

**UNIVERSITÄTSKLINIKUM HAMBURG-EPPENDORF**

Institut für Pathologie

Prof. Dr. med. Guido Sauter

**Evaluierung des quantitativen Gleason-Score Systems zur  
präziseren Beurteilung von Prostatakarzinomen**

**Dissertation**

zur Erlangung des Grades eines Doktors der Medizin / Zahnmedizin  
an der Medizinischen Fakultät der Universität Hamburg.

vorgelegt von:

Tim Julian Janssen  
aus Freiburg

Hamburg 2016

**Angenommen von der  
Medizinischen Fakultät der Universität Hamburg am: 26.10.2017**

**Veröffentlicht mit Genehmigung der  
Medizinischen Fakultät der Universität Hamburg.**

**Prüfungsausschuss, der/die Vorsitzende: Prof. Dr. Guido Sauter**

**Prüfungsausschuss, zweite/r Gutachter/in: Prof. Dr. Derya Tilki**

## **Inhaltsverzeichnis**

### **1. Einleitung**

1.1	Definition des Gleason-Scores.....	4
1.2	Historische Einordnung des Gleason-Scores.....	4
1.21	Modifikationen zur Angabe der Gleason-Grade.....	5
1.3	Herausforderungen der Gleason-Graduierung.....	7
1.4	Optimierungsansätze.....	11
1.5	Zielsetzung und Fragestellung der Arbeit.....	12

### **2. Material und Methoden**

2.1	Patienten.....	13
2.2	Biopsie-Befunde.....	13
2.3	Prostatektomie-Befunde.....	14

### **3. Ergebnisse**

3.1	Stanzbiopsie versus radikale Prostatektomie (traditionelle Methode).....	15
3.2	Quantitativer Gleason-Grad.....	17

### **4. Diskussion**

4.1	Allgemeines.....	19
4.2	Durchschnittlicher Gleason-Grad.....	20
4.3	Quantitativer Gleason-Grad.....	21

<b>5. Zusammenfassung.....</b>	<b>23</b>
--------------------------------	-----------

<b>6. Abstract.....</b>	<b>25</b>
-------------------------	-----------

<b>7. Abkürzungsverzeichnis.....</b>	<b>26</b>
--------------------------------------	-----------

<b>8. Tabellenverzeichnis u. Abbildungsverzeichnis.....</b>	<b>29</b>
---	-----------

<b>9. Literaturverzeichnis.....</b>	<b>28</b>
-------------------------------------	-----------

<b>10. Danksagung.....</b>	<b>32</b>
----------------------------	-----------

<b>11. Lebenslauf.....</b>	<b>33</b>
----------------------------	-----------

<b>12. Eidesstattliche Erklärung.....</b>	<b>34</b>
---	-----------

# 1. Einleitung

## 1.1 Definition des Gleason-Scores

Der Gleason-Score dient der histopathologischen Einteilung des Adenokarzinoms der Prostata. Die Graduierung nach Donald F. Gleason ist das einzige Graduierungssystem, das ausschließlich auf architektonischen Kriterien beruht [1, 2]. Veränderungen der Zellkerne und des Zytoplasmas spielen für die Gleason-Graduierung keine Rolle. Bereits bei der Erstbeschreibung des Gleason-Graduierungssystems 1966 durch Donald F. Gleason unterschied dieser 5 verschiedene Wachstumsmuster oder Gleason-Grade, welche die mögliche Malignisierung von normalem Prostatadrüsenepithel abbildeten [1]. Die Graduierung stellt sich prinzipiell wie folgt dar:

<b>Gleason-Grad</b>	<b>Morphologische Kriterien</b>
<b>1</b>	Kleine regelmäßige Drüsen, die scharf begrenzt sind.
<b>2</b>	Kleine weitgehend regelrechte Drüsen, die scharf begrenzt sind und nicht in die Umgebung infiltrieren.
<b>3</b>	Kleine bis mittelgroße, voneinander noch abgrenzbare Drüsen, die nicht konfluent wachsen und in die Umgebung infiltrativ wachsen.
<b>4</b>	Bestimmender Parameter für Gleason 4 ist die Konfluenz von Drüsen. Diese kann entweder in Drüsenfusionen erkennbar sein oder durch Ausbildung sogenannter kribriformer Wachstumsmuster.
<b>5</b>	Hier kommt es zur Ausbildung solider Zellverbände, zu Einzelzellinfiltraten oder zu Komedonekrosen.

**Tabelle 1. Graduierung nach Gleason (angelehnt an Egevad 2006).**

## 1.2 Historische Einordnung des Gleason-Scores

Bereits initial wurden die Gleason-Muster in einer Zeichnung dargestellt, die alle damals charakterisierten Muster beinhaltete. Da es immer wieder zu Präzisierungen des Gleason-Grades kam, gibt es „offizielle“ Gleason-Schemazeichnungen von 1966, 1974, 1992, 1994, 2005, 2010 [3-8]. Präzisierungen des Gleason-Systems erfolgten über die Zeit immer wieder (vgl. Abbildung 1).

Diese wurden v.a. durch den seit 1966 stattgefundenen Fortschritt in Diagnose und Therapie induziert. In den Jahren nach Etablierung des Graduierungsschemas durch Gleason beschränkten sich die Früherkennungsmaßnahmen (sog. sekundäre Prävention) noch maßgeblich auf die digital-rektale Untersuchung. Mit dieser Methode wurden allerdings vor allem Tumoren in späten Stadien identifiziert. 86 % der Patienten zeigten in einer Studie von 1974 bei Diagnosestellung extraprostatitisches, bzw. schon metastasiertes Tumorstadium [2]. Die Zulassung des PSA-Screenings 1986 führte zu einer sog. Stadienverschiebung, mit insbesondere steigendem Anteil früherer Stadien [9]. Aktuell werden ca. 2/3 aller Tumoren im organbeschränkten Stadium diagnostiziert [10,11]. Das ist klinisch bedeutsam, da eine kurative Behandlung des Prostatakarzinoms nur im organbegrenzten Stadium möglich ist [12].

Auch die diagnostischen Maßnahmen hinsichtlich der Biopsien haben sich dahingehend geändert, dass bei auffälligem Tastbefund u.o. PSA-Wert zur Bestätigung einer Verdachtsdiagnose nicht mehr nur ein oder zwei sondern i.d.R. mindestens zehn bis zwölf Prostataschnittbiopsien entnommen werden.

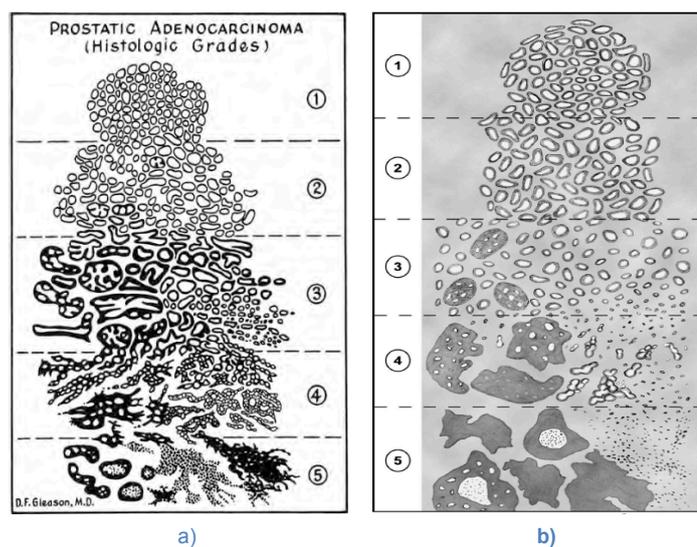


Abbildung 1. Konventionelles (a) und (b) modifiziertes Gleason-Grading System von 1966, respektive 2005. Muster 1 und 2 entsprechen dem alten Schema (wird i.d.R nicht mehr verwendet). Die wichtigsten Änderungen ergeben sich in Muster 3 und 4. In der Modifikation dominieren gleichmäßig kleine Drüsen mit nur kleinen kribriformen Anteilen. Das Muster 4 charakterisiert den Fusionsschub der kleinen Drüsen mit reichlich kribriformen und auch papillären Drüsenanteilen, z. T. resultiert ein hypernephroides klarzelliges Bild. Muster 5 unterscheidet sich nicht signifikant vom konventionellen Gleason (Helpap u. Evegad 2006).

### 1.2.1 Modifikationen zur Angabe der Gleason-Grade

Weitere Modifizierungen des Gleason-Gratings in seiner

originären Form wurden notwendig, da Muster identifiziert wurden, die in früheren Systemen/Schemata nicht eindeutig zugeordnet werden konnten. Diese zusätzlichen Wuchsformen wurden dann einem bestimmten Gleason-Grad zugeordnet. Besondere Wuchsformen von Prostatakarzinomen beinhalten beispielsweise das muzinöse Wachstum, glomeruloide Formationen oder kleine unregelmäßige Drüsen ohne eindeutige Lumenbildung. Ein weiteres Charakteristikum des Gleason-Grades ist, dass grundsätzlich 2 verschiedene Wachstumsmuster berücksichtigt werden sollen. Dies ist der Erkenntnis geschuldet, dass Prostatakarzinome häufig heterogen sind und es mehrere Arten von Wuchsformen in einem einzigen Tumor gibt. Bei der ursprünglichen Beschreibung durch Donald F. Gleason wurde festgelegt, dass immer das am häufigsten anzutreffende Gleason-Muster den primären Gleason-Grad darstellt, während das zweithäufigste Gleason-Muster als sekundärer Gleason-Grad

definiert wurde, wobei Gleason-Grad 1 den höchsten und Gleason-Grad 5 den niedrigsten Differenzierungsgrad darstellt (Abbildung 2). Falls kein sekundärer Grad 5 Prozent überschritt, wurde nur der Primärgrad gewertet, dies dann aber zwei mal, wie beispielsweise  $3 + 3 = 6$  oder  $4 + 4 = 8$ .

Modifikationen zur Art und Weise wie der Gleason-Grad angegeben wird, sind der Erkenntnis geschuldet, dass es Prostatakarzinome gibt, bei denen sehr kleine Tumoranteile maligner aussehen (höherer Gleason-Grad) als diejenigen, die aufgrund ihrer Häufigkeit als primäres und sekundäres Gleason-Muster definiert werden müssen. Um diese (schlimmsten) Tumorkomponenten im Gleason-Grading zu integrieren, wurden für verschiedene Arten von Biopsien unterschiedliche Maßnahmen ergriffen. Bei Gewebeproben mit viel Gewebematerial, welches repräsentativ für den gesamten Tumor sein könnte, wie transurethrale Resektionen und radikale Prostatektomiepräparate, wurde ein Tertiärgrad definiert. Hierbei wird neben dem Primär- und Sekundärgrad auch noch ein dritter Gleason-Grad angegeben, wenn dieser ungünstiger ist als der primäre und sekundäre Grad. Beispiel für einen solchen Fall wäre ein Tumor mit 60 % Gleason 3, 30 % Gleason 4 und 10 % Gleason 5 Anteil. Hier würde nach heutigen Kriterien die Diagnose lauten Gleason  $3 + 4 = 7$  (Tertiärgrad 5) [13,14].

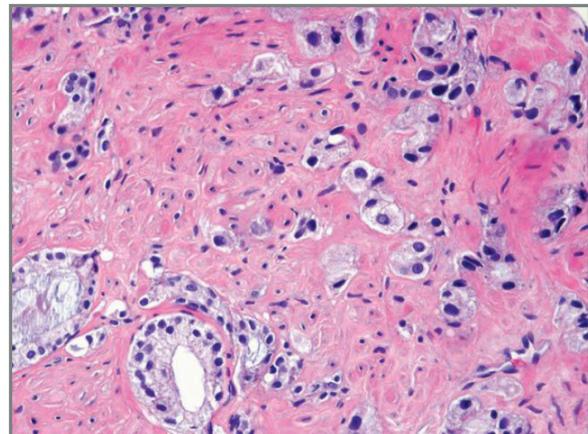


Abbildung 2. Gleason-Grad  $5 + 3 = 8$  mit individuell stehenden Gleason-Grad 5 Zellen (rechts) und fokal abgrenzbaren, weitgehend regelrechten Drüsen (unten links) (Epstein 2013).

Bei Biopsien sind die Tumoranteile häufig viel geringer, so dass andere Definitionen gefunden werden mussten. Hier wurde festgelegt, dass der primäre Gleason-Grad derjenige ist, der am häufigsten im Tumor vertreten ist und der sekundäre der maligneste (schlimmste Gleason-Grad), wenn dieser höher als der primäre und sekundäre Grad ist. Wiederum für das gleiche Beispiel wie oben besprochen, wäre ein Tumor in einer Stanzbiopsie mit 60 % Gleason 3, 30 % Gleason 4 und 10 % Gleason 5, dann ein Gleason  $3 + 5 = 8$  Karzinom. Die Art der Beschreibung eines Tertiärgrades wurde ursprünglich von Pan et al. beschrieben und vorgeschlagen [15]. Die Definition des Biopsiestanzen-Gleasons als „the most & the worst“ wurde erstmals im Jahr 2000 in die Guidelines des College of American Pathologists (CAP) aufgenommen [16] und beispielsweise in der WHO-Klassifikation 2004 so inkludiert [14].

### 1.3 Herausforderungen der Gleason-Graduierung

Die histopathologische Einteilung des Adenokarzinoms erfordert eine visuelle Beurteilung der Drüsenmorphologie durch den Pathologen. Die unterschiedliche (zwangsläufig subjektive) Interpretation der Architekturveränderungen induziert intra- und interindividuelle Unbeständigkeiten, die ein seit Jahrzehnten bekanntes und viel diskutiertes Problem darstellen. Donald F. Gleason selbst schrieb 1992: "histologic grading, no matter how well defined, is not a verifiable scalar measurement" [5]. Für die klinische Evaluation von Therapie (z.B. Active Surveillance vs. radikale operative Resektion) und Prognose ist eine verlässliche Reproduzierbarkeit jedoch essenziell. Während der prognostische Wert des Gleason-Schemas in diversen Studien über das biochemische PSA-Rezidivrisiko [17-22], der Entwicklung von Metastasen nach Prostatektomie [23,24], dem progressionsfreien Überleben [25-28] und der Gesamtüberlebenszeit [29-34] belegt werden konnte, zeigt die Datenlage zur interindividuellen Reproduzierbarkeit deutlich inhomogenere Ergebnisse [3].

*Tabelle 2* vergleicht die wichtigsten Studien zur Interobserver-Variabilität im Zeitraum von 1983 - 2013. Das typischerweise verwendete statistische Maß zur Messung der Interobserver Reliabilität ist Cohens Kappa. Hier gilt:  $\kappa$ -Werte von 0 bis 0,2 zeigen eine geringe (slight) und Werte 0,81 – 1 eine fast ideale (almost perfect) Übereinstimmung [35]. Im arithmetischen Mittel aller Studien zeigte sich eine mittelmäßige (moderate) Übereinstimmung am unteren Rand für  $\bar{\kappa} = 0,455$ , unter Partizipation speziell ausgebildeter Urologen zeigt sich ein verbesserter Wert für  $\bar{\kappa} = 0,526$ .

Hoffnungen, die Reproduzierbarkeit könne nach den Modifikationen der Konferenz der International Society of Urological Pathology (ISUP) 2005 Fortschritte machen, konnten nicht bestätigt werden. Beispielsweise ergibt sich für die jüngeren Studien nach der Konferenz unter Beteiligung der Pathologen mit besonderen Erfahrungen im Gebiet der Urologie ein durchschnittlicher  $\kappa$ -Wert von  $\bar{\kappa} = 0,475$ , während dieser vor der Konferenz noch bei  $\bar{\kappa} = 0,560$  lagen.

Möglicherweise wurde das Problem der Interobserver-Inkonsistenz durch die 2005 ISUP-Publikation von Epstein et al. sogar aggraviert [6]. 2004 erhielten zahlreiche respektierte Urologen weltweit eine Einladung an einer Konferenz teilzunehmen, mit dem Ziel die gegenwärtige Art und Weise der Gleason-Graduierung zu diskutieren. Insbesondere sollte es darum gehen, wie die Urologen bestimmte Grenzfälle beurteilen würden.

Autor (Jahr)	Pathologen		Fälle		Übereinstimmung	
	n	Fachgebiet Uropathologie	Anzahl	Typ	%	K
Ten Kate (1986)	5		50	RPE	36	0,3
Lessels (1997)	12		100	B		0,45
McLean (1997)	3		71		9,9	0,15-0,29
Allsbrook (2001)	41		38	B		0,45
Allsbrook (2001)	10	X	10	B		0,56-0,70
Oyama (2005)	8	X	37	B		0,44-0,49
Helin (2005)	3	X	62	B		0,55-0,62
Burchardt (2008)	29		278	RPE	45,7	0,58
Rodriguez(2011)	4	X	50	B		0,54
Egevad (2013)	337	X	15		46,3 – 89,3	0,41-0,43

Tabbelle 2 Abk: RPE, radikale Prostatektomie; B, Stanzbiopsie; X, alle Teilnehmer sind ausschließlich Uropathologen. Ein Vergleich der Studien ist schwierig, da eine Vielzahl von Störgrößen das Ergebnis beeinflussen (Allsbrook 2001): (1.) Die unterschiedlich Anzahl (n) der teilnehmenden Pathologen, (2.) die Unterschiedliche Anzahl der Präparate, (3.) verschiedene Arten von Gewebeproben und (4.) die unterschiedlich Auslegung von Graduierungskriterien.

Das Ziel der Konferenz war eine Bestandsaufnahme, wie gegenwärtig die erfahrensten Uropathologen Problembereiche des Gleasons interpretieren. Obwohl es ursprünglich nicht beabsichtigt war, eine Veränderung des Gleasons-Systems zu etablieren, sondern nur den Status-Quo zu beschreiben, tauchte am Ende in der Publikation das Wort Guideline auf und es wurde in der Konsequenz von vielen Pathologen von einem „neuen Gleason-Grad“ gesprochen. Jonathan Epstein, der Organisator des Meetings und Spiritus Rector der Konferenz schrieb selber 5 Jahre später, dass die Konferenz zu keinerlei Neuerungen im Gleason-System geführt habe, sondern nur den gegenwärtig üblichen Gebrauch des Gleason-Grades beschreibe [8]. Dennoch haben offenbar viele Pathologen aufgrund ihres Studiums des Manuskriptes der ISUP-Konferenz ihr Gleason Graduierungsverhalten maßgeblich verändert. Burkhard Helpap und Egevad beschrieben beispielsweise in einer 2006 erschienenen Publikation eine Reduktion der Gleason  $\leq 6$  Fälle in Prostatastanzbiopsien von beinahe 50 % auf 22 %, bei einem gleichzeitigem Anstieg der Gleason 7 Fälle nach der Modifizierung von 2005 von 25,5 auf 67,9 % [36]. Es ist davon auszugehen, dass viele andere Pathologen, wie zum Beispiel auch Jonathan Epstein selbst, ihr Graduierungsverhalten nach der Konferenz nicht geändert haben. Auch am Institut für Pathologie des Universitätsklinikums Hamburg-Eppendorf (UKE) wurde das Graduierungsverhalten von Prostatakarzinomen beibehalten. Die einzige Empfehlung des ISUP-Papers, die von dem am UKE geübten Verfahren abwich, war die Klassifizierung von kleinen, unregelmäßigen Drüsen mit manchmal nicht deutlich

erkennbarem Lumen als Gleason 4, statt Gleason 3. Diese Empfehlung wurde offenbar nur von einer knappen 2/3 Mehrheit in der ISUP-Konferenz unterstützt und die Autoren erkannten an, dass diese Form von Veränderung meistens auch mit einer Drüsenkonfluenz verbunden ist, welche ohnehin einem Gleason 4 zugeordnet wird. Diese Veränderung des Gleason-Systems dürfte dafür verantwortlich sein, dass bei bestimmten Gruppen die Zahl der Gleason 4 Tumoren so sprunghaft anstieg. Wahrscheinlich deswegen empfiehlt Jonathan Epstein auch 2013 in einer Übersichtsarbeit zum Gleason-System im Zweifelsfall derartige Veränderungen eher zurückhaltend zu beurteilen [37]. Dieses Verhalten könnte auch als „Rolle rückwärts“ beschrieben werden.

In einer weiteren eindrucksvollen Studie, die das Problem erläutert, hatten Burchardt und Mitarbeiter 2008 von 331 Karzinomen Bilder hergestellt, welche umschriebene Abschnitte der Karzinome beinhalteten [38]. Die Bilder wurden an 29 Pathologen in Deutschland verteilt und die Pathologen wurden um eine Gleason-Graduierung gebeten. Eine komplette Übereinstimmung fand sich mit dem von den Experten als richtig empfundenen Referenz-Grades nur bei durchschnittlich 45,7 % der Fälle. Die deutliche Mehrheit der Pathologen (75,9 %) gaben einen zu niedrigen Gleason-Grad an. Allgemein zeigten sich vor allem low-grade-Karzinome mit den Gleason-Scores 5, 6 und 7 in Studien anfällig für Fehlinterpretationen durch Untergraduierung [38-43]. Eine andere richtungsweisende Studie von Egevad et al. ist aus dem Jahr 2013 und damit mehr oder weniger aktuell [44]. In dieser Untersuchung wurden 25 Grenzfälle zwischen Gleason 6 und Gleason 7 von einer Gruppe spezialisierter Uropathologieexperten auf einem Treffen in Paris gemeinsam beurteilt. Nur bei 15 dieser 25 Fälle konnten sich 60 % der Teilnehmer auf eine Diagnose einigen. Diese Konsens-Fälle wurden dann an 337 Pathologen mit einem besonderem Interesse an Uropathologie geschickt und die Resultate dieser vergleichenden Untersuchung sind in Tabelle 3 dargestellt. Die Befunde zeigen, dass die Übereinstimmung mit der Konsensus-Diagnose zwischen 46,3 % und 89,3 % ( $\bar{x}$  62,4 %) lag, - mit teilweise gravierenden Differenzen zwischen den einzelnen Pathologen. Bei praktisch allen Fällen wurden 3 oder mehr Kombinationen von Gleason-Graden diagnostiziert.

Case	Consensus GS	Expert difficulty score	% Type 1	% Type 2	% Type 3	Most common member GS	Member agreement (%)
1	3 + 3 = 6	1.08	26.7	80.0	33.3	3 + 3 = 6	80.7
2	3 + 4 = 7	1.67	0.0	66.7	66.7	3 + 4 = 7	62.6
3	3 + 4 = 7	1.33	6.7	86.7	66.7	3 + 4 = 7	57.0
4	3 + 4 = 7	1.33	13.3	60.0	60.0	3 + 4 = 7	52.2
5	3 + 3 = 6	1.50	6.7	60.0	40.0	3 + 3 = 6	89.3
6	3 + 4 = 7	1.33	0.0	6.7	33.3	3 + 4 = 7	63.5
7	4 + 3 = 7	1.50	20.0	66.7	60.0	4 + 3 = 7	62.9
8	3 + 4 = 7	1.58	86.7	86.7	60.0	3 + 4 = 7	49.0
9	3 + 4 = 7	1.58	0.0	20.0	20.0	4 + 3 = 7	50.1
10	3 + 3 = 6	1.67	0.0	80.0	46.7	3 + 3 = 6	64.4
11	3 + 3 = 6	1.25	46.7	53.3	13.3	3 + 3 = 6	72.7
12	3 + 3 = 6	1.58	0.0	6.7	26.7	3 + 3 = 6	58.2
13	4 + 3 = 7	1.42	0.0	13.3	26.7	4 + 4 = 8	63.8
14	3 + 4 = 7	1.17	0.0	0.0	20.0	4 + 3 = 7	46.3
15	3 + 3 = 6	1.33	0.0	20.0	13.3	3 + 3 = 6	62.9

Tabelle 3 zeigt den von Experten bestimmten Gleason-Score (GS) in 15 ausgewählten Konsensfällen. Der (Difficulty-Score) dient als Maß für den subjektiv empfundenen Schwierigkeitsgrad, aufsteigend von 1-3. Gleason-Muster 4 Komponenten wurde klassifiziert als Typ 1 (kribrifrom), Typ 2 (Drüsenfusionen), Typ 3 (undifferenzierte (poorly formed) Drüsen). Die Teilnehmer sind ausschließlich Mitglieder der European Network of Urothology (ENUP) (Egevad et al. 2013).

Die Gesamtheit der Probleme bei der „Standardisierung“ des Gleason-Grades wird auch in einer eindrucksvollen Studie aus Schweden gezeigt. Hier hat die Arbeitsgruppe um Frau Dannemann die Gleason-Graduierungen in dem nationalen Prostatakrebsregister in 97168 Fällen von 1998-2011 untersucht [45]. Es zeigt sich über die Zeit eine konstant fortlaufende Zunahme von höheren Gleason-Graden. Bemerkenswerterweise findet diese Inflation kontinuierlich statt ohne besonders markante Veränderungen des Graduierungsverhaltens, bspw. zum Zeitpunkt wichtiger Veröffentlichungen wie der 2005 erschienene ISUP-Publikation. Es handelt sich um einen kontinuierlichen Veränderungsprozess, der wahrscheinlich damit zu tun hat, dass die Pathologen zunehmend besser ausgebildet wurden und es immer mehr junge Pathologen gibt, die nach aktuellen Kriterien diagnostizieren, während ältere, die die Veränderungen nicht mitgemacht haben, „aussterben“. Tatsächlich ist auch nicht zu erwarten, dass eine 15-seitige Publikation wie die 2005 erschienene ISUP-Darstellung<sup>[6]</sup> von allen Pathologen gleichzeitig gelesen und in die Praxis umgesetzt wird.

Zusammengefasst ist die fehlende Reproduzierbarkeit der Gleason-Grade in 2 besonderen Punkten gravierend:

Problem Nr. 1 ist die große Bedeutung des Gleason-Grades für die Therapiewahl. Nur geringe Veränderungen in der Beurteilung eines Tumors, z.B. 3 + 3 versus 3 + 4 oder 3 + 4 versus 4 + 3, führen zu dramatischen Änderungen in den therapeutischen Optionen. Bei Gleason 3 + 3 = 6 steht Active Surveillance im Vordergrund, engmaschige Überwachung und Re-Staging sollen vorschnelle operative Eingriffe mit entsprechenden Komplikationen minimieren. Bei Gleason 3 + 4 = 7 wird Active Surveillance deutlich weniger präferiert. Beispielsweise hält das deutsche Leitlinienprogramm Gleason 3 + 4 für nicht geeignet als Routineaufnahmekriterium für die Aktive Überwachung [12]. Für andere Autoren stellt die Strategie der Aktiven Überwachung bei 3 + 4 = 7 Karzinomen trotzdem eine stadiengerechte Therapieoption dar [46, 47], die Rationale hierfür ist v.a. die durch die ISUP Konferenz verursachte Stadienverschiebung von Gleason 3 + 3 = 6 zu 3 + 4 = 7 [12]. Bei Gleason 4 + 3 = 7 ist Active Surveillance keine Option mehr. Die Wahl des „richtigen Gleason-Grades“ ist somit für den Patienten entscheidend.

Problem Nr.2 ist, dass es Grenzbereiche gibt, bei denen der Pathologe Schwierigkeiten hat zwischen 3 + 3 und 3 + 4, oder zwischen 3 + 4 und zwischen 4 + 3 zu entscheiden.

#### 1.4 Optimierungsansätze

Um diese Probleme besser in den Griff zu bekommen, wurde am Institut für Pathologie des UKE vor etwa 10 Jahren (2005) entschieden, für jeden einzelnen Prostatakrebs den prozentualen Anteil der Gleason-Komponenten abzuschätzen. Dies in der Idee, nicht nur einen Gleason-Score anzugeben, wie bspw. 3 + 3 oder 3 + 4, sondern dem Empfänger der Diagnose auch mitzuteilen, ob und in welcher Weise diese Entscheidung grenzwertig war. Die Ausgangshypothese war damals, dass ein Gleason 3 + 4 mit 5 % Gleason 4 sich sehr ähnlich wie ein Gleason 3 + 3 Tumor verhält und, dass ein Gleason 3 + 4 mit 45 % Gleason 4 sich ähnlich verhalten sollte wie ein Gleason 4 + 3 mit 55 % Gleason 4. In einer früheren Untersuchung von Cheng et al. 2005 über die Bedeutung des prozentualen Gleason 4/5 Anteils als Prädiktor für die Tumorprogression bei Prostatakarzinompatienten, wurde tatsächlich festgestellt, dass die prozentualen Anteile der Gleason-Grade bei Prostatektomiepräparaten einen dramatischen Effekt auf die Patientenprognose haben, so dass eine „Dose-Response“ Kurve der PSA-Rezidive mit zunehmendem Anteil Gleason 4 entstand [29]. Abbildung 3 zeigt die Beziehung zwischen dem relativen Tumoranteil von Gleason-Muster 4 und 5 in Prostatektomiepräparaten und dem postoperativen biochemischen Rezidivrisiko.

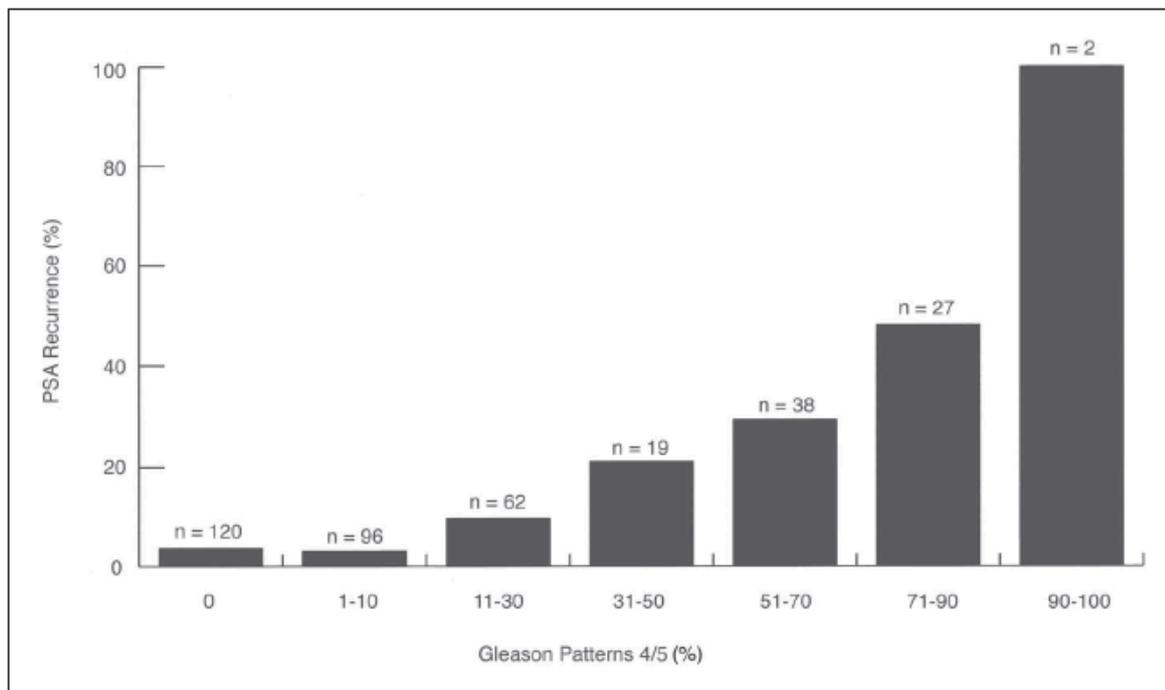


Abbildung 3, „Dosis-Wirkungs-Beziehung“ von Gleason-Muster 4/5 in % und der PSA-Rezidiv-Rate. Mit wachsendem Anteil Gleason-Muster 4/5 (Abszisse), steigt die biochemische Rezidivrate (Ordinate) überproportional an. Als PSA-Rezidiv wurde ein PSA-Wiederanstieg nach RPE von 0,1 ng/mL definiert (Cheng et al. 2005).

Ein besonderes Problem bei der Interpretation von Stanzbiopsie Gleason-Graden ist auch die Tatsache, dass die meisten Nomogramme und Prädiktionssysteme nur einen einzigen Gleason-Grad pro Patient kennen und nicht bspw. 6 oder 7 verschiedenen Gleason-Grade, gewonnen aus Stanzen verschiedener Herkunftsorte, von einem Patienten beinhalten. Die meisten Urologen verwenden deswegen für ihre Prognosebeurteilung den schlechtesten Gleason. Dies ist aber fragwürdig, wenn 6 Stanzen Gleason 3 + 3 zeigen und nur eine Stanze einen 0,1 mm messenden Gleason 4 + 4 Herd. Am Institut für Pathologie des UKE wurde deswegen schon seit mehr als 10 Jahren (2005) bei jeder Stanzbiopsie-Diagnose eine zusammenfassende Beurteilung erstellt, welche die Gleason-Grade der einzelnen Stanzen beinhalten und einen Durchschnittswert bilden, wobei auch die Tumormenge jedes Zylinders mit einfluss.

### 1.5 Zielsetzung und Fragestellung der Arbeit

Hypothese dieser Studie ist, dass die prädiktive Aussage dieses Durchschnittswertes größer ist als die des „schlimmsten“ Gleasons. Und, dass die graduelle Quantifizierung der Gleason Anteile, im Vergleich zur traditionellen Kategorisierung

der Tumoren, ein feinerer Prädiktor für den Gleason-Grad am nachfolgenden Prostatektomiepräparat sei.

Zur Klärung dieser Fragestellungen und Hypothesen wurden im Rahmen dieser Studie die prozentualen Gleason-Anteile in Stanzbiopsien mit den Resultaten an Prostatektomien in gematchten Biopsieprobenpaaren von 1657 Patienten untersucht. Die Ergebnisse zeigen tatsächlich, dass der durchschnittliche Gleason-Grad von entscheidender Bedeutung bei der Voraussage von Pathologiebefunden an einem Prostatektomiepräparat ist.

## 2. Material und Methode

### 2.1 Patienten

In die Studie flossen 1657 Patienten ein, bei denen im Zeitraum von 2009 - 2014 eine radikale Prostatektomie (RPE) in der Martiniklinik durchgeführt wurde und bei denen vorher ein Prostatakarzinom in einer am UKE untersuchten Stanzbiopsie diagnostiziert wurde.

### 2.2 Biopsie-Befunde

Die Biopsie-Befunde beinhalteten eine separate Diagnose für jeden einzelnen Entnahmeort. Die Diagnose umfasst für jeden Entnahmeort einen Gleason-Score und die prozentualen Anteile sämtlicher Gleason-Grade, die in der genannten Stanze gefunden wurden. Zusätzlich wurde von jedem Entnahmeort in der Stanzbiopsie der Tumorbefall in Millimetern erfasst. Unter Verwendung der unten angegebenen Formel wurden aus diesen Daten für jeden einzelnen Patienten die prozentualen Anteile für Gleason 3, 4 und 5 ermittelt.

Berechnung des prozentualen Anteils eines bestimmten Gleason-Grades über alle Stanzen mit  $x = \text{Gleason 3, Gleason 4 oder Gleason 5}$ ;  $S_1 = 1.$  Stanze;  $S_n = \text{letzte Stanze}$ ;  $\text{Tumor (mm)} = \text{Gesamtlänge des tumorösen Gewebes in allen Stanzen}$ .

$$\% \text{ Gleason } x = \frac{\sum_{S_1}^{S_n} \text{Gleason } x \text{ (mm)}}{\sum_{S_1}^{S_n} \text{Tumor (mm)}} * 100$$

Diese Werte wurden im Anschluss verwendet um einen Durchschnittswert (Summen-Gleason) aus allen Stanzen zu bestimmen. Der Summen-Gleason über alle Stanzen

wird durch Addition des häufigsten Gleason-Grades plus des zweithäufigsten Gleason-Grades und gegebenenfalls unter Angabe des dritthäufigsten Gleason-Grades (Tertiärgrad; aber nur sofern dieser schlimmer ist als der zweithäufigste Gleason Grad).

Als schlimmste Stanze wurde die Stanze mit dem höchsten Gleason-Score bezeichnet. Der Gleason-Score einer individuellen Stanze wird durch Addition des häufigsten und des schlimmsten Gleason-Grades angegeben.

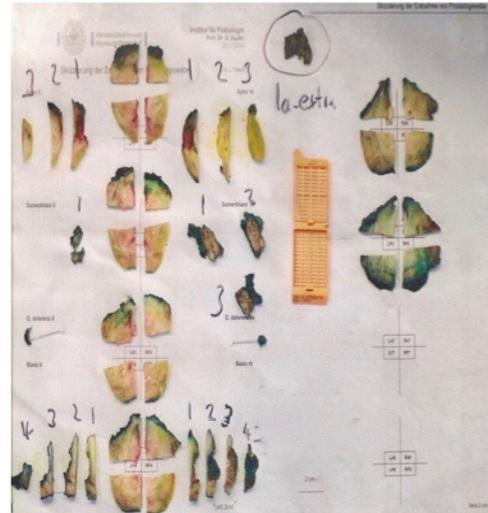


Abbildung 4. Exemplarische Skizzierung eines Prostata-Schnittpräparates (der Entnahme von Prostatagewebe) am Institut für Pathologie der Universität Hamburg-Eppendorf 11/2014

### 2.3 Radikale Prostatektomiepräparate

Die Prostatektomiepräparate wurden am Institut für Pathologie seit 2005 nach einem standardisierten Verfahren untersucht. Dabei wurden die apikalen und basalen Scheiben der Prostata entfernt und quer durchlamelliert. Die mittleren (verbliebenen) Prostataanteile wurden in Scheiben geschnitten und jede Scheibe wurde in der Regel in 4 Abschnitte (Viertel) unterteilt. Von allen Tumoren wird seit September 2009 eine Fotodokumentation angefertigt, auf der die tumorbefallenen Areale eingezeichnet werden (Abbildung 4). Die histologische Befundung umfasste seit 2005 die Erfassung aller im Präparat vorhandenen Gleason-Grade mit ihrer prozentualen Verteilung. Diese Gleason-Grad-Anteile wurden jeweils für jeden Patienten in einen Gleason-Score (einschließlich Tertiärgrad) übersetzt.

### 2.3 Statistik

In allen statistischen Analysen wurde der Chi<sup>2</sup>-Test angewendet.

### 3. Ergebnisse

#### 3.1 Stanzbiopsie versus radikale Prostatektomie (traditionelle Methode)

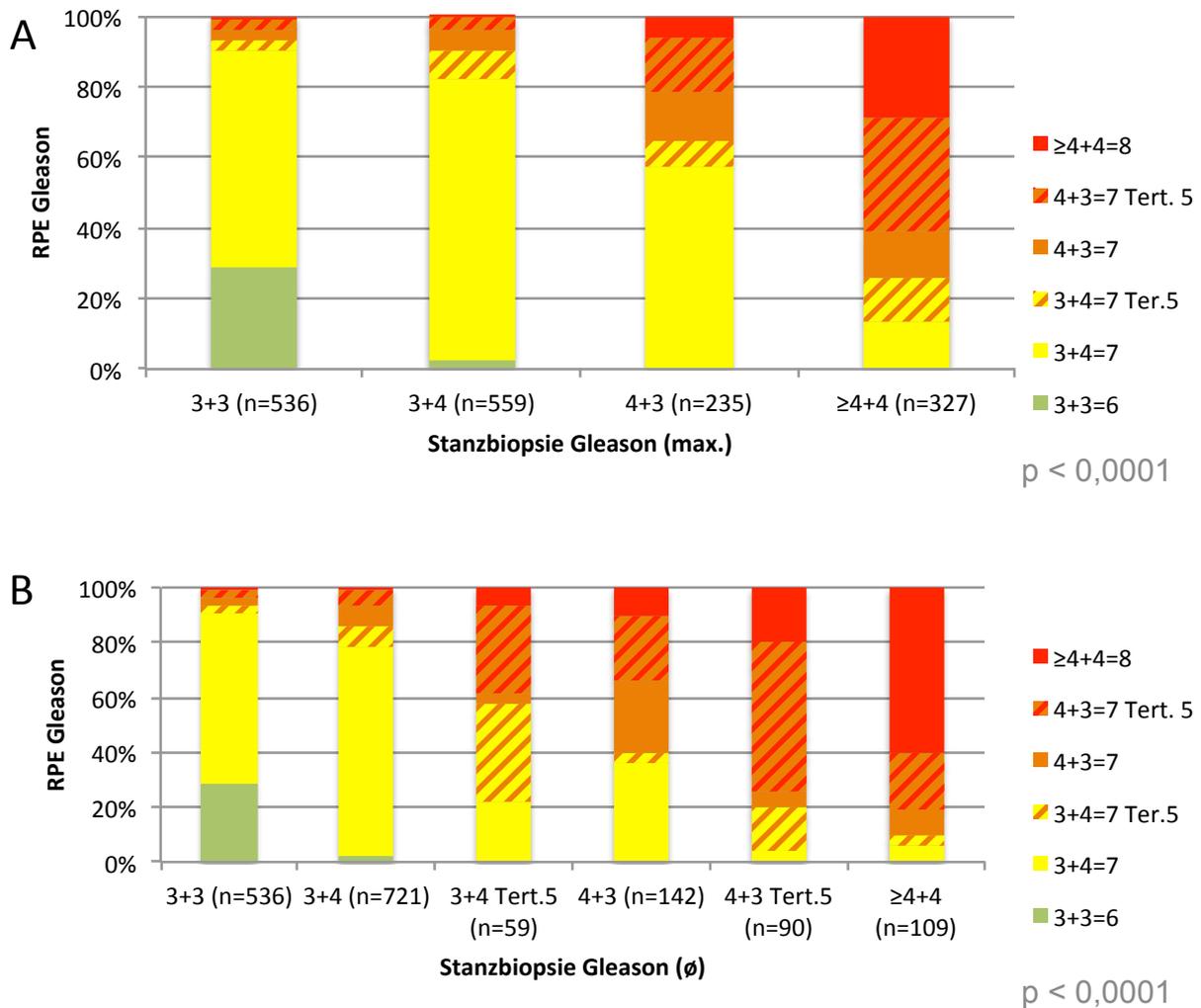


Abbildung 5 – Vergleich der Gleason-Beurteilungen an Stanzbiopsien versus radikale Prostatektomien bei Klassifizierung des Stanzbefundes nach der Stanze mit dem schlimmsten (maximalen) Gleason-Score (A) und nach dem Durchschnittswert (Summen-Gleason, Ø) aller Stanzen (B).

Die Abbildung 5 zeigt den Vergleich der Gleason-Beurteilungen an Stanzbiopsien versus radikale Prostatektomien bei Klassifizierung des Stanzbefundes nach der schlimmsten Stanze (Abbildung 5 A) und nach dem Durchschnittswert aller Stanzen (Abbildung 5 B). Naturgemäß führen beide Verfahren bei Gleason 3 + 3 Karzinomen in Stanzbiopsien zu den gleichen Ergebnissen. Hier zeigen 90 % der Tumoren im Prostatektomiepräparat entweder einen Gleason 3 + 3 oder Gleason 3 + 4. In lediglich 10 % der Fälle kommt es zu einem prognostisch deutlich ungünstigeren Tumor mit entweder führendem Gleason 4 oder einem Tertiärgrad Gleason 5.

Tumoren, die als Gleason 3 + 4 aufgrund der Stanzbiopsien klassifiziert wurden, weisen eine insgesamt günstige Risikokonstellation, unabhängig davon, ob der Gleason 3 + 4 die schlimmste Stanze war oder der Durchschnittswert. Und in beiden Fällen zeigen circa 80 % der Patienten einen eher günstigen Befund in der radikalen Prostatektomie (Gleason 3 + 3 oder Gleason 3 + 4) und es kommt bei 18 % bzw. 22 % zur Diagnose eines führenden Gleason 4 oder eines Tertiärgrades Gleason 5.

Für alle höhergradigen Tumoren scheint der durchschnittliche Gleason, gebildet aus den Befunden in allen tumorhaltigen Stanzbiopsien, aussagekräftiger zu sein als die Klassifikation eines Patienten nach seiner schlimmsten Biopsie. Bei Betrachtung des Durchschnittswertes fällt insbesondere auf, dass diejenigen Tumoren mit einem tertiären Gleason 5 in der Stanzbiopsie in 80 % bis über 90 % einen ungünstigen Tumor in sich tragen, meist Gleason 4 + 3 oder 3 + 4 mit Tertiärgrad Gleason 5. Interessanterweise beeinflusst die Klassifizierung nach dem Durchschnittswert aller Stanzen auch die Häufigkeitsverteilung der Gleason-Muster mit insbesondere hohem Anteil der Gleason-Muster 3 + 3 und 3 + 4. Die Anzahl der Patienten (n) mit einem durchschnittlichen Gleason 4 + 3 bzw. Gleason 4 + 4 im Biopsie-Befund ist ungefähr halb so groß (n=251) wie diejenigen, die in mindestens einer Stanze einen derartigen Befund haben (n=562). Dafür ist die Zahl der Patienten mit noch relativ gutartigen 3 + 4 Karzinomen in der Biopsie, welche eine günstigere Risikobewertung zulassen, im Vergleich um 29 % höher. Patienten mit einem durchschnittlichem Gleason 4 + 3 oder 4 + 4 Karzinom in der Stanzbiopsie haben entsprechend deutlich ungünstigere Befundkonstellationen in den radikalen Prostatektomien als diejenigen mit einem Maximalbefund von 4 + 3 oder 4 + 4. Nur 35 % der Patienten mit durchschnittlichem Gleason 4 + 3 aber fast 60 % derjenigen mit einem maximalen Gleason 4 + 3 in der Prostatastanzbiopsie haben in der nachfolgenden Prostatektomie einen Befund eines Gleason 3 + 4 Karzinoms. Ein noch relativ günstiges Gleason 3 + 4 Karzinom in der RPE befindet sich bei nur 5 % der Patienten mit Gleason 4 + 4 als Durchschnittswert im Stanzbiopsie-Befund aber immerhin bei fast 15 % der Patienten mit einem maximalen Gleason 4 + 4.

### 3.2 Quantitativer Gleason-Grad

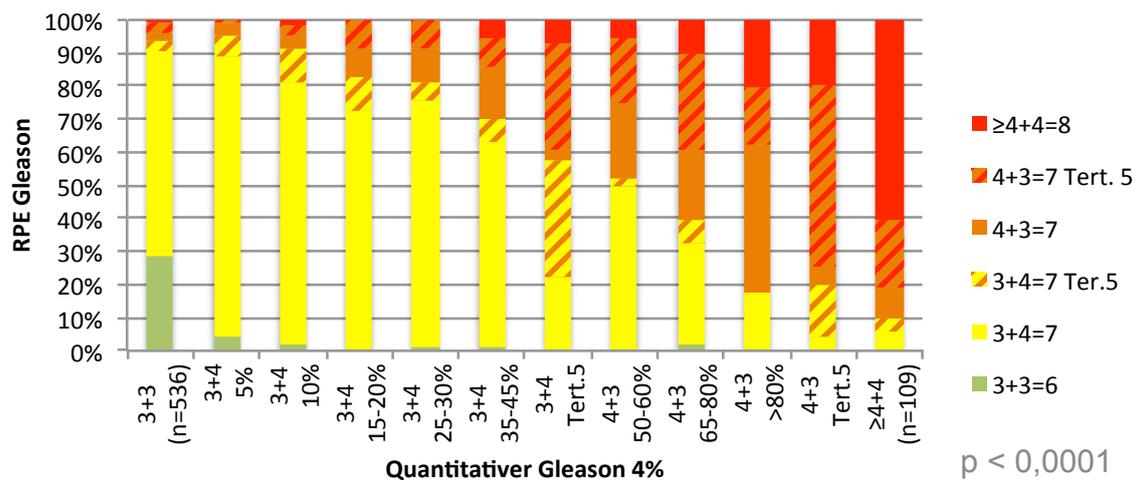


Abbildung 6 Vgl. Quantitativer Gleason-Grad 4 % in der Stanzbiopsie vs. Gleason-Score in der RPE.

Die Ausdehnung der Klassifikation auf den prozentualen Anteil ungünstiger Gleason-Teile ergibt eine noch subtilere Vorhersage der Befunde in der Prostatektomie (Abbildung 6). Hier zeigt sich, dass die Chance eines ungünstigen Tumors in der RPE mit entweder Gleason 5 oder führendem Gleason 4 bei Tumoren mit weniger als 5 % Gleason 4 Anteilen in den Stanzen, sich nicht wesentlich von den Befunden in Gleason 3 + 3 Stanzen unterscheidet. 90 % dieser Patienten haben einen günstigen Befund in der Prostatektomie mit einem maximalen Gleason 3 + 4 = 7. Die Betrachtung der weiteren Befundkonstellationen zeigt, dass ungünstige Prostatektomie-Befunde graduell mit dem prozentualen Anteil von Gleason 4 in den Stanzbiopsien ansteigen. Besonders ungünstig sind die Befunde, wenn neben einem Gleason 4 auch noch Tertiärgrad Gleason 5 gefunden wird. In solchen Fällen findet sich bei Gleason 3 + 4 Tert. 5 Karzinomen in 80 % und bei Gleason 4 + 3 Tert. 5 Karzinomen in 95% der Fälle im Prostatektomie-Befund ein schwerwiegendes Karzinom mit entweder führendem Gleason 4 oder Tertiärgrad Gleason 5.

Um abschließend die Relevanz des quantitativen Gleason-Grades im Verhältnis zur meist praktizierten Klassifikation eines Tumors nach der schlimmsten Stanze besser beurteilen zu können, wurden in der Folge die Ergebnisse des quantitativen Gleason-Grades in Subgruppen untersucht, welche von Tumoren mit einheitlicher „schlimmster Stanze“ gebildet wurden. Diese Untersuchung zeigt, dass der quantitative Gleason-Grad in der Subgruppe der Tumoren mit maximal einem Gleason 3 + 4 relativ geringe Bedeutung hat (Abbildung 7). Das Risiko eines ungünstigen Karzinoms in der RPE ist bei Tumoren mit maximal 5 % Gleason 4 Anteilen allerdings geringer als bei Tumoren mit einem höheren Gleason 4 Anteil. Am größten ist der Anteil ungünstiger Tumoren in der nachfolgenden Prostatektomie,

wenn Gleason 3 + 4 Tumoren mit 35 – 45 % Gleason 4 in der Stanzbiopsie zu finden sind. Hier sind über 30 % der Tumoren ungünstig, im Vergleich zu nur 11 % bei Gleason 3 + 4 Karzinomen mit 5 %.

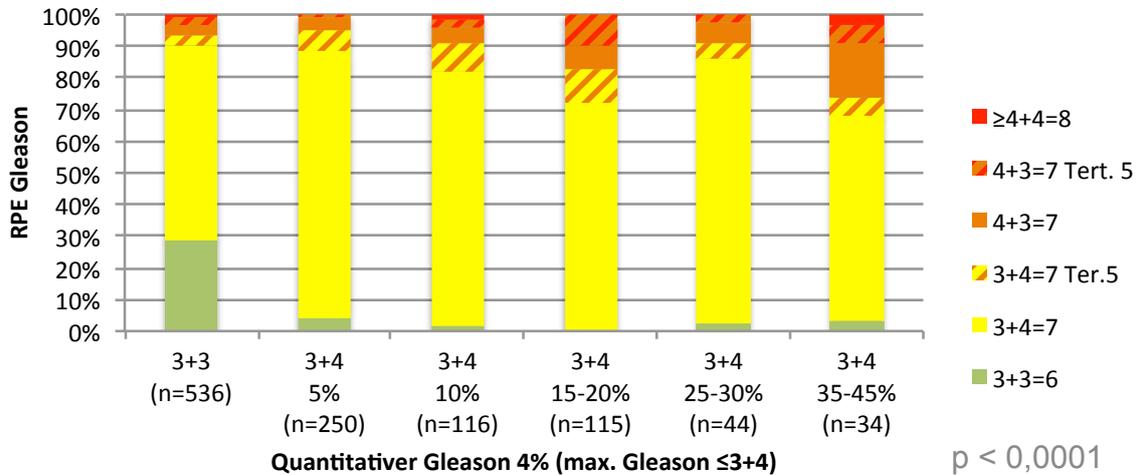


Abbildung 7 Vgl. Quantitativer Gleason-Grad 4 % in der Stanzbiopsie vs. Gleason-Score in der RPE (Subset: Schlimmste Stanze  $\leq 3 + 4$ ).

Wesentlich deutlicher ist der Einfluss des durchschnittlichen Gleason-Grades bei Patienten mit einer „schlimmsten Stanze“ mit einem Gleason-Score  $4 + 3 = 7$  (Abbildung 8). Hier zeigt sich eine graduelle Zunahme der Patienten mit ungünstigem Gleason-Befund in der RPE (Gleason 4-führend oder mindestens Tertiärgrad Gleason 5) von Tumoren mit 5 - 20 % Gleason 4 Anteilen zu Tumoren mit über 50 % oder gar über 80 % Gleason 4 Anteilen.

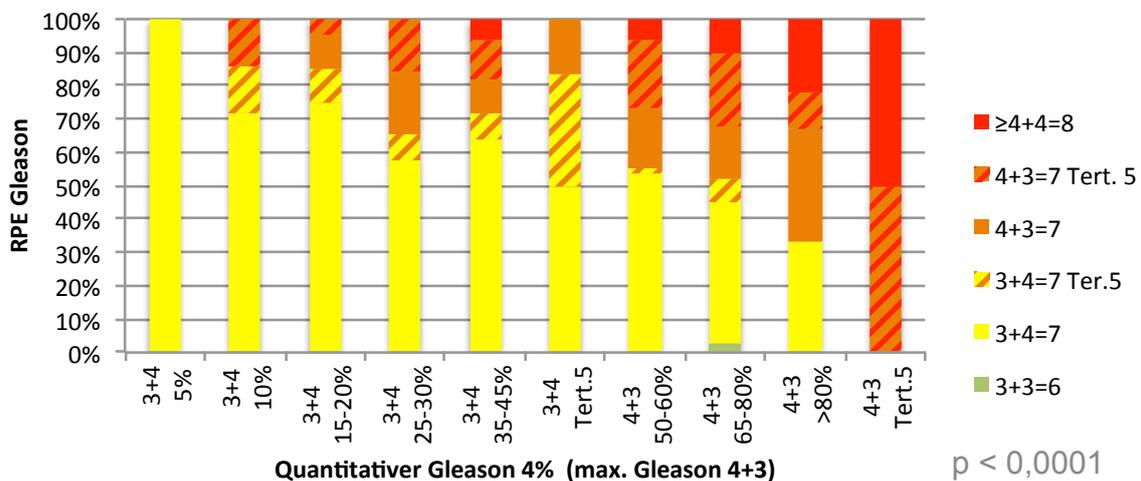


Abbildung 8 Vgl. Quantitativer Gleason-Grad 4 % in der Stanzbiopsie vs. Gleason-Score in der RPE (Subset: Schlimmste Stanze  $4 + 3$ ).

Und auch in der kleinen Gruppe von Patienten mit einer „schlimmsten Stanze“ und einem Gleason-Score von mindestens 8 im Biopsiebefund erwies sich der durchschnittliche Gleason-Grad als hochsignifikant mit dem Prostatektomie-Befund assoziiert (Abbildung 9). So haben die meisten Patienten (72 %) in dieser Gruppe auch im Durchschnitt hohe Gleason 4 Anteile von meist über 65 % im Prostatastanzbiopsiebefund, gab es doch eine kleine Gruppe von Patienten (8 %; bei max. 15 – 20 % Gleason 4), die trotz „maximaler Stanze“ mit einem Gleason  $\geq 8$ , einen relativ niedrigen prozentualen Anteil von Gleason 4 Anteilen in ihrem Tumor aufwiesen. Von diesen (wenigen) Patienten zeigten viele (78,6%) in ihrer Prostata einen insgesamt noch eher günstigen Befund mit einem beispielsweise Gleason 3 + 4 Karzinom.

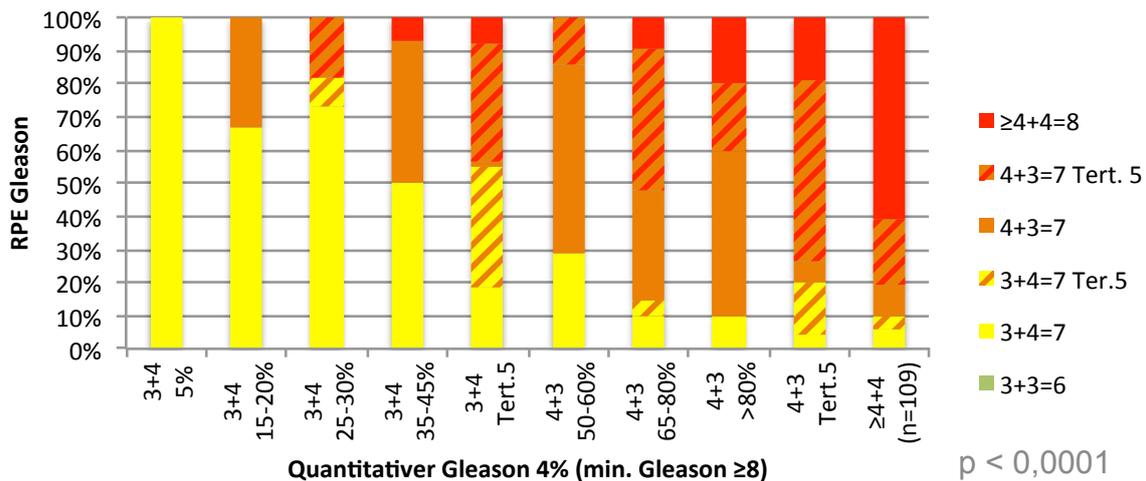


Abbildung 9 Vgl. Quantitativer Gleason-Grad 4 % in der Stanzbiopsie vs. Gleason-Score in der RPE (Subset: Schlimmste Stanze  $\geq 8$ ).

## 4. Diskussion

### 4.1 Allgemeines

Die Befunde dieser Studien zeigen die Überlegenheit des quantitativen Gleason-Grades gegenüber dem traditionellen „kategorisierenden“ Gleason-Grad. Sie zeigen auch, dass bei verschiedenen Biopsie-Befunden in mehreren tumorbehafteten Stanzzyllindern die Kategorisierung nach dem „schlimmsten“ Gleason-Score nicht optimal für die Therapieplanung eines Patienten sein dürfte.

In der vorliegenden Studien wurden zwei wesentliche Fragen verfolgt. Frage Nummer 1 war, ob ein Patient mit verschiedenen Gleason-Graden in mehreren

tumorhaltigen Stanzbiopsien nach dem schlimmsten oder dem durchschnittlichen Gleason-Grad kategorisiert werden soll. Frage Nummer 2 war, ob eine graduelle Quantifizierung der Gleason 4 Anteile bei Tumoren mit Gleason 7 Befunden eine bessere prädiktive Bedeutung haben könnte als die bisher übliche Kategorisierung der Tumore in 3 + 4 oder 4 + 3.

#### 4.2 Durchschnittlicher Gleason-Grad

Die Frage nach der Einordnung eines Patienten nach dem maximalen oder dem durchschnittlichen Gleason Grad ist von hoher praktischer Bedeutung, da alle angewandten prädiktiven Normogramme nur einen Gleason-Grad pro Patienten akzeptieren. In der Praxis scheint es so zu sein, dass die meisten Urologen, beispielsweise 81 % der US-amerikanischen Urologen, ihre Patienten nach dem maximalen Gleason-Grad klassifizieren [48]. Auch die Urologie der Martini-Klinik Hamburg-Eppendorf empfiehlt in einem im Internet zur Verfügung gestellten Risikoberechnungs-Normogramm die Angabe des schlimmsten Gleason-Grades [49].

Ein direkte Vergleich der traditionellen kategorisierenden Gleason-Scores ergab eindeutig eine Überlegenheit des Durchschnittswertes im Vergleich zum Maximalwert. Dieser Unterschied war weniger deutlich bei Tumoren, die durch beide Methoden als Gleason 3 + 4 = 7 klassifiziert wurden, aber deutlich evidenter bei Karzinomen mit einer abschließenden Stanzbiopsie-Beurteilung als Gleason 4 + 3 = 7. Und in dieser wichtigen Untergruppe lag der prozentuale Anteil von Tumoren, die im Prostatektomiepräparat noch als Gleason 3 + 4 = 7 gewertet wurden, im Fall der maximalen Stanze bei 57 % (Abb. 5 A), im Fall des Durchschnittswertes aber nur bei 35,2 % (Abb. 5 B). Die in dieser Studie dargelegte Überbewertung des Stanzbiopsie-Gleasons im Vergleich zur RPE in fast 2/3 der Fälle durch Kategorisierung nach dem Maximalbefund hat sensible Auswirkungen auf die Therapioptionen der Patienten, da konservative Behandlungsstrategien i.d.R. einen Stanzbiopsie-Befund mit führendem Gleason 3 voraussetzen. Die Verwendung des durchschnittlichen Gleason zur Erhöhung des prognostischen Werts ist dringend zu empfehlen, um die Gefahr der Übertherapie von nicht behandlungsbedürftigen Prostatakarzinomen in der Praxis zu reduzieren.

Insgesamt liegt der prozentuale Anteil der Patienten mit entweder einem Gleason 3 + 3 oder 3 + 4 bei Klassifikation nach dem durchschnittlichem Gleason bei 76 %, im Vergleich zu nur 66 % bei Verwendung der schlimmsten Stanze. Daher darf angenommen werden, dass bei Klassifikation nach der „schlimmsten“ Stanze in der Biopsie zu einer „Grad-Verschiebung“ der Gleason-Werte führt, mit insbesondere steigendem Anteil der höheren Gleason-Grade  $\geq 4 + 4$ . Eine interessante Studie von Kuroiwa et. al. verifiziert die Daten der vorliegenden Arbeit in sofern, dass in einem direkten Vergleich der kategorisierenden Gleason-Scores die Klassifizierung nach

dem höchsten Gleason-Grad in 565 retrospektiv überarbeiteten Prostata-Stanzbiopsien in 33 % der Fälle zu einer Erhöhung der Gleason-Werte führte [50]. Für den klinischen Nutzen ist eine möglichst genaue Übereinstimmung der Daten aus der Biopsie vs. RPE entscheidend. Studien, welche die prognostische Bedeutung der „ungünstigsten Stanze“ und dem „Durchschnittswert“ bei Patienten mit multiplen verschiedenen Gleason-Graden vergleichend gegenübergestellt hatten, sind spärlich. Kunz et al. empfiehlt in einer 2003 erschienenen Studie lediglich, dass Patienten mit multiplen, voneinander differierenden Gleason-Scores, beispielsweise einem Gleason-Score von 4 + 4 in einer Stanze und einem Gleason-Muster 3 in einer anderen Stanze einen abschließenden Gleason-Score von 4 + 4 erhalten sollten, anstatt einen Gleason-Score 4 + 3 [51]. Die Studienpopulation, in dieser einzigen mir bekannten Untersuchung zu dieser Fragestellung, ist mit 127 Probanden allerdings deutlich geringer als die der vorliegenden Arbeit.

#### 4.3 Quantitativer Gleason-Grad

Die besondere Bedeutung der prozentualen Anteile von Gleason-Graden in Stanzbiopsien ist das herausragende Ergebnis dieser Studie. Intuitiv war zu erwarten gewesen, dass Gleason 3 + 4 Karzinome, bei denen der Pathologe einzelne fragliche Befunde sicherheitshalber zu einem Gleason 4 kategorisiert hat und Tumoren, welche ausgedehnte Gleason 4 Anteile beinhalten und „fast“ die Hürde zum Gleason 4 + 3 = 7 Karzinom geschafft hätten sich biologisch unterscheiden. Die große Zahl, der in unserem Kollektiv vorliegenden Fälle erlaubt nun eine eindrucksvolle Verifizierung dieser erwarteten Befunde. Vor allem bei der Analyse aller in die Studie eingeschlossener Tumoren zeigt sich eine eindrucksvolle, kontinuierliche, graduelle Zunahme der ungünstigen Gleason-Befunde im Prostatektomiepräparat mit zunehmenden Gleason 4 Anteil in der Stanze. Das ähnliche graduelle Zunahmen ungünstiger Befunde auch innerhalb der heute üblichen klinischen Gruppen gefunden wurden, welche nach dem „maximalen Gleason-Grad“ gebildet wurden, unterstreicht eindrucksvoll die Bedeutung des quantitativen Gleason-Gradings. Nur wenige Autoren hatten früher den Einfluss prozentualer Anteile von Gleason 4 auf klinische Parameter untersucht. Der prognostische Wert der prozentualen Ausbreitung von Gleason-Muster 4 und 5 wurde erstmalig in einer Studie von Stamey et al. erfasst und beschrieben (1999) [52]. Die besondere Bedeutung der prozentualen Anteile von Gleason-Muster 4 zur präziseren Vorhersage des tatsächlichen Gleason-Scores der Ektomie fand zum ersten Mal in einer Studie von Reese et al. im Jahr 2012 Erwähnung, allerdings blieb in dieser Studie die Empfehlung zur Quantifizierung des Gleason-Scores nur auf den Gleason-Score 7 beschränkt, da keine Aussage über Tertiär-Grade sowie Gleason-Scores  $\geq 8$  getroffen werden konnte [53].

Insgesamt zeigen die Befunde dieser Studie die große Bedeutung der prozentualen Anteile der unterschiedlichen Gleason-Grade in Stanzbiopsien. Diese Ergebnisse sind nicht überraschend, gelten doch Gleason 3 Anteile als generell günstig und Gleason 4 Anteile als ungünstig bei Prostatakarzinomen. Es verwundert deshalb nicht, dass Tumoren mit einem höheren Anteil an ungünstigem Gleason sich weniger gutartig verhalten als andere Tumoren. Dabei ist es auch nicht überraschend, dass die biologische Aggressivität des Tumors kein „An/Aus-Phänomen“ ist, welches exakt bei einem Grenzwert von 50 % eintritt, so dass sich alle Gleason 4 + 3 Karzinome identisch maligne und alle Gleason 3 + 4 Karzinome identisch weniger maligne verhalten. Dabei ist es auch wichtig zu verstehen, dass es für den Pathologen nicht immer einfach ist, in jedem einzelnen Gesichtsfeld den Gleason-Grad exakt zu bestimmen. Es gibt hinlänglich Studien, bei denen gezeigt wurde, dass hohe Interobserver-Variationen hinsichtlich der genauen Einstufung des Gleason-Grades auftreten können (s. Tabelle 3). Aufgrund derartiger Studien wurde auch schon vorgeschlagen, dass Pathologen ihr Gleason-Grading zusammen mit einer Angabe zur Sicherheit der Diagnose abgeben sollen. Es erscheint aber fraglich, wie Urologen dann mit der Situation umgehen sollen, wenn sie bei sechs Stanzbiopsien fünf tumorbefallene Regionen haben, in denen sie nicht nur fünf verschiedene Gleason-Grade, sondern auch noch fünf verschiedene Sicherheitsfaktoren dazu bekommen. Im Gegensatz dazu ist die Angabe der Gleason Prozentanteile, sowohl ein eindeutiger in Nomogramme einbringbarer Parameter, als auch ein indirekter Marker für die Sicherheit einer bestimmten Gleason-Graduierung. Falls der Gleason 4 Anteil 5 % oder 10 % beträgt, muss davon ausgegangen werden, dass nur dezente Anteile des Tumors einem Gleason 4 Grad zugeordnet werden können. Es ist davon auszugehen, dass bei derartigen Tumoren einige andere Pathologen auch einen Gleason-Grad von  $3 + 3 = 6$  diagnostiziert hätten. Auf der anderen Seite ist ein Tumor mit einem Gleason-Grad von  $3 + 4 = 7$  und 40 % Gleason 4 offenbar ein Grenzfall zu einem Gleason 4 + 3 und es ist auch hier davon auszugehen, dass einige Pathologen diesen Tumor mit Gleason 4 + 3 = 7 klassifiziert hätten. Vermutlich mit einem Gleason-Grad-4 Prozentanteil von 50 – 60 % oder maximal 70 %.

## 5. Zusammenfassung

Der Gleason-Score ist der heute am weitverbreiteste morphologische Prognose-Parameter beim Prostatakarzinom. Allerdings ist die Einschätzung des weiteren Krankheitsverlaufes insbesondere bei „mittleren“ Gleason Kategorien (3 + 4 und 4 + 3) schwierig. Dazu kommt, dass die Bestimmung des Gleason Grades an der Stanzbiopsie oft ungenau ist, was sich in einer schlechten Übereinstimmung des Gleason Grades in der Stanze mit der Ektomie zeigt. Die Bestimmung des Gleason-Scores hat jedoch weitreichende Folgen für die Therapieentscheidung. Insbesondere die Entscheidung zwischen der „Active Surveillance“ Strategie und einer invasiven Therapie (z.B. radikale Prostatektomie) basiert unter anderem auf dem Gleason-Score. Aus diesem Grund ist es essentiell, die Präzision des Gleason-Scorings weiter zu verbessern.

Ziel der vorliegenden Studie war es daher zu prüfen, ob der prozentuale Gleason-Anteil in allen Biopsiepräparaten eine bessere Übereinstimmung mit dem tatsächlichen Gleason-Score im Ektomiepräparat erzielt als der heute meist verwendete „schlimmste“ Gleason-Score in der Stanzbiopsie. Dazu wurden die prozentualen Gleason-Anteile in allen Stanzbiopsien von 1657 Patienten bestimmt und mit dem Gleason-Score in den dazugehörigen Prostatektomien verglichen. Grundlage dafür war die seit 2005 am Institut für Pathologie des UKE etablierte Tumordatenbank, in der für jede Stanzbiopsie der Tumorbefall in Millimetern und der jeweilige Gleason Grad in Millimetern erfasst wurde. Die Ergebnisse der vorliegenden Arbeit zeigen eindrucksvoll, dass eine Mittlung des Gleason-Scores über alle Stanzbiopsien zur Prognose des finalen Gleason-Scores in der Prostatektomie aussagekräftiger ist, als die dem bisherigen Usus folgendene Verwendung des schlimmsten Stanzbefundes. In der wichtigen Untergruppe der Tumoren mit einem Gleason-Score von  $4 + 3 = 7$  in der „schlimmsten“ Stanze, lag bei 57 % der Patienten nur ein Gleason-Score von  $3 + 4 = 7$  im Prostatektomiepräparat vor. Im Fall des durchschnittlichen Gleason-Scores ist der prozentuale Anteil der fehlerhaft zu hoch eingeschätzten Fälle mit 35,2 % hingegen deutlich geringer. Insgesamt ist der Anteil der Patienten mit einem niedrigen Gleason-Score (3 + 3 oder 3 + 4) bei Klassifikation nach dem durchschnittlichen Gleason mit 76 % etwas höher als mit 66 % bei Verwendung der schlimmsten Stanze. Folglich führt die Klassifikation nach der „schlimmsten“ Stanze in der Biopsie zu einer „Grad-Verschiebung“ mit insbesondere steigendem Anteil der höheren Gleason-Grade  $\geq 4 + 4$ . Die zusätzliche Quantifizierung des Gleason 4 Anteils lässt eine noch subtilere Voraussage des nachfolgenden Gleason-Grades in der radikalen Prostatektomie zu. Vor allem bei der Analyse aller in die Studie eingeschlossener Tumoren zeigt sich eine eindrucksvolle und stetige Zunahme der ungünstigen Gleason-Befunde im Prostatektomiepräparat mit zunehmendem Gleason 4 Anteil in der Stanze. Dass ähnliche graduelle Zunahmen ungünstiger Befunde auch innerhalb der heute üblichen klinischen Gruppen gefunden wurden, welche nach dem „schlimmsten Gleason-Grad“ gebildet wurden, unterstreicht die Bedeutung des

quantitativen Gleason-Gradings. Im Hinblick auf die Problematik der häufig auftretenden Interobserver-Variabilität bleibt es zukünftigen Studien vorbehalten, ob die prozentuale Angabe der Gleason Anteile - als indirekter Marker für die Sicherheit der Diagnose - auch einen positiven Einfluss auf die Reproduzierbarkeit zeigen kann.

## 6. Abstract

Today, the Gleason score is the strongest prognostic parameter in prostate cancer. However, estimation of the risk for progression remains difficult especially in "moderate" grade tumors (3 + 4 and 4 + 3). In addition, determination of the Gleason score is subject to interobserver variances, which can be seen by the low overall agreement between the score of biopsy and subsequent radical prostatectomy specimens. A more subtle risk prediction using quantitative aspects of the Gleason grading is of particular interest in moderate grade tumors as these patients have a wide variety of treatment options ranging from active surveillance to radical prostatectomy. It is, therefore, of crucial importance to assess the Gleason score in biopsy specimens as accurate as possible.

The aim of this study was to evaluate whether the average Gleason score across all biopsies yields a better concordance with the definite Gleason score in radical prostatectomy specimen than the Gleason score of the worst biopsy, which is used by many pathologist. For this purpose, percentage of Gleason patterns across all biopsies was determined and compared to the Gleason score of the radical prostatectomy in a cohort of 1657 patients. Data for the analysis have been prospectively collected since 2005 at the Institute of Pathology at the University Medical Center Hamburg-Eppendorf. In summary, these data demonstrate impressively that the average Gleason score across all prostate biopsies is better correlated with the final pathological Gleason score of the radical prostatectomy than the Gleason score of the worst biopsy. In the important subgroup of tumors with a Gleason score of 4 + 3 = 7 in the "worst" biopsy, there was only a Gleason score of 3 + 4 = 7 in 57 % of prostatectomy specimens. If the average Gleason score was used, the percentage of erroneously overestimated cases was distinctly smaller (35.2 %). Overall, the proportion of patients with a low Gleason score (3 + 3 or 3 + 4) according to the average Gleason score (76 %) was slightly higher as if the worst biopsy Gleason was used (66%). Consequently, classification according to the Gleason of the worst biopsy leads to a shift in tumor stage and an increasing proportion of cancers with higher Gleason grade ( $\geq 4 + 4$ ). The additional quantification of the Gleason 4 percentage allows for an even more subtle prediction of the subsequent Gleason grade in a radical prostatectomy. Especially in the analysis of all tumors in the study, an impressive and steady increase in unfavorable Gleason findings is shown in prostatectomy specimens with an increasing Gleason 4 proportion in the biopsy specimen.

As similar gradual increases in adverse findings were also found within actual clinical groups, which were defined by the "worst Gleason grade", underlines the importance of quantitative Gleason grading. It remains to be shown in future studies whether the percentage of Gleason patterns - as an indirect marker for diagnostic accuracy - can also show a positive effect on the reproducibility.

## 7. Abkürzungsverzeichnis

B: Stanzbiopsie

ISUP: International Society of Urological Pathology

$\kappa$ -Wert: Cohens Kappa

n: Anzahl

PSA: Prostaspezifisches Antigen

RPE: radikale Prostatektomie

Tert.: Tertiärgrad

UKE: Universitätsklinikum Hamburg-Eppendorf

WHO: World Health Organisation

## 8. Tabellenverzeichnis

Tabelle 1: Graduierung nach Gleason

Tabelle 2: Übersicht früherer Studien zur Interobserver-Variabilität (1986 - 2013)

Tabelle 3: Interobservervariabilität in 15 ausgewählten Konsensus-Fällen

## Abbildungsverzeichnis

Abbildung 1: Das konventionelle und modifizierte Gleason-Grading System im Vergleich

Abbildung 2: Pathohistologie Prostatagewebe: Gleason-Score  $5 + 3 = 8$

Abbildung 3: Verhältnis von Gleason-Muster 4/5 in Prozent zur PSA-Rezidiv-Rate

Abbildung 4: Exemplarische Skizzierung eines Prostata-Schnittpräparates am Institut für Pathologie, Universitätsklinikum Hamburg-Eppendorf

Abbildung 5: Vergleich der Gleason-Beurteilungen an Stanzbiopsien versus radikale Prostatektomien bei Klassifizierung des Stanzbefundes nach der maximalen Stanze (A) und nach dem Durchschnittswert ( $\emptyset$ ) aller Stanzen (B).

Abbildung 6: Vergleich zwischen quantitativem Gleason-Grad 4% in der Stanzbiopsie versus dem Gleason-Score in der RPE

Abbildung 7: Vergleich zwischen quantitativem Gleason-Grad 4% in der Stanzbiopsie versus dem Gleason-Score in der RPE (Subset: Schlimmste Stanze  $\leq 3+4$ )

Abbildung 8: Vergleich zwischen quantitativem Gleason-Grad 4% in der Stanzbiopsie versus dem Gleason-Score in der RPE (Subset: Schlimmste Stanze  $4+3$ )

Abbildung 9: Vergleich zwischen quantitativem Gleason-Grad 4% in der Stanzbiopsie versus dem Gleason-Score in der RPE (Subset: Schlimmste Stanze  $\geq 8$ )

## 9. Literaturverzeichnis

1. Gleason, D. F. (1966). "Classification of prostatic carcinomas." Cancer Chemother Rep **50**(3): 125-128.
2. Gleason DF, Mellinger GT; Veterans Administration Cooperative Urological Research Group. Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. J Urol. 1974; 111; 58–64.
3. Delahunt, B., et al. (2012). "Gleason grading: past, present and future." Histopathology 60(1): 75-86.
4. 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.
5. Gleason DF. Histologic grading of prostate cancer: a perspective. Hum. Pathol. 1992; 23; 273–279.
6. Epstein, J. I., et al. (2005). "The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma." Am J Surg Pathol **29**(9): 1228-1242.
7. Bostwick, D. G. (1994). "Grading prostate cancer." Am J Clin Pathol **102**(4 Suppl 1): S38-56
8. Epstein, J. I. (2010). "An update of the Gleason grading system." J Urol **183**(2): 433-440.
9. Hoedemaeker RF, Rietbergen JB, Kranse R, Schroder FH, van der Kwast TH: Histopathological prostate cancer characteristics at radical prostatectomy after population based screening. J Urol 2000; 164(2): 411–15.
10. Catalona, W. J., et al. (1995). "Evaluation of percentage of free serum prostate-specific antigen to improve specificity of prostate cancer screening." JAMA **274**(15): 1214-1220.
11. Luboldt, H. J., et al. (1999). "[Early recognition of prostate carcinoma. Initial results of a prospective multicenter study in Germany. Project Group for Early Detection DGU-BDU Laboratory diagnosis Professional Circle]." Urologe A **38**(2): 114-123.
12. Interdisziplinäre Leitlinie der Qualität S3 zur Früherkennung, Diagnose und Therapie der verschiedenen Stadien des Prostatakarzinoms, 2014
13. Epstein, J. I. (2010). "An update of the Gleason grading system." J Urol **183**(2): 433-440.
14. Eble JN, Sauter G, Epstein JI, et al. *Pathology and Genetics of Tumours of the Male Genitourinary Organs*. Lyon, France: IARC Press; 2004, S.179-182.
15. Pan, C. C., et al. (2000). "The prognostic significance of tertiary Gleason patterns of higher grade in radical prostatectomy specimens: a proposal to modify the Gleason grading system." Am J Surg Pathol **24**(4): 563-569.
16. Srigley, J. R., et al. (2000). "Updated protocol for the examination of specimens from patients with carcinomas of the prostate gland: a basis for checklists. Cancer Committee." Arch Pathol Lab Med 124(7): 1034-1039.

17. Anderson PR, Hanlon AL, Patchefsky A et al. Perineural invasion and Gleason 7–10 tumors predict increased failure in prostate cancer patients with pretreatment PSA <10 ng / ml treated with conformal external beam radiation therapy. *Int. J. Radiat. Oncol. Biol. Phys.* 1998; 41; 1087–1091
18. Ennis RD, Malyszko BK, Rescigno J et al. Biologic classification as an alternative to anatomic staging for clinically localized prostate cancer; a proposal based on patients treated with external beam radiotherapy. *Urology* 1998; 51; 265–270.
19. D'Amico AV, Whittington R, Malkowicz SB et al. A multivariate analysis of clinical and pathological factors that predict for prostate specific antigen failure after radical prostatectomy for prostate cancer. *J. Urol.* 1995; 154; 131–138.
20. D'Amico AV, Whittington R, Malkowicz SB et al. The combination of preoperative prostate specific antigen and postoperative pathological findings to predict prostate specific antigen outcome in clinically localized prostate cancer. *J. Urol.* 1998; 160; 2096–2101.
21. Carvalhal G, Humphrey PA, Thorson P et al. Visual estimate of percentage of cancer is an independent predictor of prostate cancer recurrence after radical prostatectomy. *Cancer* 2000; 89; 1308–1314.
22. Koch MO, Foster RS, Bell S et al. Characterization and predictors of prostate specific antigen progression rates after radical prostatectomy. *J. Urol.* 2000; 164; 749–753.
23. Sgrignoli AR, Walsh PC, Steinberg GD et al. Prognostic factors in men with stage D1 prostate cancer; identification of patients less likely to have prolonged survival after radical prostatectomy.
24. Cheng L, Seho RJ, Slezak J et al. Predictors of survival for prostate carcinoma patients treated with salvage radical prostatectomy after radiation therapy. *Cancer* 1998; 83; 2164–2171
25. Humphrey PA. Gleason grading and prognostic factors in carcinoma of the prostate. *Mod Pathol.* 2004;17:292–306.
26. Norberg M, Holmberg L, Wheeler T et al. Five year follow-up after radical prostatectomy for localized prostate cancer: a study of impact of different tumor variables on progression. *Scand. J. Urol. Nephrol.* 1994; 28; 391–399.
27. Epstein JI, Partin AW, Sauvageot J et al. Prediction of progression following radical prostatectomy: a multivariate analysis of 721 men with long-term follow-up. *Am. J. Surg. Pathol.* 1996; 20; 286–292.
28. Han M, Partin AW, Pound CR et al. Long-term biochemical disease-free and cancer-specific survival following radical retropubic prostatectomy. The 15-year Johns Hopkins experience. *Urol. Clin. North Am.* 2001; 281; 555–565.
29. Cheng L, Koch MO, Juliar BE, et al. The combined percentage of Gleason patterns 4 and 5 is the best predictor of cancer progression after radical prostatectomy. *J Clin Oncol.* 2005;23:2911–2917.
30. Tefilli MV, Gheiler EL, Tiguert R et al. Prognostic indicators in patients with seminal vesicle involvement following radical prostatectomy for clinically localized prostate cancer. *J. Urol.* 1998; 160; 802–806.

31. Jhaveri FM, Zippe CD, Klein EA et al. Biochemical failure does not predict overall survival after radical prostatectomy for localized prostate cancer: 10-year results. *Urology* 1999; 54; 884–890.
32. Sweat SD, Bergstralh EJ, Slezak J et al. Competing risk analysis after radical prostatectomy for clinically nonmetastatic prostate adenocarcinoma according to clinical Gleason score and patient age. *Urology* 2002; 168; 525–529.
33. Eggener SE, Scardino PT, Walsh PC et al. Predicting 15-year prostate cancer specific mortality after radical prostatectomy. *J. Urol.* 2011; 185; 869–875.
34. Andren O, Fall K, Fransen L, Andersson SO, Johansson JE, Rubin MA. How well does the Gleason score predict prostate cancer death? A 20-year follow up of a population based cohort in Sweden. *J. Urol.* 2006; 175; 1337–1340.
35. J. R. Landis, G. G. Koch: *The measurement of observer agreement for categorical data*. In: *Biometrics*. 33, 1977, 159–174
36. Helpap, B. and L. Egevad (2006). "The significance of modified Gleason grading of prostatic carcinoma in biopsy and radical prostatectomy specimens."
37. Epstein JI. *The Gleason Grading System: A Complete Guide for Pathologists and Clinicians*. Philadelphia, PA: Lippincott Williams & Wilkins; 2013
38. Burchardt, M., et al. (2008). "Interobserver reproducibility of Gleason grading: evaluation using prostate cancer tissue microarrays." *J Cancer Res Clin Oncol* **134**(10): 1071-1078.
39. King CR (2000) Patterns of prostate cancer biopsy grading: trends and clinical implications. *Int J Cancer* 90:305–311
40. Kronz JD, Silberman MA, Allsbrook WC (2000) A web-based tutorial improves practicing pathologists' Gleason grading of images of prostate carcinoma specimens obtained by needle biopsy: validation of a new medical education paradigm. *Cancer* 89:1818–1823
41. Egevad L, Norlen BJ, Norberg M (2001) The value of multiple core biopsies for predicting the Gleason score of prostate cancer. *BJU Int* 88:716–721
42. Mikami Y, Manabe T, Epstein JI, Shiraishi T, Furusato M, Tsuzuki T, Matsuno Y, Sasano H (2003) Accuracy of Gleason grading by practicing pathologists and the impact of education on improving agreement. *Hum Pathol* 34(7):658–665
43. Smith EB, Frierson HF Jr, Mills SE, Boyd JC, Theodorescu D (2002) Gleason scores of prostate biopsy and radical prostatectomy specimens over the past 10 years: is there evidence for systematic upgrading? *Cancer* 94:2282–2287
44. Egevad, L., et al. (2013). "Standardization of Gleason grading among 337 European pathologists." *Histopathology* **62**(2): 247-256.
45. Danneman, D., et al. (2014). "Gleason Inflation 1998-2011. A Registry Study of 97 168 Men." *BJU Int*.
46. Cooperberg MR, Cowan JE, Hilton JF, et al. Outcomes of active surveillance for men with intermediate-risk prostate cancer. *J Clin Oncol* 2011;29:228–34.
47. Ploussard G, Isbarn H, Briganti A, et al. Can we expand active surveillance criteria to include biopsy Gleason 3+4 prostate cancer? A multi-institutional study of 2,323 patients. *Urol Oncol* 2015;33, 71.e1–9.

48. Rubin MA, Bismar TA, Curtis S et al: Prostate needle biopsy reporting: how are the surgical members of the Society of Urologic Oncology using pathology reports to guide treatment of prostate cancer patients? *Am J Surg Pathol* 2004; 28: 946
49. <https://www.mskcc.org/nomograms/prostate/pre-op>, 20.12.2015
50. K. Kuroiwa, H. Uchino, A. Yokomizo, et al. Impact of reporting rules of biopsy Gleason score for prostate cancer *J Clin Pathol*, 62 (2009), p. 260
51. Kunz GM Jr, Epstein JI: Should each core with prostate cancer be assigned a separate Gleason score? *Hum Pathol* 2003, 34 :911-914
52. Stamey TA, McNeal JE, Yemoto CM et al: Biological determinants of cancer progression in men with prostate cancer. *JAMA* 1999; 281: 1395.
53. Reese AC, Cowan JE, Brajtbord JS et al: The quantitative Gleason score improves prostate cancer risk assessment. *Cancer* 2012; 118: 6046.

## **10. Danksagung**

Ich danke meinem Doktorvater Prof. Dr. med. G. Sauter für die Ermöglichung und die außerordentliche Unterstützung bei der Fertigstellung dieser Arbeit.

Desweiteren danke ich denjenigen, die mich während der Anfertigung der Dissertation betreut haben: Frau Dr. Martina Kluth, Herr PD. Dr. Simon, Frau Christina Koop.

Zum Schluss danke ich meiner Familie, Annette, Peter, Lisa, Nicolai und Ömer.

**11. Lebenslauf**

## **12. 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.

Ich erkläre mich einverstanden, dass meine Dissertation vom Dekanat der Medizinischen Fakultät mit einer gängigen Software zur Erkennung von Plagiaten überprüft werden kann.

Unterschrift: .....