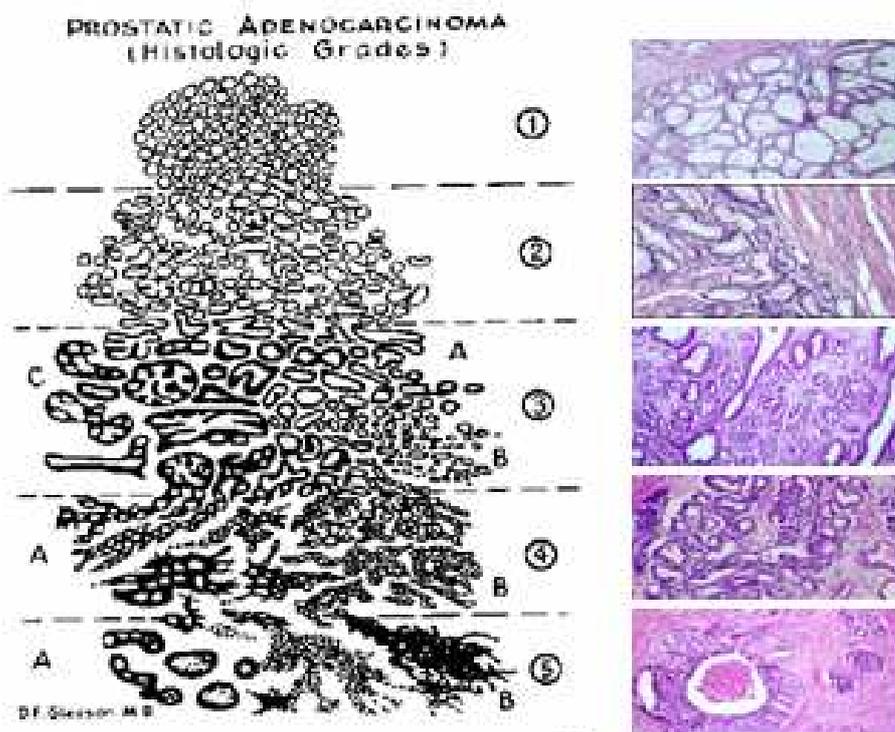
besteht eine direkte Korrelation zwischen der Prostatagröße und dem Serum-PSA (Chun, Perrotte et al. 2006).

Leichte Erhöhungen des PSA-Wertes sind also keinesfalls krankheitsspezifisch, sondern können vielmehr Ausdruck einer benignen, malignen oder auch inflammatorischen Prostataerkrankung sein. Auch mechanische Manipulationen wie die DRU, eine ultraschallgesteuerte Prostatastanzbiopsie oder eine Cystoskopie führen zu einer PSA-Erhöhung. Besonders zwischen dem PCa und der BPH gibt es hierbei große Überschneidungsbereiche. Werte zwischen 4,1 und 10 ng/ml sind zwar mit einer benignen Prostatahyperplasie zu vereinbaren, schließen das PCa jedoch keinesfalls aus (Malignität in 20-30 % der Fälle). PSA ist somit kein krankheitsspezifischer bzw. karzinomspezifischer, sondern vielmehr ein organ-spezifischer Serummarker. Wenn man aber die Einschränkungen in der Diagnostik des PCa anhand des tPSA-Wertes beachtet, muss man abschließend sagen, dass das tPSA derzeit der beste verfügbare Tumormarker für das PCa darstellt (Kirby 2001) und das %fPSA in Zukunft wohl noch an Bedeutung innerhalb der Prostatadiagnostik gewinnen wird.

### **1.3 Histologie des Prostatakarzinoms**

Der weitaus größte Anteil der Karzinome der Prostata, nämlich 97%, sind Adenokarzinome. Plattenepithelkarzinome, endometroide, kleinzellige oder verschleimende Karzinome sind Raritäten. Bevorzugte Lokalisation des PCa sind mit bis zu 80% die peripheren Zonen, 20% entstehen daneben in der Transitionalzone der Prostata. Zur Beurteilung von Adenokarzinomen der Prostata bewährt sich seit langem das Gleason-System, welches nach dem amerikanischen Arzt Dr. Donald Gleason benannt ist. Gleason hatte Mitte der 70er Jahre durch mikroskopische Beobachtung an zahlreichen Gewebeproben herausgefunden, dass das Auftreten unterschiedlicher PCa-Zellen mit unterschiedlicher Aggressivität der Erkrankung

vergesellschaftet ist (Gleason and Mellinger 1974). Der Gleason-Score (Abbildung 1) unterscheidet fünf verschiedene histologische Differenzierungs- und Wachstumsmuster. Je höher der Wert (zwischen 1-5) ist, desto undifferenzierter sind die Zellen und desto ungünstiger ist damit die Prognose des Patienten. Da erwiesenermaßen mehr als 85% der Tumoren multifokal wachsen (McNeal 1969; Byar and Mostofi 1972; McNeal, Redwine et al. 1988; Epstein, Walsh et al. 1994), werden hierbei die zwei größten Tumorareale mittels Gleason-System beurteilt und zum Gleason-Score addiert. Die Summe kann dabei Werte zwischen 2 (1+1) und 10 (5+5) annehmen. Zu bedenken ist, dass z.B. eine Summe 5, die sich aus den Ziffern 3+2 zusammensetzt ungünstiger für den Patienten ist, als eine Kombination der Ziffern 2+3, da im ersten Fall der Anteil an undifferenzierten und damit aggressiveren Zellen höher liegt.



**Abbildung 2: Differenzierungsgrad nach Gleason**

Zusammenfassend kann man sagen, dass der Gleason Grad zur histologischen Beschreibung der Tumorcharakteristika (Heterogenität, Ausdehnung...) herangezogen werden kann, ein „Undergrading“ des Gleason Grades der Biopsie im Vergleich mit der endgültigen Pathologie

jedoch häufig zu beobachten ist und bedacht werden muss (Bostwick 1994; King and Long 2000; Cam, Yucel et al. 2002; Chun, Briganti et al. 2006; Chun, Karakiewicz et al. 2006). Umfangreichere Biopsieschemata ( $\geq 10$  Biopsien) scheinen den Umstand des „Undergradings“ signifikant zu verbessern (King, Patel et al. 2005). Letztlich ist das Gleason System vor allem im Bereich der prognostischen Einschätzung des Krebses allgemein akzeptiert und verbreitet.

#### **1.4 Klinisch-pathologische Einteilung des Prostatakarzinoms**

Die Stadieneinteilung des PCa (Staging) entspricht der TNM-Klassifikation von 2002 (Tabelle 1). Im Rahmen dieser Klassifikation werden die lokale Ausdehnung des Primärtumors (T-Stadium), der regionäre Lymphknotenbefall (N-Stadium) und das Vorliegen von Fernmetastasen (M-Stadium) beurteilt. Es wird zwischen einer klinischen Stadieneinteilung (cTNM) auf der Basis der digital-rektalen Untersuchung, des PSA-Wertes und bildgebender Verfahren sowie einer pathologischen Klassifizierung (pTNM) anhand des endgültigen histologischen Befundes des Prostatektomiepräparates unterschieden. Der Sinn des Stagings liegt in der Möglichkeit, Patienten individuell einer entsprechenden Therapieform zuzuführen (klinisches Staging) oder prognostische Aussagen zu treffen (pathologisches Staging). Mittels der DRU wird eine Einschätzung des cT-Stadiums, der lokalen Tumorausdehnung eines PCa, vorgenommen. Wie bereits zuvor erwähnt, ist die DRU stark vom jeweiligen Untersucher und dessen Erfahrung abhängig, so dass im Vergleich zum endgültigen pathologischen (pT)-Stadium oft falsche Vorhersagen getroffen werden. So zeigte eine Studie zur Genauigkeit der DRU (Partin, Kattan et al. 1997), dass die klinische Vermutung eines auf die Prostata begrenzten Karzinoms in 52% zutraf. Bei 31% der Patienten zeigte sich ein extrakapsuläres Wachstum oder eine Samenblaseninfiltration, obwohl dies durch den Tastbefund nur bei 17% prognostiziert wurde. Ein cT4-Stadium (organüberschreitender Tastbefund) konnte nur in 36%

bestätigt werden, bei 19% war das Karzinom auf die Prostata beschränkt und bei 45% in die Samenblase infiltriert.

Erwähnenswert ist noch die Beobachtung mehrerer Autoren, dass mit einem Anstieg des klinischen Stadiums auch das Risiko für das Auftreten eines biochemischen Rezidivs nach radikaler Prostatektomie zunimmt (Partin, Pound et al. 1993; Hull, Rabbani et al. 2002; Chun, Graefen et al. 2006).

<b>T</b>	<b>Lokale Ausdehnung des Primärtumors</b>
TX	Primärtumor nicht bestimmbar
T0	Kein Anhalt für einen Primärtumor
T1	Klinisch nicht erkennbarer Tumor
T1a	Tumor zufälliger histologischer Befund (<5% des Resektionsgewebes)
T1b	Tumor zufälliger histologischer Befund (>5% des Resektionsgewebes)
T1c	Tumor durch Nadelbiopsie diagnostiziert
T2	Primärtumor ist auf Prostata beschränkt
T2a	Tumor befällt ≤50% eines Seitenlappens
T2b	Tumor befällt ≥50% eines Seitenlappens
T2c	Tumor befällt beide Seitenlappen
T3	Primärtumor durchbricht die Prostatakapsel
T3a	Extrakapsulärer Tumor
T3b	Tumor infiltrierte die Samenblase
T4	Primärtumor ist fixiert oder infiltrierte benachbarte Strukturen
<b>N</b>	<b>Regionäre Lymphknoten</b>
NX	Regionäre Lymphknoten können nicht beurteilt werden
N0	Kein Anhalt für regionäre Lymphknoten
N1	Regionärer Lymphknotenbefall
<b>M</b>	<b>Fernmetastasen</b>
MX	Fernmetastasen können nicht beurteilt werden
M0	Kein Anhalt für Fernmetastasen
M1	Vorliegen von Fernmetastasen
M1a	Extraregionärer Lymphknotenbefall
M1b	Knochenmetastasen
M1c	Andere Manifestationen

**Tabelle 1: TNM-Klassifikation**

### 1.5 Therapeutische Möglichkeiten des Prostatakarzinoms

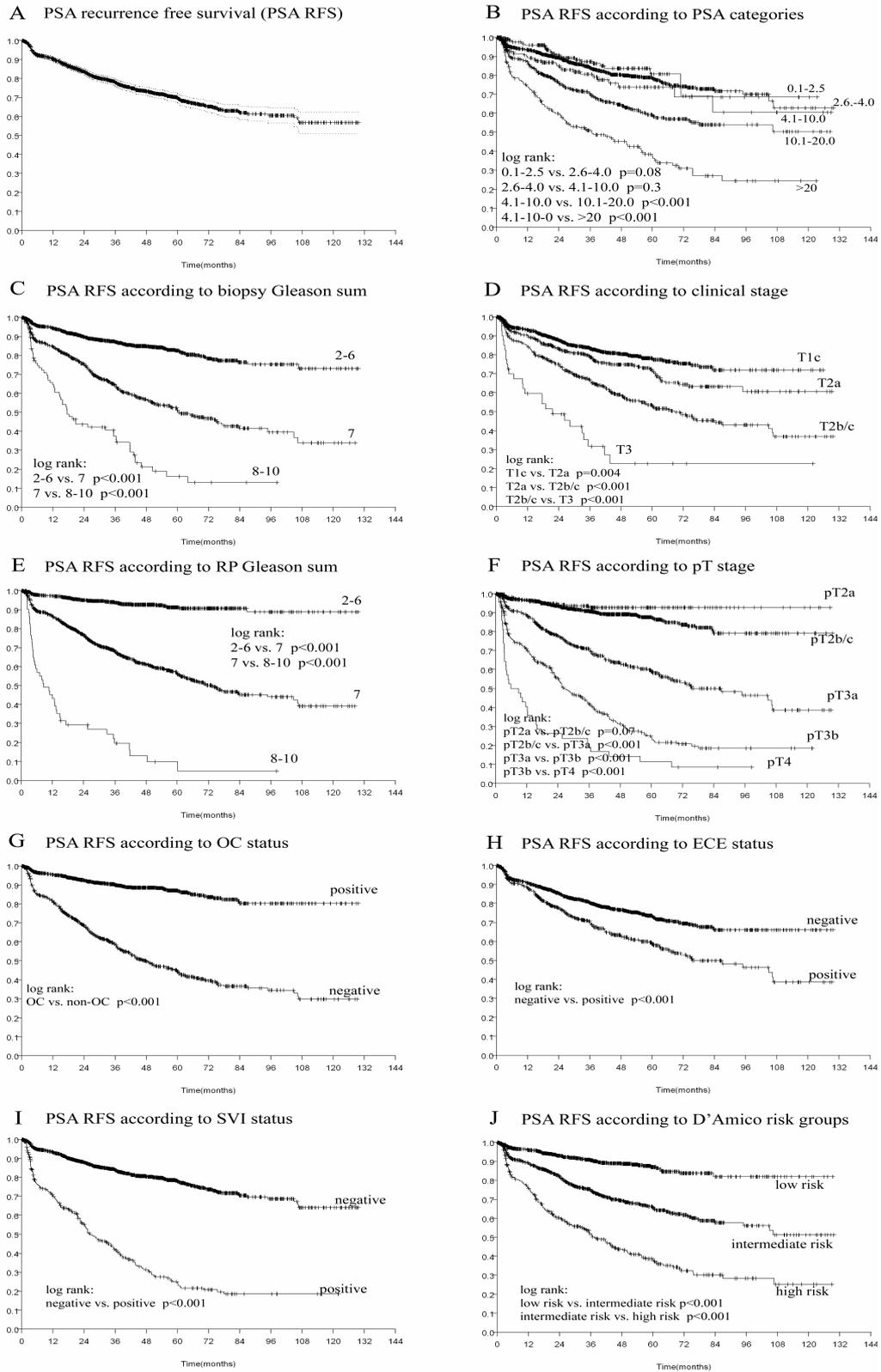
Verschiedene Therapieoptionen zur Behandlung eines klinisch lokalisierten, lokal fortgeschrittenen oder metastasierten PCa stehen zur Verfügung. Allen Therapieoptionen gemein ist neben der Krebskontrolle eine Aufrechterhaltung der Lebensqualität bezüglich der Potenz und/oder der Kontinenz.

Die Frage, welche Methode für den jeweiligen Patienten am ehesten in Frage kommt, muß individuell anhand unterschiedlicher Faktoren wie Alter, Komorbidität mit entsprechender Lebenserwartung, Tumorausdehnung sowie Differenzierungsgrad geklärt werden.

## **Lokalisiertes bzw. lokal fortgeschrittenes PCa**

Den heute am häufigsten gewählten Eingriff zur Behandlung des PCa stellt bei unter 70-jährigen die radikale Prostatektomie dar. 1889 wurde von Czerny die erste totale perineale Prostatektomie bei einem Patienten mit PCa durchgeführt. Heute wird meist der offene, retropubische Zugang durchgeführt, für den exzellente onkologische Ergebnisse nachgewiesen sind und der dazu niedrige Komplikationsraten aufweist (Hull, Rabbani et al. 2002; Augustin, Hammerer et al. 2003; Michl, Graefen et al. 2003; Chun, Graefen et al. 2006). Wie in Abbildung 3 dargestellt, zeigten Chun et al., dass nach radikaler retropubischer Prostatektomie das rezidivfreie Überleben bei einem pathologisch nachgewiesenen Organbegrenzten Stadium nach 10 Jahren bei 85% lag. Wenn kein suspekter klinischer Tastbefund vorliegt (cT1-Stadium), ist das rezidivfreie Überleben 10 Jahre nach radikaler retropubischer Prostatektomie zu 75% gewährleistet. Die anatomischen Studien von Walsh und Donker stellen außerdem die Basis für die Variante der nerverhaltenden radikalen retropubischen Prostatektomie dar (Walsh and Donker 1982). Bei dieser Operationsform bemüht man sich, die Nervi erigentes zu schonen, die für die Gliedsteife zuständig sind und der Prostatakapsel, die bei der Operation entfernt wird, direkt anliegen. Auf diese Weise möchte man die Erektionsfähigkeit der Patienten erhalten; die Entfernung aller Krebszellen bleibt jedoch vorrangig.

Weitere operative Optionen neben dem offen-retropubischen Zugang stellen der offen-perineale Zugang, der laparoskopische Ansatz: transperitoneal vs. extraperitoneal (Stolzenburg, Truss et al. 2003) und die Roboter-assistierte Methode dar.



**Abbildung 3: Gesamt-PSA freies Überleben sowie stratifiziert nach klinischen Charakteristika nach radikaler retropubischer Prostatektomie aus Chun et al. WJU 2006 (Chun, Graefen et al. 2006)**

Eine andere Therapieoption stellt die Strahlentherapie dar. Sie ist eine gute Alternative zum operativen Vorgehen für Patienten, für die z.B. eine Operation aus Altersgründen nicht in Frage kommt, die eine Operation generell verweigern oder bei denen Risikofaktoren (wie z.B. Narkoseunverträglichkeit oder kreislaufinsuffiziente Zustände) vorhanden sind. Neben der herkömmlichen perkutanen Bestrahlung, die mit einer Gesamtdosis von maximal 70 Gy durchgeführt werden kann, steht noch die Brachytherapie (Kurzdistanzbestrahlung) zur Verfügung (Deger, Bohmer et al. 2001), welche als Hoch-Dosis-Brachytherapie mit 50,4 Gy („high-dose“, HDR) durchgeführt werden kann oder alternativ als Niedrig-Dosis-Verfahren („low-dose“, LDR). Bei der LDR-Brachytherapie wird über mehrere Monate hinweg von den implantierten Quellen eine verhältnismäßig schwache radioaktive Strahlung an das Prostatagewebe abgegeben. Dabei wird eine Gesamtdosis von 145 Gy erreicht. Die Brachytherapie ist eine Strahlentherapie „von innen“, bei der kleinste radioaktive Teilchen (Seeds) unter Ultraschallkontrolle durch den Damm über Hohlnadeln in die Prostata permanent (LDR) oder transient (HDR) eingebracht werden. Permanente Strahlenquellen stellen dabei Palladium-103 und Jod-125 dar, wohingegen als temporäre Strahlenquelle in der Regel Iridium-192 angewandt wird.

Die LDR-Brachytherapie des Prostata-Karzinoms wurde in den sechziger Jahren in Europa entwickelt, setzte sich in den USA durch und hat in den letzten Jahren auch in Deutschland immer mehr Anwendung gefunden (Siegsmund, Musial et al. 2001). Das Verfahren dauert ungefähr eine Stunde und wird unter Narkose durchgeführt. Die radioaktiven Seeds aus Palladium-103 oder Jod-125, bestrahlen das umliegende Prostatagewebe gleichmäßig von innen, bis sie im Laufe eines Jahres ihre Radioaktivität verlieren. Sie können nach der Therapie in der Prostata verbleiben. Daneben existiert noch die Hochdosis(HDR)-Brachytherapie - auch Afterloading genannt. Hierbei handelt es sich um ein strahlentherapeutisches Therapieverfahren, das auch bei aggressiver Tumorbildung

angewendet werden kann, zumindest solange noch keine Metastasen nachgewiesen werden können. Auch hierbei wird eine radioaktive Strahlenquelle (Iridium 192 oder Palladium 103) in die Prostata eingebracht, nach der Behandlung jedoch wieder entfernt. Die Tumorzellen werden in mehreren operativen Sitzungen kurzzeitig radioaktiv bestrahlt und somit irreparabel geschädigt. Im Anschluss erfolgt noch eine ergänzende tägliche perkutane Bestrahlung für die Dauer mehrerer Wochen.

Eine weitere Alternative für Arzt und Patient stellt die Methode des abwartenden Beobachtens - "active surveillance" - dar. Dieser Therapieansatz sieht vor, dass Patienten nach Diagnosestellung zunächst keine kurativ angelegte Therapie erhalten. Es werden vielmehr engmaschig Kontrolluntersuchungen durchgeführt, um ein lokales Fortschreiten des Tumors oder eine Metastasierung möglichst frühzeitig zu entdecken. Dieser Ansatz hat sich unter der Annahme entwickelt, dass manche Tumoren so langsam wachsen, dass sie die Lebensqualität des Patienten voraussichtlich nicht gefährden.

### **Metastasiertes PCa**

Als weitere Alternative steht noch die medikamentöse Therapie zur Verfügung. Mittels einer Hormonentzugstherapie kann das Wachstum des Tumors über einen längeren Zeitraum aufgehalten und kontrolliert werden – eine Heilung ist auf diese Weise jedoch nicht möglich (Ebert 2001). Das männliche Geschlechtshormon Testosteron, welches von den Leydig-Zellen produziert wird, scheint einen wesentlichen Einfluß auf das beginnende Wachstum von Krebszellen zu haben. Man kann nun auf verschiedenen Arten eingreifen. Zum einen kann die Wirkung des Testosterons durch Gabe von sogenannten Antiandrogenen blockiert werden. Und zum anderen ist es möglich, GnRH-Analoga (Gonadotropin Releasing Hormone Gonadoliberin) als Depotspritzen zu verabreichen, die die Testosteronproduktion direkt

drosseln (Wedding 2000; Wolff 2004) . Die Auswahl der Medikamente richtet sich dabei ganz nach den Charakteristika des Tumors und des Lebensalters des Patienten.

## **1.6 Risikovorhersagemodelle**

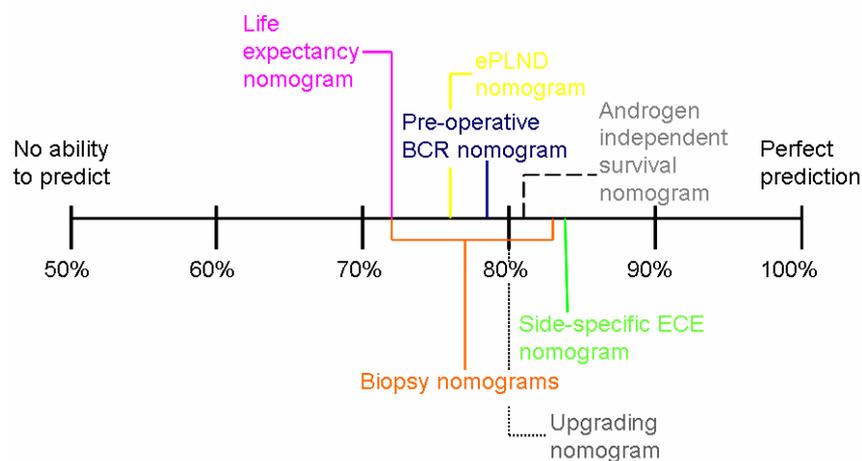
Das Feld der Vorhersagemöglichkeiten in Bezug auf das Prostata-Karzinom hat sich in den letzten Jahren sehr stark ausgeweitet. Klinikern steht bei ihrer Arbeit eine ganze Reihe an evidenz-basierten Modellen zur Verfügung. Diese Vorhersagemodelle kombinieren unterschiedliche prä- oder post-operative Risikofaktoren, um einen gewünschten Endpunkt vorherzusagen. Somit stellen sie wichtige Entscheidungshilfen für den Kliniker und den Patienten dar, um eine fundierte Beratung zu ermöglichen (Chun, Karakiewicz et al. 2006; Karakiewicz, Chun et al. 2006). Es gibt jedoch unterschiedliche statistische Ansätze, um Ereignisse zu prädictieren. Hierzu gehören so genannte artifizielle neuronale Netzwerke (ANN), Nomogramme und Classification and Regression Trees (CART), die auch als Entscheidungsbaum-Modelle bezeichnet werden.

### **1.6.1 Kriterien**

*Komplexität:* Bei der Entwicklung und Nutzung von Risiko-Vorhersagemodellen sollte man beachten, dass die Modelle nicht übermäßig kompliziert sein sollten, um eine Anwendung im klinischen Alltag zu gewährleisten. Dafür sollten sie über eine überschaubare Anzahl und leicht zugängliche klinisch-pathologische Vorhersageparameter verfügen.

*Prädiktive Genauigkeit:* Die Genauigkeit in der Vorhersage stellt eine der entscheidenden Kriterien bei der Bewertung von Risiko-Vorhersagemodelle dar. Für gewöhnlich wird die Vorhersagegenauigkeit mittels ROC (Receiver operating characteristics) und der AUC (Area under the curve) in Prozent dargestellt. In Kombination mit einer Kalibrierung, wie es in dieser Arbeit angegeben ist, bildet die ROC die prädiktive Genauigkeit. Somit kombiniert die prädiktive Genauigkeit sowohl Diskriminierung als auch Kalibrierung. Werte für die

prädiktive Genauigkeit schwanken zwischen 50 und 100%, wobei 50% dem Werfen einer Münze entspricht und 100% perfekter Vorhersage. Da jedoch kein Modell in der Anwendung perfekt ist und zu 100% richtig vorhersagen kann, werden in der Regel Genauigkeitswerte zwischen 70 und 80% akzeptiert (Chun, Karakiewicz et al. 2006). Die Vorhersagewerte sollten idealerweise anhand einer externen Patientenkohorte auf ihre Genauigkeit hin überprüft werden. Abbildung 4 veranschaulicht anhand einer Skala die Einordnung der prädiktiven Genauigkeit und stellt diese für verschiedene PCa-Nomogramme dar.



**Abbildung 4: Prädiktive Genauigkeit unterschiedlicher PCa-Nomogramme aus Chun et al. WJU 2007 in press**

*Anwendbarkeit/Kalibrierung:* Wenn auch die prädiktive Genauigkeit als Beurteilungskriterium über allen anderen Punkten steht und einen Endpunkt vorhersagen kann, erhält der Anwender dennoch keinerlei Informationen über die prädiktive Genauigkeit des Risiko-Vorhersagemodells, wenn man es auf bestimmte Patienten-Untergruppen anwendet. So wird es beispielsweise Modelle geben, die gute Vorhersagen bei Hochrisiko-Patienten liefern, auf Niedrigrisiko-Patienten angewandt hingegen schlechte Ergebnisse erzielen. Und wieder andere werden durchgehend gute prädiktive Genauigkeiten liefern können (Karakiewicz, Chun et al. 2006). Um dem Anwender einen Überblick über die prädiktive Genauigkeit der Modelle in

Bezug auf verschiedene Risikostufen zu geben, gibt es die Möglichkeit, dies graphisch mittels eines „Kalibrierungsdiagrammes“ darzustellen und abzulesen (Steyerberg, Harrell et al. 1998; Kattan 2006).

*Allgemeingültigkeit:* Vorhersagemodelle entstehen auf der Grundlage der Patientencharakteristika, anhand derer sie entwickelt wurden. Bevor ein Kliniker ein Modell benutzt, sollte von ihm sichergestellt werden, dass es anhand eines Patientenkollektivs entwickelt wurde, welches über ähnliche Krankheitseigenschaften verfügt wie der aktuell einzuschätzende Patient (Bradley 1993; Steyerberg, Harrell et al. 2001). Darüber hinaus muss bedacht werden, dass Patientencharakteristika im internationalen Vergleich Unterschiede aufweisen können. Nachdem Studien ergeben haben, dass ethnische Faktoren Einfluss auf die Prävalenz des PCa haben, besteht die Notwendigkeit, die Allgemeingültigkeit der Risiko-Vorhersagemodelle zu prüfen, wenn sie auf Patienten mit gleichen Krankheitseigenschaften aber unterschiedlicher Nationalität angewendet werden. Dieser Tatsache folgend wurde 2006 als Neuerung ein Erstbiopsie-Nomogramm eingeführt, welches ethnisch-spezifisch Unterschiede in der Prävalenz des PCa vorhersagt vor allem im südasiatischen Raum (Suzuki, Komiya et al. 2006). Und auch für den afro-amerikanischen Raum wurde ein solches individualisiertes Risiko-Vorhersagemodell eingeführt (Yanke, Carver et al. 2006).

*Direktvergleich der Modelle:* Der Vergleich einzelner Modelle untereinander ist auch von besonderer Relevanz, wenn man bedenkt, dass für einige vorherzusagende Endpunkte mehrere Risiko-Vorhersagemodelle angewendet werden könnten. Dem Kliniker kann durch den direkten Vergleich bei der Frage geholfen werden, welches der Vorhersagemodelle für ihn und seinen Patienten am besten geeignet erscheint.

Wenn etablierte Vorhersagemodelle auf ihre prädiktive Genauigkeit und Gültigkeit beurteilt werden, sollte dies im direkten Vergleich mit einer externen Kohorte stattfinden. So kann objektiv eingeschätzt werden, welches Modell die höchste prädiktive Genauigkeit und beste

Kalibrierung aufweist und somit das geeignetste für die Patientenberatung ist. (Kattan 2002; Kattan 2003; Chun, Karakiewicz et al. 2006).

### **1.6.2 Limitationen der Risikovorhersagemodelle**

Wenn man sich an Vorteilen der Vorhersagemodelle orientiert, etwa der Tatsache, dass Erfahrungswerte von tausenden von Patienten in ihnen zusammengefasst werden können und das Modell ein klinisches Ereignis damit präziser vorhersagen kann als ein Kliniker (Ross, Scardino et al. 2001; Specht, Kattan et al. 2005), gilt es auch deren Limitation zu beachten. Allgemein lässt sich sagen, dass der Nachteil der Modelle in der Abhängigkeit von ihrem jeweiligen Datensatz besteht, auf deren Basis sie einst entwickelt wurden. In der Folge mag es sein, dass die Daten der Patienten recht ausgefallen sind, da die Kohorten oftmals in hoch spezialisierten Zentren erhoben wurden (Kattan, Eastham et al. 1998; Kattan, Wheeler et al. 1999; Chun, Steuber et al. 2006). Desweiteren beruhen Nomogramme stets auf retrospektiv zusammengestellten Daten, so dass daher eine Notwendigkeit zur ständigen Aktualisierung der Risiko-Vorhersagemodelle gemäß den Veränderungen des Krebses oder der Untersuchungsmethoden besteht. Um zum Beispiel eine Vergleichbarkeit von Risiko-Vorhersagemodelle zu erreichen, die zu unterschiedlichen Zeiten entwickelt wurden, ist es daher notwendig, Modelle, die auf Ergebnissen von Sextantenbiopsien beruhen, auf den Goldstandard der ausgedehnteren Mehrfachbiopsien ( $\geq 10$  Biopsien) zu bringen (Chun, Briganti et al. 2006). Ein weiterer Nachteil der Risiko-Vorhersagemodelle stellt der Umstand dar, dass sie nicht auf alle Patienten gleichermaßen anwendbar sind. Wenn Patienten zum Beispiel eine neoadjuvante Hormontherapie erhalten haben, ist dies ein Ausschlusskriterium für die meisten Vorhersagemodelle. Auch der Einsatz des tPSA-Wertes in der Vorhersage des PCa ist nicht unkritisch zu sehen. Obwohl man sagen kann, dass seine Einführung die Vorhersage des PCa mittels Risiko-Vorhersagemodellen revolutionierte und der überwiegende Anteil der Modelle den tPSA-Wert als Parameter beinhaltet, ist seine Spezifität –wie bereits

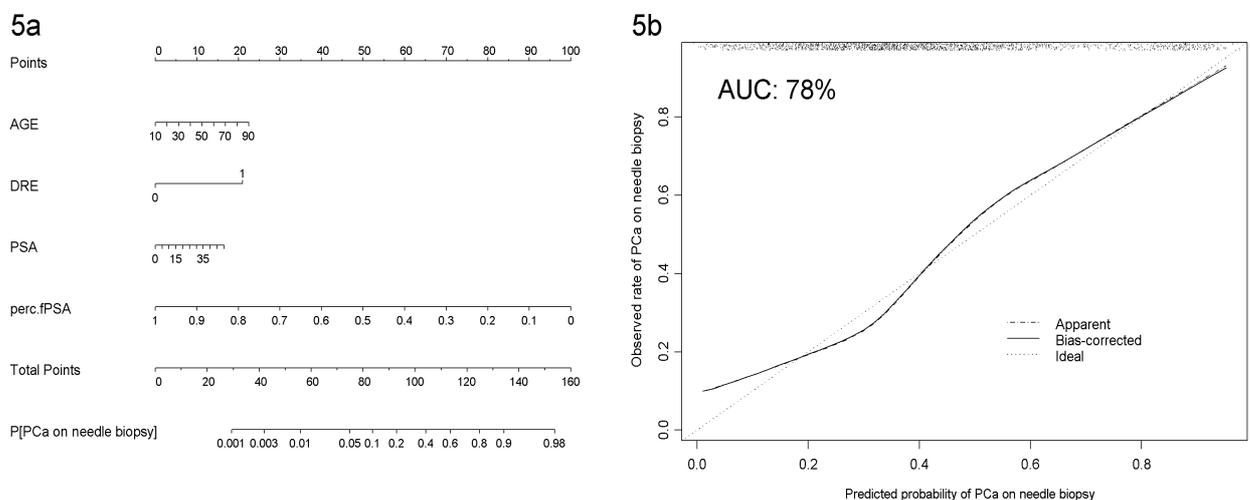
zuvor beschrieben- limitiert. Daher sollte besonderes Augenmerk auf die Entdeckung neuer Vorhersageparameter gelegt werden (Shariat, Canto et al. 2004). Als aussichtsreiche neue serologische Marker sind in diesem Zusammenhang Interleukin-6 sowie hK2 und TGF Beta im Gespräch, die die Vorhersage eines biochemischen Rezidivs nach radikaler Prostatektomie verbessern sollen (Kattan 2003; Steuber, Vickers et al. 2006). In gleicher Weise müssen die von den Modellen vorherzusagenden Endpunkte überdacht werden. Beispielsweise stellt das PSA-Rezidiv als Endpunkt lediglich einen Surrogatmarker des Scheiterns einer lokalen Therapieform dar. In diesem Zusammenhang sind Daten bezüglich lokaler oder distaler Metastasen bzw. dem Prostatakarzinom-spezifischen Überleben aussagekräftiger. Diese Daten liegen jedoch wegen des natürlichen Verlaufs eines Prostatakarzinoms noch nicht vor. Abschließend muß noch erwähnt werden, dass kein Risiko-Vorhersagemodell perfekt vorhersagt, da kein statistisches Modell in der Lage ist, alle biologisch relevanten Risikofaktoren zu berücksichtigen (Kattan 2002).

### **1.6.3 Nomogramme**

Auf dem Gebiet der evidenz-basierten Vorhersage von Prostata-Karzinomen gibt es, wie zum Teil schon zuvor beschrieben, eine ganze Reihe an „Entscheidungshilfen“, die dem Kliniker zur Verfügung stehen. Es hat sich jedoch im Alltag eingebürgert, verschiedenste statistische Modelle (z.B. ANN, CART-Analyse, Look-up-Tafeln) mit dem Begriff „Nomogramm“ zu versehen.

Dabei ist der Begriff „Nomogramm“ im engeren Sinne für eine ganz spezielle Art der statistischen Repräsentation reserviert. Es handelt sich hierbei um graphische Vorhersagemodelle, bei denen als Ausgangsbasis mit Ziffern markierte Balken benutzt werden, die mittels traditioneller statistischer Methoden ermittelt wurden. Multivariable logistische Regressionsanalysen, wenn es darum geht, ein binäres Vorhandensein im Sinne „Ja“ oder „Nein“ vorauszusagen oder Cox-Regressionsanalysen, um ein prognostisches

Auftreten innerhalb einer gewissen Zeitspanne zu klären. Nomogramme tauchten das erste Mal um 1775 herum in den statistischen Atlanten von William Playfair, Diagrammen von James Watt und Schriften von Johann Heinrich Lambert auf (Lambert 1779). Das erste speziell medizinische Nomogramm wurde hingegen im Jahr 1928 von dem früheren Physiologen und Präsident der Gesellschaft für Wissenschaftsgeschichte L.J. Henderson veröffentlicht (Henderson 1928), der es dazu nutzte, das komplizierte chemische System des Blutes zu analysieren. Heutzutage sind weit mehr als 1700 Publikationen in Medline registriert, die sich auf Nomogrammen beziehen. Etwa 1100 dieser Publikationen stammen aus den 90er Jahren, bedingt durch die Einführung und Weiterentwicklung der Digitalisierung. Besonders wichtig ist in diesem Zusammenhang die S-Programmiersprache, die die Basis für statistische Anwendungen darstellt und es dem Anwender ermöglicht, komplexe mathematische Funktionen graphisch darzustellen (Bianco 2006).



**Abbildung 5: Schematische Darstellung eines Nomogramms unter Verwendung prä-operativer Risikofaktoren zur Vorhersage eines Prostatakarzinoms in der Erstbiopsie aus Karakiewicz et al. JUrol 2005 (Karakiewicz, Benayoun et al. 2005) mit Kalibrierung**

Um die Aussage des Nomogramms zu erhalten, platziert man der Reihe nach die Werte des Patienten auf den dazugehörigen Achsen/Balken, wie es in der Abbildung 5a dargestellt ist.

Anschließend zeichnet man jeweils eine vertikale Linie zur „Punkteachse“, um

herauszufinden, wie viele Punkte für jeden Wert des Patienten vergeben werden. Diese einzelnen Punktwerte addiert man dann zu einem Gesamtwert, dessen Ergebnis zuletzt auf der Achse für die Gesamtpunktezahl eingetragen wird. Der Nutzer des Nomogrammes ist nun in der Lage, das Risiko (in %) für den Patienten in Bezug auf sein individuell vorherzusagendes Ereignis von der Gesamtpunkteskala abzulesen (Abbildung 5a).

Innerhalb des Kalibrierungsdiagrammes (Abbildung 5b) repräsentieren die x-Achsen die jeweilige Vorhersage der Modelle für das Vorhandensein eines PCa und ihre y-Achsen stellen die Anzahl an tatsächlich in Biopsien beobachteten PCa dar. Perfekte Übereinstimmung stellt hierbei eine 45° Linie dar. Vorhersage und tatsächliches Vorhandensein eines PCa entsprechen sich zu 100%. Punkte, die oberhalb der 45° Linie angesiedelt sind, zeigen eine Unterschätzung des Modells an, wohingegen Punkte, die sich unterhalb der 45° Linie befinden anzeigen, dass das Risiko vom Vorhersagemodell zu hoch eingeschätzt wurde. Die Ergebnisse sind nach Kalibrierung zusätzlich graphisch darstellbar und so im Hinblick auf ihre prädiktive Genauigkeit für unterschiedliche Risikobereiche gut zu beurteilen.

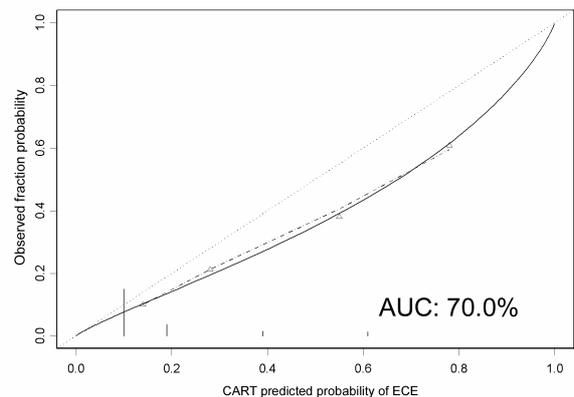
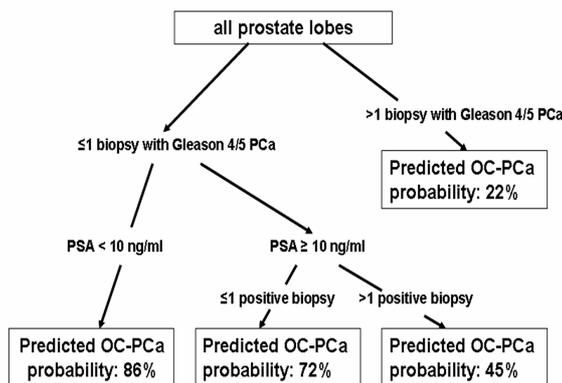
#### **1.6.4 CART (Classification and Regression Tree)**

CART ist ein Algorithmus, der zur Entscheidungsfindung dient. Er wurde erstmals 1984 von Leo Breiman publiziert (Breiman 1984). Die CART-Analyse kann man im Grunde mit dem Begriff „Entscheidungsbaum“ gleichsetzen.

Ein bedeutendes Merkmal des CART-Algorithmus stellt die Tatsache dar, dass es ein binäres, kategorisches System ist. Demzufolge beginnt die CART-Analyse mit einem Grundstamm, an dessen Ende sich zwei Verzweigungen befinden, verbunden mit den Antworten „Ja“ oder „Nein“ in Bezug auf eine gestellte Frage. Es wird nun eine ganze Reihe von Fragen gestellt, die allesamt mit „Ja“ oder „Nein“ beantwortet werden können. Diese Serie ergibt dann eine Klassifikation - ein Resultat -, welches durch eine Regel bestimmt ist,

die einfach ablesbar ist, wenn man von der Wurzel her den Ästen des Baumes bis zum Endpunkt folgt.

Bei dieser Analyse ist es sehr wichtig, bei jedem Schritt genau das Attribut zu erfragen, mit welchem man die Daten am besten klassifizieren kann. Um diese optimale kategorische Trennung zu ermitteln, muss der beste „Split“ gefunden werden, d.h. die Aufteilung der Daten muss so gewählt werden, dass sie nach der Aufteilung möglichst rein sind und eine Maximierung des Informationsgehaltes vorliegt. Das Maß für die Reinheit der Daten stellt hierbei die Entropie dar. Aus der Entropie lässt sich dann berechnen, welches Attribut für den Split den höchsten Info-Gehalt bietet. In der Abbildung 6 ist ein Beispiel für einen CART-Algorithmus dargestellt.



**Abbildung 6: Schematische Darstellung einer CART-Analyse unter Verwendung prä-operativer Risikofaktoren zur Vorhersage eines extraprostatatischen PCa-Wachstums im endgültigen pathologischen Präparat mit Kalibrierung. Aus Karakiewicz et al: CanJUrol 2006 (Karakiewicz, Chun et al. 2006)**

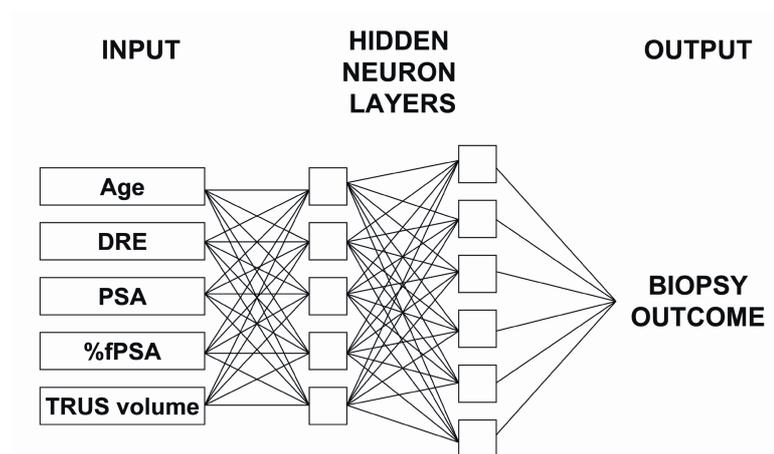
### 1.6.5 Künstliche neuronale Netzwerke (ANN)

Künstliche neuronale Netzwerke sind komplexe mathematische Modelle, die mittlerweile über mehr als 50 Jahre ausgiebig untersucht und analysiert wurden. Auf dem Gebiet der PCa-Forschung werden Künstliche neuronale Netzwerke seit dem Jahre 1994 eingesetzt, was auf

P.B. Snow zurückgeht (Snow, Smith et al. 1994). Snow demonstrierte damals den Nutzen eines ANN für die Prädiktion eines späteren PCa-Rezidivs anhand des Einsatzes von PSA und anderer präoperativer Parameter.

ANN sind Modelle, die ganz entfernt mit der Arbeitsweise des menschlichen Nervensystems und seiner Neuronenschaltungen verglichen werden können, da sie bestimmte Funktionen und Topologien eines zentralen Nervensystems simulieren. Wegen ihrer Fähigkeit zu generalisieren und Abstraktionen zu generieren, gewinnen sie zunehmend an Popularität (Wei, Zhang et al. 1998; Revold 2001).

In den meisten Fällen ist ein ANN ein angepasstes System, welches seine Struktur je nach äußeren und inneren Informationen, die sein Netzwerk durchlaufen, ändert. Hierbei werden mehrere neuronale Stationen durchlaufen, welche sowohl Interaktionen der einzelnen Parameter als auch bestimmte Muster in Datensätzen erkennen. Ein ANN besitzt - vereinfacht gesagt - die Fähigkeit anhand von Informationen und Datensätzen zu lernen. In Abbildung 7 ist ein Beispiel für ein ANN angegeben.



**Abbildung 7: Schematische Darstellung eines künstlichen neuronalen Netzwerks**

## 1.7 Arbeitshypothese

In der vorliegenden Arbeit wurden vier existierende Prostata-Erstbiopsie-Vorhersagemodelle (Stephan, Cammann et al. 2002; Garzotto, Hudson et al. 2003; Garzotto, Beer et al. 2005; Karakiewicz, Benayoun et al. 2005) anhand eines externen Patientenkollektivs (n = 3980) miteinander verglichen. Die verglichenen Erstbiopsie-Vorhersagemodelle basieren auf verschiedenen statistischen Methoden. Das Modell von Karakiewicz (Karakiewicz, Benayoun et al. 2005) stellt ein auf logistischen Regressionsanalysen basierendes Nomogramm dar, ebenso eines der Garzotto-Modelle (Garzotto, Hudson et al. 2003). Auch eine CART-Analyse (Garzotto, Beer et al. 2005) und ein ANN (Stephan, Cammann et al. 2002) wurden miteinander verglichen.

In unserer Arbeit untersuchten wir jedes einzelne Modell bezüglich seiner prädiktiven Genauigkeit. Neben der prädiktiven Genauigkeit der Modelle wurden ebenfalls Vergleiche der sogenannten „performance characteristics“ oder Kalibrierungen durchgeführt. Es sollte beachtet werden, dass nach unserem Wissensstand noch nie zuvor ein derartiger Vergleich von konkurrierenden Erstbiopsie-Vorhersagemodellen der Prostata stattgefunden hat.

Die Hypothese lautet dabei, dass eines der Verfahren den anderen in Bezug auf die prädiktive Genauigkeit und die „performance characteristics“ oder Kalibrierungen überlegen ist.

## **2 Methoden**

### **2.1 Analyisierte Modelle**

In dieser Arbeit wurden insgesamt vier etablierte, auf unterschiedlichen statistischen Methoden beruhende Erstbiopsie-Vorhersagemodelle zur Einschätzung des Prostatakarzinomrisikos bezüglich ihrer prädiktiven Genauigkeit verglichen. Es handelte sich dabei um folgende Modelle:

#### **2.1.1 Nomogramme**

Ein von Karakiewicz et al. publiziertes Nomogramm (Karakiewicz, Benayoun et al. 2005) mit den vier Eingabevariablen Alter, DRU, Serum-PSA (0,1-10 ng/ml) und %fPSA wurde abgeglichen. Es wurde mittels Daten aus Hamburg, Deutschland (1762 Patienten) entwickelt und anhand von 514 Patienten aus Montreal, Kanada extern validiert. Seine ermittelte prädiktive Genauigkeit lag bei 78,0%, nach externer Validierung durch den kanadischen Patientenstamm bei 77%. Die Patienten erhielten hier eine Sextantenbiopsie.

In der vorliegenden Arbeit wurde noch ein weiteres Nomogramm verglichen (Garzotto, Hudson et al. 2003). Es beinhaltet an Eingabevariablen Alter, DRU, TRUS, Serum-PSA-Wert  $\leq 10$  ng/ml und die PSA-Dichte. Es wurde an einer Population von 1237 Patienten entwickelt, die mindestens eine Sextantenbiopsie erhielten und erreichte eine Vorhersagegenauigkeit von 73,0 %. Von den 1237 Patienten wurden 20% (247 Patienten) zur externen Validierung eingesetzt.

#### **2.1.2 CART-Analyse (Classification and Regression Tree)**

Die in dieser Arbeit untersuchte CART-Analyse von Garzotto (Garzotto, Beer et al. 2005) beinhaltet an Eingabevariablen: Alter, den Serum-PSA-Wert ( $\leq 10$  ng/ml), die PSA-Dichte, den TRUS und das Prostatavolumen. Die CART-Analyse wurde ursprünglich anhand eines Patientenkollektivs von 1433 Männern entwickelt, von denen 20 % (287 Patient) zur externen

Validierung ausgewählt wurden. Die Patienten erhielten mindestens eine Sextantenbiopsie (Bereich 6-13 Biopsiezylinder) und wiesen alle einen Serum-PSA-Wert  $\leq 10$  ng/ml auf; die Vorhersagegenauigkeit betrug dort 73,0%.

### **2.1.3 Artifizielles neuronales Netzwerk (ANN)**

Das in dieser Arbeit getestete Artifizielle neuronale Netzwerk (ANN) geht auf Stephan zurück (Stephan, Cammann et al. 2002), und beinhaltet fünf Eingabevariablen: Alter, DRU, Serum-PSA (2-20 ng/ml), %fPSA und das Prostatavolumen. Seine ursprünglich ermittelte Vorhersagegenauigkeit wurde mit 78% angegeben (Stephan, Cammann et al. 2002). Es wurde 2002 anhand eines Patientenkollektivs von 1188 Männern entwickelt, bei denen entweder eine Sextanten- oder Oktantenbiopsie durchgeführt wurde. Zur externen Validierung durchliefen mehrere Screeningkohorten das ANN.

## **2.2 Patientenpopulation**

Das Patientenkollektiv zur externen Validierung der vier Vorhersagemodelle bestand ursprünglich aus 4661 männlichen Patienten, die aufgrund eines suspekten Tastbefundes und/oder einem abnorm veränderten PSA-Wert auffällig geworden waren und im Zeitraum Januar 2000 bis Juni 2005 der Urologischen Universitätsklinik Hamburg-Eppendorf mit dem Verdacht auf PCa zugewiesen und anschließend einer Prostata-Erstbiopsie unterzogen wurden. Von den ursprünglich 4661 Patienten wurden 681 von der Überprüfung ausgeschlossen, da für sie nicht alle klinischen Parameter vorlagen, die für die vier verschiedenen Vorhersagemodelle benötigt wurden. Somit erfolgten die Validierungen an 3980 Patienten.

Das mittlere Alter der Patientenpopulation lag bei 65 Jahren. Der mittlere PSA-Wert lag bei dem Patientenkollektiv bei einem Wert von 7,2 ng/ml, der mittlere %fPSA-Wert bei einem Wert von 19,2.

Die für die Nomogramme notwendigen klinischen Angaben umfassten insgesamt sieben Parameter: das Patientenalter, das Prostatagesamtvolumen, das Ergebnis der DRU, den präoperativen Serum-PSA-Gesamtwert, den prozentualen freien PSA Anteil (%fPSA), die PSA-Dichte und die Ergebnisse des transrektalen Ultraschalles.

### **2.3 Histologische und klinische Daten**

Auf unser Patientenkollektiv wurde der Abbott Axym PSA Assay (Abbott Park, IL, USA) angewendet und der PSA-Wert wurde zu Beginn der Behandlung gemessen, bevor die digital rektale Untersuchung und der transrektale Ultraschall durchgeführt wurden. Für jeden Patienten wurde das klinische Stadium anhand der 6.Auflage des AJCC (Greene 2002) festgelegt. Die mittels transrektalem Ultraschall gesteuerten Biopsien wurden nach den aktuellen Standards durchgeführt (Presti, Chang et al. 2000; Chun, Briganti et al. 2006). Das durch transrektalen Ultraschall ermittelte Prostatavolumen wurde durch Anwendung der Formel ( $0,52 \times \text{Länge} \times \text{Breite} \times \text{Höhe}$ ) erzielt. Alle Biopsie-Proben wurden mittels des Gleason-Grading Systems (Gleason and Mellinger 1974) durch einen Mitarbeiter des Pathologischen Instituts der Universitätsklinik Hamburg-Eppendorf histologisch aufgearbeitet

### **2.4 Statistische Analysen**

#### **2.4.1 Regressionsanalyse: Externe Validierung**

Mittels 3980 Patienten wurden die Vorhersagemodelle auf ihre jeweilige prädiktive Genauigkeit hin extern überprüft. Dazu wurden im speziellen die jeweiligen Patientendaten für jedes einzelne Vorhersagemodell gemäß den Originalpublikationen in das Vorhersagemodell "eingespeist". Die jeweiligen Punktwerte und resultierenden Wahrscheinlichkeiten wurden graphisch exploriert. Die hieraus resultierende externe Validierung und graphische Darstellung wurde mittels S-PLUS Professional, version 1 (MathSoft Inc., Seattle, Washington) errechnet.

Der Vorteil dieser externen Validierung liegt in der gleichzeitigen Ermittlung von ROC und Kalibrierung. Hieraus folgt die prädiktive Genauigkeit. Ein Maß für die Diskriminierungsfähigkeit eines Vorhersagemodells des Tests stellt die Fläche unter der ROC-Kurve (AUC- Area under the curve) dar. Die Fläche kann Werte zwischen 0,5 und 1 annehmen, wobei ein höherer Wert die bessere Güte anzeigt. In der Regel wird eine  $AUC \geq 0,70$  als Zeichen prädiktiver Genauigkeit gewertet. Die Vorhersagegenauigkeit der einzelnen Modelle verglichen wir mittels des Mantel-Haenszel-Tests. Abschließend stellten wir die Ergebnisse graphisch dar, indem wir jeweils die vorhergesagte Möglichkeit gegenüber der tatsächlich beobachteten Anzahl an PCa in den Biopsaten darstellten. Alle Tests wiesen ein Signifikanzniveau von  $p < 0,05$  auf.

#### **2.4.2 Kalibrierungskurven**

Für die verschiedenen Erstbiopsie-Vorhersagemodelle wurden jeweils Kalibrierungsdiagramme angefertigt. Hier sind die vorhergesagten Wahrscheinlichkeiten des jeweiligen Modelles für ein PCa gegen die beobachteten Zahlen an tatsächlich entdeckten PCa gegeneinander aufgetragen. Dabei stellt die x-Achse die für das jeweilige Modell errechnete Wahrscheinlichkeit (in %) einer PCa-Diagnose dar. Die y-Achse stellt die tatsächlich beobachteten Diagnosen dar. Die 45°-Linie entspricht dabei dem Zustand perfekter Vorhersagen. Hier fallen vorhergesagte Wahrscheinlichkeiten mit der tatsächlich beobachteten Zahl an PCa zusammen. Alle Punkte des Graphen, die sich oberhalb der 45°-Linie befinden, zeigen eine Unterschätzung der Anzahl an PCa-Diagnosen durch die Modelle an, wohingegen Punkte unterhalb der 45°-Linie eine Überschätzung darstellen. Innerhalb des Graphen werden bestimmte Bereiche voneinander abgegrenzt. Zwischen 0-20% befindet sich der „low risk“-Bereich, zwischen 60 und 80% der „intermediate risk“-Bereich und zwischen 80 und 100% der „high risk“-Bereich.

### **3 Ergebnisse**

#### **3.1 Deskriptive Analysen**

Tabelle 2 zeigt die klinischen Patientencharakteristika der 3980 Männer, welche für unsere Arbeit zur externen Validierung der vier Erstbiopsie-Vorhersagemodelle verwendet wurden.

Die Altersspannbreite der Patientenkohorte reichte von 34 bis 84 Jahre (Altersdurchschnitt: 64,8 Jahre, Median: 65). Die Werte für das PSA reichten von 2 bis 20 ng/ml (Durchschnittswert: 7,2, Median: 6,2) und für das %fPSA von 1,1-91,2 (Durchschnittswert: 19,2, Median: 17,7). Das Gesamt-Prostatavolumen lag zwischen 12 und 150 cm<sup>3</sup> (Durchschnittswert: 62,5, Median: 58). Von allen Patienten wiesen 206 Männer (5,2%) bei der digitalen rektalen Untersuchung einen suspekten Befund auf. Bei 169 (82%) von diesen Männern mit positivem Tastbefund konnte bei der Prostata-Erstbiopsie ein PCa nachgewiesen werden. Insgesamt wurde bei 1688 der 3980 Patienten ein PCa entdeckt (42,2%). Die Anzahl der gewonnenen Biopsiezylinder reichte von minimal 6 bis zu maximal 24 Biopsaten (Durchschnittswert: 8,7).

<b>Variable</b>	<b>N (%)</b>
<b>Patientenanzahl</b>	3980 (100%)
<b>Alter (Jahre)</b>	
<b>Durchschnittsalter</b>	64,8
<b>Median</b>	65
<b>Bereich</b>	34-85
<b>PSA (ng/ml)</b>	
<b>Durchschnittswert</b>	7,2
<b>Median</b>	6,2
<b>Bereich</b>	2,0-20,0
<b>%fPSA</b>	
<b>Durchschnittswert</b>	19,2
<b>Median</b>	17,7
<b>Bereich</b>	1,1-91,2
<b>Suspekte DRU (%)</b>	206 (5,2%)
<b>Prostatagesamtvolumen(cm<sup>3</sup>)</b>	
<b>Durchschnittswert</b>	62,5
<b>Median</b>	58,0
<b>Bereich</b>	12-150
<b>Anzahl der Biopsien</b>	
<b>Durchschnitt</b>	8,7
<b>Median</b>	10,0
<b>Bereich</b>	6-24
<b>PCa durch Biopsie erkannt (%)</b>	1688 (42,4%)
<b>DRU: Digital rektale Untersuchung; PSA:Prostata-spezifisches Antigen; %fPSA: prozentualer Anteil an freiem PSA</b>	

**Tabelle 2: Grundlegende Charakteristika der 3980 Patienten, die einer Prostata-Erstbiopsie unterzogen wurden und die Studienkohorte bilden**

In Tabelle 3 sind die Ergebnisse der externen Validierung sowie die jeweils ursprünglich publizierten Ergebnisse der untersuchten Erstbiopsie-Vorhersagemodelle dargestellt. Insgesamt zeigten alle vier Modelle in der externen Validierung einen Abfall ihrer prädiktiven Genauigkeit im Vergleich zu ihren Originalpublikationen. Die genaueste Vorhersage für eine positive Prostata-Erstbiopsie konnte das Karakiewicz-Nomogramm (Karakiewicz, Benayoun et al. 2005) mit einer prädiktiven Genauigkeit von 71% liefern. Dahinter folgt das Garzotto-Nomogramm (Garzotto, Hudson et al. 2003), dessen prädiktive Genauigkeit nach externer

Validierung 70% betrug. Die CART-Analyse (Garzotto, Beer et al. 2005) kommt auf 68% prädiktive Genauigkeit und das ANN (Stephan, Cammann et al. 2002) auf 67%. Die prädiktive Genauigkeit des Artifiziiellen neuronalen Netzwerkes (Stephan, Cammann et al. 2002), welches die Vorhersageparameter Alter, DRU, PSA (2-20ng/ml), %fPSA und Prostatagesamtvolumen beinhaltet, wurde von seinen Autoren ursprünglich mit 78% angegeben. Nach der Überprüfung mittels unserer externen Patientenkohorte lag die dabei ermittelte prädiktive Genauigkeit bei 67%. Somit betrug der Abfall in der Vorhersagegenauigkeit 11%. Für das Nomogramm von Karakiewicz (Karakiewicz, Benayoun et al. 2005), welches auf den Parametern Alter, DRU, PSA (0,1-50 ng/ml) und %fPSA basiert, wurde ursprünglich ebenfalls eine prädiktive Genauigkeit von 78% angegeben. Gemäß unserer externen Validierung musste diese von uns auf 71% abgesenkt werden. Die Differenz betrug somit 7%. Das Nomogramm von Garzotto et al. (Garzotto, Hudson et al. 2003) mit den Vorhersageparametern Alter, DRU, TRUS, PSA (<10 ng/ml) und PSA-Dichte konnte nach der Anwendung auf unseren externen Datensatz noch eine prädiktive Genauigkeit von 70% aufweisen. Die Differenz lag hier bei 3%, bei einer ursprünglich angegebenen prädiktiven Genauigkeit von 73%. Die CART-Analyse von Garzotto et al. (Garzotto, Beer et al. 2005) beinhaltete zur Vorhersage die Parameter Alter, PSA ( $\leq 10$  ng/ml), PSA-Dichte, TRUS und Prostatagesamtvolumen. Die prädiktive Genauigkeit wurde von den Autoren mit 73% ausgewiesen und lag nach der externen Validierung bei 68%.

	QUELLEN	ART DES VORHERSAGEMODELLS	PARAMETER ZUR VORHERSAGE EINES PROSTATAKARZINOMS IN DER ERSTBIOPSIE	PATIENTENANZAHL	ANZAHL DER BIOPSIEN	VALIDIERUNG	PRÄDIKTIVE GENAUIGKEIT (%)
Aktuelle externe Validierung	Karakiewicz et al.(Karakiewicz , Benayoun et al. 2005)	Nomogramm	Alter, DRU, PSA (0,1-50 ng/ml), %fPSA	3980	6+	Extern: Direktvergleich	71
	Garzotto et al.(Garzotto, Hudson et al. 2003)	Nomogramm	Alter, DRU, TRUS, PSA<10ng/ml, PSA-Dichte				70
	Garzotto et al.(Garzotto, Beer et al. 2005)	Classification and Regression Tree	Alter, PSA (≤10 ng/ml), PSA-Dichte, TRUS, Prostatagesamtvolumen				68
	Stephan et al.(Stephan, Cammann et al. 2002)	Artifizielles neuronales Netzwerk	Alter, DRU, PSA (2-20ng/ml), %fPSA, Prostatagesamtvolumen				67
Ursprünglich berichtete Version	Karakiewicz et al.(Karakiewicz , Benayoun et al. 2005)	Nomogramm	Alter, DRU, PSA (0,1-50 ng/ml), %fPSA	1762	6	Extern.	78
	Garzotto et al.(Garzotto, Hudson et al. 2003)	Nomogramm	Alter, DRU, TRUS, PSA<10ng/ml, PSA-Dichte	1237	6+	Split sample	73
	Garzotto et al.(Garzotto, Beer et al. 2005)	Classification and Regression Tree	Alter, PSA (≤10 ng/ml), PSA-Dichte, TRUS, Prostatagesamtvolumen	1433	6	Split sample	73
	Stephan et al.(Stephan, Cammann et al. 2002)	Artifizielles neuronales Netzwerk	Alter, DRU, PSA (2-20ng/ml), %fPSA, Prostatagesamtvolumen	1188	6-8	Split sample	78

**DRU: Digitale rektale Untersuchung, PSA: Prostata-spezifisches Antigen, %fPSA: Anteil an freiem Prostata-spezifischen Antigen in Prozent, TRUS: Transrektaler Ultraschall**

**Tabelle 3: Vergleich der Ergebnisse von externer Überprüfung und ursprünglich angegebenen Werten der Vorhersagemodelle**

In den Diagrammen 1A-D sind die Kalibrierungskurven für das jeweilige Erstbiopsie-Vorhersagemodell dargestellt. Wie bereits in der Methodik näher erläutert, sind hier die vorhergesagten Wahrscheinlichkeiten gegen die beobachteten Zahlen an tatsächlich entdeckten PCa gegeneinander aufgetragen. Die Diagramme zeigen durchgehend eine Unterschätzung der Anzahl an PCa-Diagnosen an. Die Ausnahme bildet das Karakiewicz-Nomogramm (Karakiewicz, Benayoun et al. 2005) mit einem nahezu parallel zur Idealkurve verlaufenden Graphen.

### **3.2 Direktvergleich der einzelnen Erstbiopsie-Vorhersagemodelle**

Tabelle 4 zeigt den Direktvergleich der prädiktiven Genauigkeiten der unterschiedlichen Erstbiopsie-Vorhersagemodelle. Hier fiel auf, dass auf multivariablen logistischen Regressionsanalysen beruhende Nomogramme (Garzotto, Hudson et al. 2003; Karakiewicz, Benayoun et al. 2005) verglichen mit der CART-Analyse (Garzotto, Beer et al. 2005) und dem ANN (Stephan, Cammann et al. 2002) am besten abschnitten.

Das Karakiewicz-Nomogramm (Karakiewicz, Benayoun et al. 2005) war im direkten Vergleich der CART-Analyse von Garzotto (Garzotto, Beer et al. 2005) signifikant ( $p=0,004$ ) und dem ANN (Stephan, Cammann et al. 2002) hochsignifikant überlegen ( $p=0,0001$ ). Im Direktvergleich zum Garzotto-Nomogramm (Garzotto, Hudson et al. 2003) konnte hingegen kein signifikanter Unterschied nachgewiesen werden ( $p=0,3$ ).

Das Garzotto-Nomogramm (Garzotto, Hudson et al. 2003) wies nur im direkten Vergleich zum ANN (Stephan, Cammann et al. 2002) eine signifikante Überlegenheit auf ( $p=0,004$ ).

Festzustellen ist, dass das ANN (Stephan, Cammann et al. 2002) am schlechtesten abschnitt. Es wies gegenüber keinem der übrigen Erstbiopsie-Vorhersagemodelle signifikante Überlegenheit auf.

Direktvergleich	Prädiktive Genauigkeit	p-Werte
Karakiewicz-Nomogramm vs. ANN	71% vs. 67%	0.0001
Karakiewicz-Nomogramm vs. CART	71% vs. 68%	0.004
Karakiewicz-Nomogramm vs. Garzotto-Nomogramm	71% vs. 70%	0,3
Garzotto-Nomogramm vs. ANN	70% vs. 67%	0,004
Garzotto-Nomogramm vs. CART	70% vs. 68%	0,05
CART vs. ANN	68% vs. 67%	0,4

**Tabelle 4: Direktvergleich der prädiktiven Genauigkeiten der einzelnen Erstbiopsie-Modelle mittels des Mantel-Haenszel-Test**

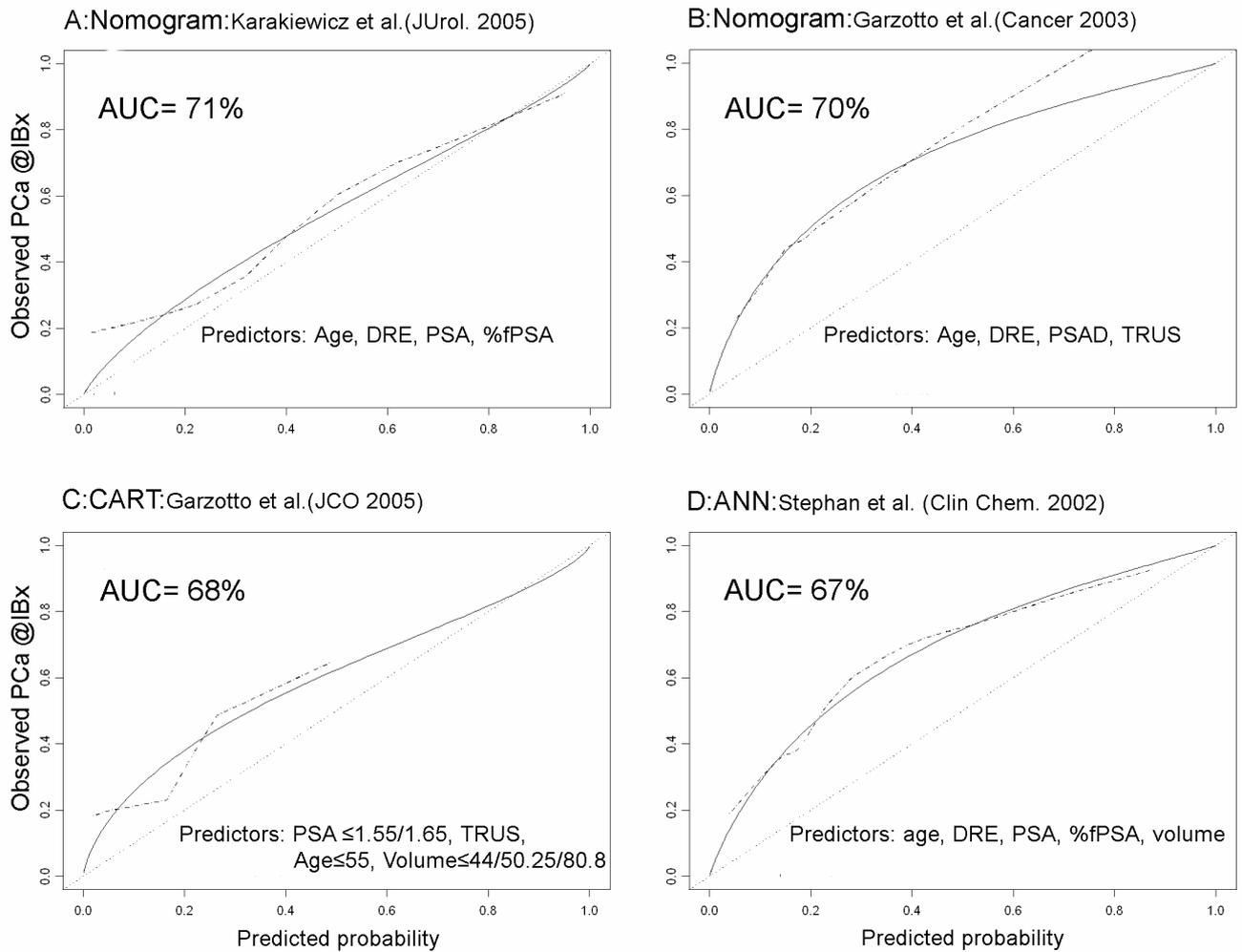
Beim Vergleich der Kalibrierungskurven für die verschiedenen Erstbiopsie-Vorhersagemodelle in den Diagrammen der Abbildung 8A-D zeigt sich ebenfalls die Überlegenheit der Nomogramme (Garzotto, Hudson et al. 2003; Karakiewicz, Benayoun et al. 2005) gegenüber CART-Analyse (Garzotto, Beer et al. 2005) und ANN (Stephan, Cammann et al. 2002). Aus dem Diagramm 8D geht hervor, dass das ANN in seiner prädiktiven Genauigkeit stark vom Ideal der perfekten Vorhersage abweicht und in erheblichem Maße das Risiko eines PCa in der Biopsie unterschätzt. Diese Unterschätzung zieht sich durch alle Vorhersagebereiche, gilt also sowohl für Niedrigrisiko- als auch für Hochrisiko-Patienten. Am stärksten ist diese Abweichung jedoch im Vorhersagebereich zwischen 10-70% ausgeprägt. Wenn das ANN beispielsweise das Risiko des Vorhandenseins eines PCa mit 40% angibt, liegt die Anzahl der tatsächlich entdeckten Karzinome bei 60%. Die AUC liegt bei 67%.

Einen ähnlichen graphischen Verlauf können wir im Diagramm 8B für das Garzotto-Nomogramm beobachten. Auch hier findet sich eine alle Bereiche umfassende starke Risiko-Unterschätzung für das Vorhandensein eines PCa, besonders ausgeprägt

zwischen 10 und 70%. Lediglich im Hochrisikobereich zwischen 80 und 100% ist dieses Modell dem ANN überlegen. Die AUC liegt bei 70%.

Im Diagramm 8C wird deutlich, dass die CART-Analyse (Garzotto, Beer et al. 2005) das Vorhandensein von Prostatakrebs in der Biopsie insgesamt eher unterschätzt, vor allem in dem Bereich zwischen 2 und 40%. Im Hochrisikobereich, in dem das Modell die Wahrscheinlichkeit des Vorhandenseins eines PCa mit 80-100% angibt, werden hingegen annähernd perfekte Vorhersagen getroffen. Die AUC liegt bei 68%.

Das Diagramm 8A zum Nomogramm von Karakiewicz (Karakiewicz, Benayoun et al. 2005) weist nahezu perfekte Übereinstimmung zwischen Vorhersage und tatsächlichem Vorhandensein eines PCa in der Biopsie auf. Die vorhergesagte Wahrscheinlichkeit eines PCa verläuft parallel in unmittelbarer Nähe zur 45°-Linie, die innerhalb des Graphs perfekte Vorhersage darstellt. Im Hochrisiko-Bereich, in dem das Vorhandensein eines PCa mit einer Wahrscheinlichkeit von 80-100% angegeben wird, verschmelzen die Kurven bezüglich Vorhersage und tatsächlich entdeckten PCa vollständig innerhalb der 45°-Linie miteinander. Die Vorhersagen sind innerhalb dieses Bereiches noch exakter als es die CART-Analyse von Garzotto (Garzotto, Beer et al. 2005) zu leisten vermag. Die AUC liegt bei 71%.



**Abbildung 8: Kalibrierungskurven der externen Validierung der etablierten Prostataerstbiopsie Vorhersagemodelle.**

## 4 Diskussion

Nomogramme stellen etablierte Werkzeuge dar, denn sie helfen dem Kliniker bei der Risikoeinschätzung von Patienten, die sich einer Prostata-Erstbiopsie unterziehen. Aufgrund der Vielfalt von etablierten, statistischen Erstbiopsie-Vorhersagemodellen herrscht unter Klinikern jedoch ein gewisser Grad an Verunsicherung, welches das am besten geeignete Modell für den jeweiligen Patienten darstellt. Dieser Sachverhalt stellte die Grundlage unserer Arbeitshypothese dar. Um herauszufinden, welches statistische Prostata-Erstbiopsie-Modell den größten klinischen Nutzen hat, führten wir eine externe Validierung von vier verschiedenen Vorhersagemodellen durch: darunter zwei Nomogramme (Garzotto, Hudson et al. 2003; Karakiewicz, Benayoun et al. 2005) ein CART-Modell (Garzotto, Beer et al. 2005) sowie ein ANN (Stephan, Cammann et al. 2002). Der jeweilige klinische Nutzen wurde anhand der prädiktiven Genauigkeit mittels einer großen externen Patientenkohorte aus der Abteilung Urologie des Universitätsklinikums Hamburg-Eppendorf untersucht und untereinander direkt verglichen. Zusätzlich wurden Vergleiche der sogenannten Kalibrierungen angestellt. Die untersuchten Erstbiopsie-Vorhersagemodelle sind ursprünglich auf der Grundlage von 6-8 Biopsien entstanden. Der heutige Goldstandard des Biopsieschemas umfasst dagegen acht oder mehr Biopsiezylinder. Es besteht Uneinigkeit darüber, ob aktuelle, mindestens auf Oktantenbiopsien basierende Patientendaten auf Erstbiopsie-Vorhersagemodelle angewandt werden dürfen, die nicht dem heutigen Biopsie-Goldstandard entsprechen. In der vorliegenden Arbeit haben wir dies überprüft.

Das Nomogramm von Karakiewicz et al. (Karakiewicz, Benayoun et al. 2005) basierte auf den vier Risikoparametern Alter, DRU, PSA und %fPSA und seine ursprünglich angegebene prädiktive Genauigkeit betrug 78%, innerhalb eines PSA-Spektrums von 0,1-50 ng/ml. Das Nomogramm von Garzotto (Garzotto, Hudson et al. 2003) beinhaltete die Prädiktoren Alter, DRU, TRUS, PSA <10 ng/ml und PSA-Dichte. Die angegebene

prädiktive Genauigkeit lag bei 73%. Die CART-Analyse von Garzotto (Garzotto, Beer et al. 2005) umfasst fünf Parameter (Alter, DRU, TRUS, PSA ( $\leq 10$  ng/ml), PSA-Dichte und Prostatagesamtvolumen) – die prädiktive Genauigkeit lag initial bei 73%. Das ANN der Autoren Stephan et al. (Stephan, Cammann et al. 2002) beinhaltete schließlich neben den bereits genannten Parametern noch das Prostatagesamtvolumen und wurde mit einer prädiktiven Genauigkeit von ebenfalls 78% ausgewiesen. Das PSA-Spektrum reichte dabei von 2 bis 20 ng/ml.

Die Ergebnisse nach unseren Untersuchungen zeigten, dass auf logistischen Regressionsanalysen beruhende Nomogramme [(Karakiewicz, Benayoun et al. 2005): prädiktive Genauigkeit 71% und (Garzotto, Hudson et al. 2003): prädiktive Genauigkeit 70%] sowohl CART-Analysen (Garzotto, Beer et al. 2005) (prädiktive Genauigkeit 68%) als auch ANNs (Stephan, Cammann et al. 2002) gegenüber (prädiktive Genauigkeit 67%) in der Vorhersage eines PCa in der Erstbiopsie signifikant überlegen sind. Diese Tatsache spiegelte sich auch in den Kalibrierungskurven wieder. Bei diesen zeigte das Karakiewicz-Nomogramm (Karakiewicz, Benayoun et al. 2005) einen nahezu parallelen Verlauf zur 45°-Linie, die perfekter Vorhersage einer PCa-Diagnose entspricht. Für alle übrigen Nomogramme (Stephan, Cammann et al. 2002; Garzotto, Hudson et al. 2003; Garzotto, Beer et al. 2005) zeigten die Kalibrierungskurven hingegen eine starke Risiko-Unterschätzung für das Vorhandensein eines PCa.

<b>ERSTBIOPSIEVORHERSAGE-MODELL</b>	<b>PRÄDIKTIVE GENAUIGKEIT (%)</b>
<b>Nomogramm Karakiewicz et al.(Karakiewicz, Benayoun et al. 2005)</b>	71
<b>Nomogramm Garzotto et al.(Garzotto, Hudson et al. 2003)</b>	70
<b>CART Garzotto et al.(Garzotto, Beer et al. 2005)</b>	68
<b>ANN Stephan et al.(Stephan, Cammann et al. 2002)</b>	67

**Tabelle 5: Prädiktive Genauigkeiten der einzelnen Erstbiopsievorhersage Modelle in der externen Validierung (n=3980).**

Der maximale Vorhersageunterschied zwischen diesen vier Modellen beträgt 4% (resultiert aus dem Unterschied von 67% beim ANN und 71% beim Nomogramm nach Karakiewicz). Im konkreten Fall bedeutet es, dass von 1000 Patienten 40 mehr in Bezug auf das Vorhandensein von Prostatakrebs in der Erstbiopsie richtig beurteilt werden, wenn der Nutzer das Nomogramm (Karakiewicz, Benayoun et al. 2005) anstelle des ANN (Stephan, Cammann et al. 2002) nutzt. Diese Zahl gewinnt noch an Bedeutung, wenn man sich die aktuelle epidemiologische Entwicklung vergegenwärtigt. Zusätzlich konnte durch die größte urologische Studie (Thompson, Pauler et al. 2004) gezeigt werden, dass der Nutzen von PSA-Schwellenwerten limitiert ist. Hieraus ist abzuleiten, dass in der Zukunft die Zahl der Patienten, die einer Prostata-Erstbiopsie unterzogen werden, ansteigt. Somit werden validierte Risikostratifizierungsmodelle in der Zukunft an Bedeutung gewinnen. Die Ergebnisse unserer Arbeit sind demnach sowohl für Kliniker als auch für Patienten von größter Bedeutung, da es im Interesse beider sein muss, möglichst genaue Risikoabschätzungen zu treffen, um so dem Patienten eine bestmögliche Einschätzung seines Risikos für einen Prostatakrebs geben zu können.

Als weiteres wichtiges Ergebnis kann festgehalten werden, dass alle vier Erstbiopsie-Modelle im Vergleich zu ihren Originalpublikationen innerhalb der externen Validierung einen Abfall in der prädiktiven Genauigkeit zeigten. Dieses Phänomen ist nicht überraschend (Carroll 2000; Graefen, Karakiewicz et al. 2002; Graefen, Karakiewicz et al. 2002; Kattan 2002; Kattan 2006). konnten ähnliche Abfälle der prädiktiven Genauigkeit beobachten. Insgesamt sind die in der externen Validierung ermittelten prädiktiven Genauigkeiten um den Wert 70% herum angesiedelt. Dieser Wert ist insgesamt akzeptabel und deutet an, dass sich die Modelle durchaus zur Risikostratifizierung für aktuelle Patienten eignen. Überraschenderweise zeigte sich der größte Abfall der prädiktiven Genauigkeit für das ANN (Stephan, Cammann et al. 2002), obwohl die externen Patientendaten für den PSA-Bereich 2-20 ng/ml anhand dieses Modells adaptiert wurden. In der externen Validierung war ein Abfall von 11% nachzuweisen. Die geringste Diskrepanz bestand in der prädiktiven Genauigkeit des Garzotto-Nomogrammes (Garzotto, Hudson et al. 2003). Hier wurde ursprünglich ein Genauigkeitswert von 73% angegeben, dem ein Wert von 70% in der externen Validierung gegenüber steht. Im Karakiewicz-Nomogramm (Karakiewicz, Benayoun et al. 2005) betrug die Differenz 7% und bei der CART-Analyse (Garzotto, Beer et al. 2005) 5%. Alle Modelle sagten also weniger korrekt das Vorhandensein eines PCa voraus, als sie es nach Angaben ihrer Originalarbeit sollten.

Für diesen Umstand mag die Tatsache verantwortlich sein, dass die verschiedenen Modelle in der Mehrheit anhand der Ergebnisse von Sextantenbiopsien entwickelt wurden, die jeweils an den Instituten durchgeführt wurden, an denen die Autoren zur Zeit der Entwicklung tätig waren (siehe Tabelle 3). In unserer Untersuchung wurden alle Modelle anhand einer externen Kohorte von 3980 Patienten validiert, die mindestens eine Sextantenbiopsie erhalten hatten. Mehrere Arbeiten konnten zeigen, dass durch ein erweitertes systematisches Biopsieschema ( $\geq 10$  Biopsiezylinder)

zwischen 10-20% mehr PCa entdeckt werden können (Presti, Chang et al. 2000; Matlaga, Eskew et al. 2003). Andere Arbeitsgruppen äußerten sich ebenfalls dahingehend, dass Nomogramme, die in der Sextantenbiopsie-Ära entwickelt wurden, in der heutigen Ära der Mehrfachbiopsien nicht mehr so akkurat vorhersagen können (Chun, Briganti et al. 2006; Walz, Graefen et al. 2006). Unsere Ergebnisse bestätigen diese Annahmen. Eine 10 oder >10-fache Prostata-Erstbiopsie wurde bei 2204 von 3980 Patienten (55,4%) durchgeführt. Somit wurde mehr als die Hälfte dieser externen Patientenkohorte dem aktuellen Goldstandard einer ausgedehnten Prostata-Erstbiopsie unterzogen.

Ein weiterer Punkt, der die Abweichungen in der prädiktiven Genauigkeit zwischen ursprünglich berichteten Versionen und der aktuellen externen Validierung der Modelle erklären mag, besteht in der Tatsache, dass bei der externen Validierung alle Modelle anhand eines gleichen Patientensatzes verglichen wurden. Entwickelt wurden dagegen alle Vorhersagemodelle an unterschiedlichen Patienten. Dieser Umstand bezieht sich sowohl auf die Patientencharakteristika als auch die Patientenanzahl.

Ein genauer Blick auf das Alter der Patienten in den ursprünglichen Patientenkollektiven zeigt z.B., dass beim ANN und der CART-Analyse die Patienten im Mittel 65 Jahre alt waren – dies ist auch der Mittelwert der aktuellen, auf alle Modelle angewandten Patientenkohorte. Das Garzotto-Nomogramm beinhaltete hingegen einen erheblichen Anteil jüngerer Patienten (40,7% waren hier unter 65 Jahre).

Auch im Bereich des tPSA unterscheiden sich die ursprünglichen Versionen. Das Karakiewicz-Nomogramm (Karakiewicz, Benayoun et al. 2005) beinhaltet tPSA-Werte zwischen 0,1-50 ng/ml, wohingegen das Patientenkollektiv in den Garzotto-Modellen (Garzotto, Hudson et al. 2003; Garzotto, Beer et al. 2005) auf tPSA-Werte  $\leq 10$  ng/ml beschränkt war. Für das ANN (Stephan, Cammann et al. 2002) wurden schließlich

tPSA-Werte zwischen 2-20ng/ml gefordert. In diesem Bereich zwischen 2 und 20ng/ml lagen auch die tPSA-Werte der 3980 Patienten, mit deren Hilfe die verschiedenen Modelle extern verglichen wurden. Interessanterweise zeigte das ANN (Stephan, Cammann et al. 2002), obwohl die Eigenschaften des Patientenkollektivs an ihm angelegten waren, dennoch die niedrigste prädiktive Genauigkeit.

Ein weiterer Erklärungsansatz für die unterschiedlich ermittelten prädiktiven Genauigkeiten der einzelnen Modelle zwischen Originalpublikation und externer Validierung mag in der Verwendung unterschiedlicher PSA-Assays zur Bestimmung der PSA-Werte liegen. So verwendete Garzotto für die PSA-Bestimmungen, die seinen Risikovorhersagemodellen (Garzotto, Hudson et al. 2003; Garzotto, Beer et al. 2005) zu Grunde liegen, ein Assay von Abbott Diagnostics, Abbot Park IL. Dieses Assay wurde auch für die Ermittlung der PSA-Werte in der externen Patientenkohorte verwendet, mittels derer wir die verschiedenen Modelle auf ihre prädiktive Genauigkeit verglichen. Für das ANN (Stephan, Cammann et al. 2002) wurde dagegen der IMMULITE PSA und IMMULITE Free PSA Assay benutzt. Jeder Hersteller verwendet dabei individuelle Antikörper, die ihrerseits schon unterschiedliche Ergebnisse zur Folge haben können (Semjonow, Brandt et al. 1996).

Auch für die DRU, einen Parameter aller Modelle zur Risikovorhersage eines PCa, lassen sich bezüglich der Patientencharakteristika Unterschiede feststellen, die möglicherweise Auswirkungen auf die prädiktive Genauigkeit haben. So lag der Anteil an Patienten, die eine suspekte DRU aufwiesen, innerhalb der zur externen Validierung eingesetzten Kohorte mit 206 Patienten (5,2%) recht niedrig. Im Vergleich zu den Patientenkollektiven, mittels derer die Risikovorhersagemodelle ursprünglich entwickelt wurden, lassen sich diesbezüglich klare Unterschiede feststellen. Der Anteil an Patienten mit suspekter DRU lag innerhalb des Krakiewicz-Nomogrammes (Karakiewicz, Benayoun et al. 2005) bei 37,2%, innerhalb des Garzotto-Nomogrammes

(Garzotto, Hudson et al. 2003) bei 41,9% und innerhalb der CART-Analyse (Garzotto, Beer et al. 2005) bei 41,7%. Allgemein lässt sich anmerken, dass die Ergebnisse der DRU, die in den verschiedenen Risikovorhersagemodellen mit normal/ suspekt angegeben werden vom jeweiligen Untersucher abhängig sind und es auf diesem Wege zu Unterschieden in der Beurteilung kommen kann.

Zusammenfassend kann man sagen, dass all diese beschriebenen Faktoren möglicherweise zum Umstand beigetragen haben, dass die prädiktiven Genauigkeiten der einzelnen verglichenen Modelle bei der Validierung durch eine externe Patientenkohorte niedriger lag als es von ihren Autoren ursprünglich angegeben wurde.

Erstbiopsie-Vorhersagemodelle sollten nicht zu komplex sein, um eine Anwendung im klinischen Alltag zu gewährleisten. Nun mag man denken, dass die prädiktive Genauigkeit der Erstbiopsie-Vorhersagemodelle mit der Zahl der Risikoparameter ansteigt. Unsere Ergebnisse haben genau das Gegenteil gezeigt. Das Karakiewicz-Nomogramm (Karakiewicz, Benayoun et al. 2005) mit der geringsten Anzahl von vier Vorhersageparametern (Alter, DRU, PSA: 0,1-50 ng/ml, %fPSA) sagt mit einer prädiktiven Genauigkeit von 71% das Risiko einer PCa-Diagnose am besten voraus. Erstbiopsie-Vorhersagemodelle müssen also nicht komplex sein und ein einfaches, nicht adaptiertes Modell (Karakiewicz, Benayoun et al. 2005) hat überraschenderweise die besten Ergebnisse erzielt.

Darüber hinaus sollten Vorhersagemodelle möglichst unabhängig von Faktoren wie z.B. verwendetem PSA-Assay oder dem PSA-Spektrum sein und dabei eine möglichst hohe prädiktive Genauigkeit besitzen (Chun, Karakiewicz et al. 2006). Nur so sind sie auf viele verschiedene Patientengruppen anwendbar. Unsere Ergebnisse zeigen daher auch auf, wie wichtig eine externe Validierung ist, um die Anwendbarkeit auf möglichst viele verschiedene Patientengruppen zu gewährleisten.

Neben der Komplexität ist die Anwendbarkeit eines Erstbiopsie-Vorhersagemodells von entscheidender Bedeutung im klinischen Alltag. So ist das Karakiewicz-Nomogramm (Karakiewicz, Benayoun et al. 2005) mit der besten prädiktiven Genauigkeit für den Kliniker auch einfach zu handhaben, da es neben einer Internetversion ([www.martini-klinik.de](http://www.martini-klinik.de)) auch in Papierform zur Verfügung steht. Ebenso kann das Garzotto-Nomogramm (Garzotto, Hudson et al. 2003) und die CART-Analyse (Garzotto, Beer et al. 2005) in Papierform angewendet werden. Möchte der Kliniker hingegen das ANN (Stephan, Cammann et al. 2002) für seine Patienten verwenden, ist er auf einen Computer mit Internetzugang angewiesen.

Unsere Ergebnisse haben gezeigt, dass die prädiktive Genauigkeit eines auf logistischen Regressionsanalysen beruhenden Nomogrammes weder von CART-Analysen noch von einem ANN übertroffen werden kann. Diese Tatsache wird durch Diagramme (Diagramme in den Abbildungen 6A-D), in denen die angegebenen Vorhersagewerte gegenüber den tatsächlich entdeckten PCa-Zahlen in den verschiedenen Modellen aufgetragen sind, noch verdeutlicht. Nach der aktuellen Datenlage ist dies der bisher einzige Erstbiopsie-Direktvergleich zwischen drei verschiedenen statistischen Risikovorhersagemodellen.

Den durchgeführten Vergleich der verschiedenen Erstbiopsie-Vorhersagemodelle hat möglicherweise die Tatsache ungünstig beeinflusst, dass das Karakiewicz-Nomogramm (Karakiewicz, Benayoun et al. 2005), welches die beste prädiktive Genauigkeit aufweist, ursprünglich anhand eines Patientenkollektivs der urologischen Abteilung des Universitätsklinikums Hamburg-Eppendorf entwickelt wurde. Für die externe Validierung aller Erstbiopsie-Vorhersagemodelle wurden ebenfalls Patienten der Universitätsklinik Hamburg-Eppendorf benutzt. Die Patienten, die zur Entwicklung des Modelles verwendet wurden, waren dabei jedoch andere als diejenigen, die für die externe Validierung herangezogen wurden. Limitierend könnte sich auch auswirken,

dass unsere externe Patientenkohorte aus einer Zuweiserpopulation besteht. Schon die hohe Rate der entdeckten (42,2%) weist darauf hin. Inwieweit die Erstbiopsie-Vorhersagemodelle bei einer PSA-Screeningkohorte ähnliche Ergebnisse vorhersagen würden, ist unklar.

## 5 Zusammenfassung

In der vorliegenden Arbeit wurde erstmalig ein Direktvergleich zwischen verschiedenen statistischen Erstbiopsie-Modellen zur Vorhersage des PCa-Risikos anhand einer großen externen Patientenkohorte durchgeführt. Es handelte sich dabei um zwei Nomogramme (Garzotto, Hudson et al. 2003; Karakiewicz, Benayoun et al. 2005), eine CART-Analyse (Garzotto, Beer et al. 2005) und ein ANN (Stephan, Cammann et al. 2002).

Es zeigte sich, dass Erstbiopsie-Vorhersagemodelle, die anhand einer Sextantenbiopsie entwickelt wurden, in der Ära der 8-, 10- oder noch umfassenderen Biopsieschemata nicht mehr so korrekt vorhersagen wie früher und in ihrer prädiktiven Genauigkeit abfallen.

Zusätzlich konnte gezeigt werden, dass Nomogramme innerhalb der verschiedenen statistischen Erstbiopsie-Vorhersagemodelle das Risiko einer PCa-Diagnose am zuverlässigsten vorhersagen und darin CART-Analysen sowie ANN überlegen sind. Somit objektivieren unsere Ergebnisse den Stellenwert der viel untersuchten statistischen Ansätze für die Prostataerstbiopsie.

Bisher herrschte unter Klinikern ein großes Maß an Verunsicherung, wenn es um die Frage ging, welches etablierte Erstbiopsie-Vorhersagemodell aus der Vielzahl der zur Verfügung stehenden zu benutzen sei. Unsere Ergebnisse bedeuten für die Kliniker, dass sie auf der Grundlage unserer Untersuchungen beruhigt und sicher das für sie und ihre Patienten am besten geeignete Erstbiopsie-Vorhersagemodell auswählen können.

## 6 Literaturverzeichnis

- Augustin, H., P. Hammerer, et al. (2003). "Intraoperative and perioperative morbidity of contemporary radical retropubic prostatectomy in a consecutive series of 1243 patients: results of a single center between 1999 and 2002." Eur Urol **43**(2): 113-8.
- Bianco, F. J., Jr. (2006). "Nomograms and medicine." Eur Urol **50**(5): 884-6.
- Bostwick, D. G. (1994). "Gleason grading of prostatic needle biopsies. Correlation with grade in 316 matched prostatectomies." Am J Surg Pathol **18**(8): 796-803.
- Bradley, E., Tibshirani, R.J. (1993). Monographs on statistics and applied probability: An introduction to the bootstrap., Champman and Hall, CRC,p.275.
- Breiman, L., Friedman, J.H., Olshen. R.A., Stone, C.J. (1984). Classification and Regression Trees, Belmont CA, Wodsworth.
- Byar, D. P. and F. K. Mostofi (1972). "Carcinoma of the prostate: prognostic evaluation of certain pathologic features in 208 radical prostatectomies. Examined by the step-section technique." Cancer **30**(1): 5-13.
- Cam, K., S. Yucel, et al. (2002). "Accuracy of transrectal ultrasound guided prostate biopsy: histopathological correlation to matched prostatectomy specimens." Int J Urol **9**(5): 257-60.
- Carroll, P. R. (2000). "Prostate biopsy: a wealth of information when done and interpreted correctly." J Clin Oncol **18**(6): 1161-3.
- Catalona, W. J., J. P. Richie, et al. (1994). "Comparison of digital rectal examination and serum prostate specific antigen in the early detection of prostate cancer: results of a multicenter clinical trial of 6,630 men." J Urol **151**(5): 1283-90.
- Chun, F. K., A. Briganti, et al. (2006). "Development and External Validation of an Extended 10-Core Biopsy Nomogram." Eur Urol.
- Chun, F. K., A. Briganti, et al. (2006). "Significant upgrading affects a third of men diagnosed with prostate cancer: predictive nomogram and internal validation." BJU Int **98**(2): 329-34.
- Chun, F. K., M. Graefen, et al. (2006). "Anatomic radical retropubic prostatectomy-long-term recurrence-free survival rates for localized prostate cancer." World J Urol **24**(3): 273-80.

- Chun, F. K., P. I. Karakiewicz, et al. (2006). "Prostate cancer nomograms: an update." Eur Urol **50**(5): 914-26; discussion 926.
- Chun, F. K., P. Perrotte, et al. (2006). "Prostate specific-antigen distribution in asymptomatic Canadian men with no clinical evidence of prostate cancer." BJU Int **98**(1): 50-3.
- Chun, F. K., T. Steuber, et al. (2006). "Development and internal validation of a nomogram predicting the probability of prostate cancer Gleason sum upgrading between biopsy and radical prostatectomy pathology." Eur Urol **49**(5): 820-6.
- Deger, S., D. Bohmer, et al. (2001). "[Brachytherapy of local prostatic carcinoma]." Urologe A **40**(3): 181-4.
- Ebert, T., Altwein, J.E., Wirth, M. (2001). "Primäre Hormontherapie beim fortgeschrittenen Prostatakarzinom." Dt. Ärzteblatt, Jg 98(5).
- Epstein, J. I., P. C. Walsh, et al. (1994). "Pathologic and clinical findings to predict tumor extent of nonpalpable (stage T1c) prostate cancer." Jama **271**(5): 368-74.
- Garzotto, M., T. M. Beer, et al. (2005). "Improved detection of prostate cancer using classification and regression tree analysis." J Clin Oncol **23**(19): 4322-9.
- Garzotto, M., R. G. Hudson, et al. (2003). "Predictive modeling for the presence of prostate carcinoma using clinical, laboratory, and ultrasound parameters in patients with prostate specific antigen levels < or = 10 ng/mL." Cancer **98**(7): 1417-22.
- Gleason, D. F. and G. T. Mellinger (1974). "Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging." J Urol **111**(1): 58-64.
- Graefen, M., P. I. Karakiewicz, et al. (2002). "Validation study of the accuracy of a postoperative nomogram for recurrence after radical prostatectomy for localized prostate cancer." J Clin Oncol **20**(4): 951-6.
- Graefen, M., P. I. Karakiewicz, et al. (2002). "International validation of a preoperative nomogram for prostate cancer recurrence after radical prostatectomy." J Clin Oncol **20**(15): 3206-12.
- Graefen, M., J. Walz, et al. (2006). "Open retropubic nerve-sparing radical prostatectomy." Eur Urol **49**(1): 38-48.
- Greene, F., Page, D., Fleming, I., Fritz, A., Bach, C., Haller, D., et al (2002). "American Joint Committee on Cancer Staging Manual." **ed 6**.

- Henderson, L. J. (1928). Blood: a study in general physiology. Yale University Press.  
**Vol. 3 New Haven:** p. 148, Fig. 141.
- Hudson, M. A., R. R. Bahnson, et al. (1989). "Clinical use of prostate specific antigen in patients with prostate cancer." J Urol **142**(4): 1011-7.
- Hull, G. W., F. Rabbani, et al. (2002). "Cancer control with radical prostatectomy alone in 1,000 consecutive patients." J Urol **167**(2 Pt 1): 528-34.
- Jemal, A., R. C. Tiwari, et al. (2004). "Cancer statistics, 2004." CA Cancer J Clin **54**(1): 8-29.
- Jewett, H. J. (1956). "Significance of the palpable prostatic nodule." J Am Med Assoc **160**(10): 838-9.
- Karakiewicz, P. I., S. Benayoun, et al. (2005). "Development and validation of a nomogram predicting the outcome of prostate biopsy based on patient age, digital rectal examination and serum prostate specific antigen." J Urol **173**(6): 1930-4.
- Karakiewicz, P. I., F. K. Chun, et al. (2006). "Prostate cancer nomograms are superior to neural networks." Can J Urol **13 Suppl 2**: 18-25.
- Kattan, M. W. (2002). "Nomograms. Introduction." Semin Urol Oncol **20**(2): 79-81.
- Kattan, M. W. (2003). "Judging new markers by their ability to improve predictive accuracy." J Natl Cancer Inst **95**(9): 634-5.
- Kattan, M. W. (2006). "Validating a prognostic model." Cancer **107**(11): 2523-4.
- Kattan, M. W., J. A. Eastham, et al. (1998). "A preoperative nomogram for disease recurrence following radical prostatectomy for prostate cancer." J Natl Cancer Inst **90**(10): 766-71.
- Kattan, M. W., T. M. Wheeler, et al. (1999). "Postoperative nomogram for disease recurrence after radical prostatectomy for prostate cancer." J Clin Oncol **17**(5): 1499-507.
- King, C. R. and J. P. Long (2000). "Prostate biopsy grading errors: a sampling problem?" Int J Cancer **90**(6): 326-30.
- King, C. R., D. A. Patel, et al. (2005). "Prostate biopsy volume indices do not predict for significant Gleason upgrading." Am J Clin Oncol **28**(2): 125-9.
- Kirby, R. S., Christmas, T.J., Brawer, M.K. (2001). Prostate Cancer, p.60, Mosby International.
- Kleer, E. and J. E. Oesterling (1993). "PSA and staging of localized prostate cancer." Urol Clin North Am **20**(4): 695-704.

- Kuefer, R. (2004). Tastuntersuchung, PSA und transrektaler Ultraschall. Rationale Diagnostik beim Prostatakarzinom, p.127-132.
- Labrie, F., B. Candas, et al. (2004). "Screening decreases prostate cancer mortality: 11-year follow-up of the 1988 Quebec prospective randomized controlled trial." Prostate **59**(3): 311-8.
- Lambert, J. H. (1779). Pyrometrie; oder, Vom Maasse des Feuers und der Wärme mit acht Kupfertafeln. Berlin, Bey Haude & Spener, pp. 424-88.
- Lilja, H., Piironen, T.P., Rittenhouse, H.G. et al (2000). Value of molecular forms of prostate specific antigen and related kallikrein, hK2, in diagnosis and staging of prostate cancer, Lippincott Williams & Wilkins, p. 638-650.
- Matlaga, B. R., L. A. Eskew, et al. (2003). "Prostate biopsy: indications and technique." J Urol **169**(1): 12-9.
- McNeal, J. E. (1969). "Origin and development of carcinoma in the prostate." Cancer **23**(1): 24-34.
- McNeal, J. E., E. A. Redwine, et al. (1988). "Zonal distribution of prostatic adenocarcinoma. Correlation with histologic pattern and direction of spread." Am J Surg Pathol **12**(12): 897-906.
- Michl, U., M. Graefen, et al. (2003). "[Functional results of various surgical techniques for radical prostatectomy]." Urologe A **42**(9): 1196-202.
- Mikolajczyk, S. D., L. S. Millar, et al. (2000). "A precursor form of prostate-specific antigen is more highly elevated in prostate cancer compared with benign transition zone prostate tissue." Cancer Res **60**(3): 756-9.
- Noldus, J., Z. Chen, et al. (1997). "Isolation and characterization of free form prostate specific antigen (f-PSA) in sera of men with prostate cancer." J Urol **158**(4): 1606-9.
- Noldus, J., M. Graefen, et al. (2000). "Stage migration in clinically localized prostate cancer." Eur Urol **38**(1): 74-8.
- Nurmikko, P., V. Vaisanen, et al. (2000). "Production and characterization of novel anti-prostate-specific antigen (PSA) monoclonal antibodies that do not detect internally cleaved Lys145-Lys146 inactive PSA." Clin Chem **46**(10): 1610-8.
- Partin, A. W., M. W. Kattan, et al. (1997). "Combination of prostate-specific antigen, clinical stage, and Gleason score to predict pathological stage of localized prostate cancer. A multi-institutional update." Jama **277**(18): 1445-51.

- Partin, A. W., C. R. Pound, et al. (1993). "Serum PSA after anatomic radical prostatectomy. The Johns Hopkins experience after 10 years." Urol Clin North Am **20**(4): 713-25.
- Postma, R. and F. H. Schroder (2005). "Screening for prostate cancer." Eur J Cancer **41**(6): 825-33.
- Pound, C. R., A. W. Partin, et al. (1997). "Prostate-specific antigen after anatomic radical retropubic prostatectomy. Patterns of recurrence and cancer control." Urol Clin North Am **24**(2): 395-406.
- Presti, J. C., Jr., J. J. Chang, et al. (2000). "The optimal systematic prostate biopsy scheme should include 8 rather than 6 biopsies: results of a prospective clinical trial." J Urol **163**(1): 163-6; discussion 166-7.
- (2005). "PSA-Tests helfen, Prostata-Ca-Mortalität zu senken." Ärztezeitung, 2005, 28.02.2005.
- Revdol, D. M., McLeod, D.G., Brandt, J.M., Snow, P.B., Murphy, G.P. (2001). "Introduction to artificial neural networks for physicians: taking the lid off the black box." Prostate **46**(1), p. 39-44.
- Ross, P. L., P. T. Scardino, et al. (2001). "A catalog of prostate cancer nomograms." J Urol **165**(5): 1562-8.
- Sakr, W. A., D. J. Grignon, et al. (1994). "High grade prostatic intraepithelial neoplasia (HGPIN) and prostatic adenocarcinoma between the ages of 20-69: an autopsy study of 249 cases." In Vivo **8**(3): 439-43.
- Semjonow, A., B. Brandt, et al. (1996). "Discordance of assay methods creates pitfalls for the interpretation of prostate-specific antigen values." Prostate Suppl **7**: 3-16.
- Shariat, S. F., E. I. Canto, et al. (2004). "Beyond prostate-specific antigen: new serologic biomarkers for improved diagnosis and management of prostate cancer." Rev Urol **6**(2): 58-72.
- Shinohara, K., Wheeler, T.M., Scardino, P.T. (1989). "The appearance of prostate cancer on transrectal ultrasonography: correlation of imaging and pathological examinations." J Urol: 142: 76-82.
- Siegsmond, M., A. Musial, et al. (2001). "[Ldr brachytherapy, a minimally invasive alternative in the treatment of organ-confined prostate cancer]." Onkologie **24 Suppl 5**: 46-50.
- Snow, P. B., D. S. Smith, et al. (1994). "Artificial neural networks in the diagnosis and prognosis of prostate cancer: a pilot study." J Urol **152**(5 Pt 2): 1923-6.

- Specht, M. C., M. W. Kattan, et al. (2005). "Predicting nonsentinel node status after positive sentinel lymph biopsy for breast cancer: clinicians versus nomogram." Ann Surg Oncol **12**(8): 654-9.
- Stenman, U. H., J. Leinonen, et al. (1991). "A complex between prostate-specific antigen and alpha 1-antichymotrypsin is the major form of prostate-specific antigen in serum of patients with prostatic cancer: assay of the complex improves clinical sensitivity for cancer." Cancer Res **51**(1): 222-6.
- Stephan, C., H. Cammann, et al. (2002). "Multicenter evaluation of an artificial neural network to increase the prostate cancer detection rate and reduce unnecessary biopsies." Clin Chem **48**(8): 1279-87.
- Steuber, T., A. J. Vickers, et al. (2006). "Risk assessment for biochemical recurrence prior to radical prostatectomy: significant enhancement contributed by human glandular kallikrein 2 (hK2) and free prostate specific antigen (PSA) in men with moderate PSA-elevation in serum." Int J Cancer **118**(5): 1234-40.
- Steyerberg, E. W., F. E. Harrell, Jr., et al. (2001). "Internal validation of predictive models: efficiency of some procedures for logistic regression analysis." J Clin Epidemiol **54**(8): 774-81.
- Steyerberg, E. W., F. E. Harrell, Jr., et al. (1998). "Neural networks, logistic regression, and calibration." Med Decis Making **18**(3): 349-50.
- Stolzenburg, J. U., M. C. Truss, et al. (2003). "Evolution of endoscopic extraperitoneal radical prostatectomy (EERPE)--technical improvements and development of a nerve-sparing, potency-preserving approach." World J Urol **21**(3): 147-52.
- Suzuki, H., A. Komiya, et al. (2006). "Development of a nomogram to predict probability of positive initial prostate biopsy among Japanese patients." Urology **67**(1): 131-6.
- Thompson, I. M., D. K. Pauler, et al. (2004). "Prevalence of prostate cancer among men with a prostate-specific antigen level  $\leq$  4.0 ng per milliliter." N Engl J Med **350**(22): 2239-46.
- Vaisanen, V., J. Lovgren, et al. (1999). "Characterization and processing of prostate specific antigen (hK3) and human glandular kallikrein (hK2) secreted by LNCaP cells." Prostate Cancer Prostatic Dis **2**(2): 91-97.
- Walsh, P. C. and P. J. Donker (1982). "Impotence following radical prostatectomy: insight into etiology and prevention." J Urol **128**(3): 492-7.

- Walz, J., M. Graefen, et al. (2006). "High incidence of prostate cancer detected by saturation biopsy after previous negative biopsy series." Eur Urol **50**(3): 498-505.
- Wang, M. C., L. A. Valenzuela, et al. (1979). "Purification of a human prostate specific antigen." Invest Urol **17**(2): 159-63.
- Wedding, U., Höffken, K., Hörtl, W. (2000). "Medikamentöse Therapie des hormonsensiblen und hormonresistenten Prostatakarzinoms." Der Onkologe, p.137-143 **6**(2).
- Wei, J. T., Z. Zhang, et al. (1998). "Understanding artificial neural networks and exploring their potential applications for the practicing urologist." Urology **52**(2): 161-72.
- Wolff, J., Altwein, J.E. (2004). Prostatakarzinom-Grundlagen und Therapie, Springer, p. 61ff.
- Yanke, B. V., B. S. Carver, et al. (2006). "African-American race is a predictor of prostate cancer detection: incorporation into a pre-biopsy nomogram." BJU Int **98**(4): 783-7.
- Zincke, H., J. E. Oesterling, et al. (1994). "Long-term (15 years) results after radical prostatectomy for clinically localized (stage T2c or lower) prostate cancer." J Urol **152**(5 Pt 2): 1850-7.

## 7 Danksagung

An dieser Stelle möchte ich mich bei verschiedenen Personen bedanken, die zum Gelingen dieser Arbeit beigetragen haben.

Mein besonderer Dank gilt:

Prof. Dr. med. Markus Graefen, Chefarzt der Martini-Klinik am UKE, Hamburg-Eppendorf, für die Überlassung des Themas und seine Unterstützung, besonders in der Anfangsphase der Arbeit.

Dr. med. Jongi Chun aus der Urologischen Klinik des UKE, Hamburg-Eppendorf, für seine unermüdliche und exzellente Betreuung.

Meinem Bruder Daniel Hopp für seine Hilfe in Gestaltungsfragen.

Richard von Lossow für die ständigen Ermutigungen während des Verfassens dieser Arbeit.

Meinen Eltern Clemens und Gunhild Hopp für die jahrelange Unterstützung und die Ermöglichung nicht nur dieses Studiums. Ihnen ist diese Arbeit gewidmet.

## 8 Lebenslauf

### *Angaben zur Person*

**Name:** Julia Marie Hopp  
**Geburtsdatum:** 29.12.1978  
**Geburtsort:** Buchholz in der Nordheide  
**Familienstand:** ledig  
**Anschrift:** Sillemstraße 36, 20257 Hamburg

### *Ausbildung*

**Schule:**

1986-1990	Grundschule Steinbeck, Buchholz
1990-1992	Orientierungsstufe Am Kattenberge, Buchholz
1992-1999	Gymnasium Am Kattenberge, Buchholz

**Studium:**

WiSe 99	Magisterstudium der Geschichte an der Uni Bremen (Abschluß Grundstudium SoSe 2001)
SoSe 02	Beginn des Studiums der Humanmedizin an der Uni Hamburg
SoSe 04	Physikum
Febr. 07	Beginn des Praktischen Jahres in der Asklepios Klinik St.Georg ( Innere Medizin)
Juni 07	Chirurgie-Tertial PJ (Waid-Spital, Zürich)
Okt. 07	PJ Wahlfach Gynäkologie in der Asklepios Klinik Barmbek
Apr. 08	Staatsexamen der Humanmedizin

## **9 Erklärung**

### **EIDESSTATTLICHE VERSICHERUNG:**

Ich versichere ausdrücklich, dass ich die Arbeit selbständig und ohne fremde Hilfe verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die aus den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen einzeln nach Ausgabe (Auflage und Jahr des Erscheinens), Band und Seite des benutzten Werkes kenntlich gemacht habe.

Ferner versichere ich, dass ich die Dissertation bisher nicht einem Fachvertreter an einer anderen Hochschule zur Überprüfung vorgelegt oder mich anderweitig um Zulassung zur Promotion beworben habe.

Unterschrift: .....

(Julia Marie Hopp)