

Aus dem
CharitéCentrum für Diagnostische und Interventionelle Radiologie und Nuklearmedizin
Klinik für Nuklearmedizin
Arbeitsbereich Klinische Nuklearmedizin
Leiter: Prof. Dr. med. Holger Amthauer

Habilitationsschrift

Optimierung visueller und quantitativer Bildparameter in der FDG-
PET/CT Diagnostik bei Patienten mit nicht-kleinzelligem Lungenkarzinom

zur Erlangung der Lehrbefähigung
für das Fach Experimentelle Nuklearmedizin

vorgelegt dem Fakultätsrat der Medizinischen Fakultät
Charité-Universitätsmedizin Berlin

von

Dr. med. Julian Manuel Michael Rogasch

Eingereicht: Dezember 2020

Dekan: Prof. Dr. med. Axel R. Pries

1. Gutachter/in: Frau Prof. Dr. med. Susanne Klutmann, Hamburg

2. Gutachter/in: Prof. Dr. med. Peter Bartenstein, München

Inhaltsverzeichnis

<i>Inhaltsverzeichnis</i>	2
<i>Abkürzungen</i>	3
1. Einleitung	4
1.1 Das Lungenkarzinom	4
1.1.1 Epidemiologie	4
1.1.2 Das nicht-kleinzellige Lungenkarzinom (NSCLC)	4
1.1.3 Patienten mit potenziell kurativer Therapieintention	6
1.2 FDG-PET/CT Diagnostik.....	7
1.2.1 Prinzip.....	7
1.2.2 Bildqualität: Räumliche Auflösung und Signal-Rausch-Abstand (SNR).....	8
1.2.3 Visuelle Bildbefundung	10
1.2.4 Quantifizierung und quantitative Bildparameter	10
1.3 FDG-PET/CT Diagnostik beim NSCLC.....	15
1.3.1 Prätherapeutische Lymphknotendiagnostik	15
1.3.2 Quantifizierung, Risiko- und Prognoseabschätzung	18
1.4 Eigene Ziele und Fragestellungen	20
1.4.1 Prätherapeutische Lymphknotendiagnostik	20
1.4.2 Quantifizierung, Risiko- und Prognoseabschätzung	20
2. Eigene Arbeiten	21
2.1 Prätherapeutische Lymphknotendiagnostik.....	21
2.1.1 Dual Time Point Imaging für die prätherapeutische Lymphknotendiagnostik.....	21
2.1.2 Interobserver-Übereinstimmung der standardisierten visuellen Befundung	33
2.1.3 Einfluss neuartiger PET-Bildrekonstruktionsverfahren auf die Bildqualität.....	40
2.2 Quantifizierung, Risiko- und Prognoseabschätzung.....	56
2.2.1 Prognostischer Stellenwert der ASP beim NSCLC: Explorative Studie.....	56
2.2.2 Prognostischer Stellenwert der ASP beim NSCLC: Konfirmatorische Studie.....	67
2.2.3 Variabilität der ASP unter dem Einfluss der Bildrekonstruktion	78
3. Diskussion	92
3.1 Prätherapeutische Lymphknotendiagnostik.....	92
3.2 Quantifizierung, Risiko- und Prognoseabschätzung.....	95
4. Zusammenfassung	99
4.1 Prätherapeutische Lymphknotendiagnostik.....	99
4.2 Quantifizierung, Risiko- und Prognoseabschätzung.....	99
5. Literaturangaben	100
<i>Danksagung</i>	107
<i>Erklärung</i>	108

Abkürzungen

ADC	adenocarcinoma (Adenokarzinom)
ASP	asphericity
ATS	American Thoracic Society
AUC	area under the curve
COV	coefficient of variation (Variationskoeffizient)
CT	Computertomographie
DFS	disease-free survival (krankheitsfreies Überleben)
DTPI	Dual Time Point Imaging
EARL	EANM Research Ltd.
EBUS	endobronchialer Ultraschall
EBUS-TBNA	endobronchialer Ultraschall-gestützte transbronchiale Nadelaspiration
ECOG	Eastern Cooperative Oncology Group
ERS	European Respiratory Society
ESMO	European Society for Medical Oncology
FDG	Fluorodesoxyglucose
FWHM	full width at half maximum (Halbwertsbreite)
GE	General Electric
GLCM	grey-level co-occurrence matrix (Grauwertematrix)
GLUT	Glukosetransporter
IASLC	International Association for the Study of Lung Cancer
LOL	linker Oberlappen
LUL	linker Unterlappen
ML	Mittellappen
MTV	metabolic tumor volume (metabolisches Tumolvolumen)
NGTDM	neighbourhood grey-tone difference matrix
NSCLC	nicht-kleinzelliges Lungenkarzinom
OS	overall survival (Gesamtüberleben)
OSEM	ordered subset expectation maximization
PET	Positronen-Emissions-Tomographie
PFS	progression-free survival (progressionsfreies Überleben)
PSF	point spread function (Punktabbildungsfunktion)
RI	Retentionsindex
ROL	rechter Oberlappen
RUL	rechter Unterlappen
SCC	squamous cell carcinoma (Plattenepithelkarzinom)
SNR	signal-to-noise ratio (Signal-Rausch-Abstand)
SUV	standardized uptake value
TBR	tumor-to-background ratio (Läsion-zu-Hintergrund-Verhältnis)
TLG	tumor lesion glycolysis
TOF	time of flight
UICC	Union for International Cancer Control
VATS	videoassistierte Thorakoskopie

1. Einleitung

1.1 Das Lungenkarzinom

1.1.1 Epidemiologie

Das Lungenkarzinom ist in Deutschland (Stand: 2016) die dritthäufigste Krebsart bei Frauen und zweithäufigste Krebsart bei Männern. Aufgrund der geringen 5-Jahres-Überlebensrate von – im Durchschnitt – nur ca. 15-20% ist das Lungenkarzinom zudem die zweithäufigste Krebstodesursache bei Frauen und die häufigste Krebstodesursache bei Männern (1). Von grundlegender Bedeutung ist die histologische Differenzierung zwischen kleinzelligem (ca. 15-20% der Fälle) und nicht-kleinzelligem Lungenkarzinom (ca. 80-85% der Fälle). Beide Subtypen unterscheiden sich entscheidend hinsichtlich prognostischer Betrachtungen, diagnostischer und therapeutischer Algorithmen (2).

1.1.2 Das nicht-kleinzellige Lungenkarzinom (NSCLC)

Soweit nicht anders spezifiziert, beschränken sich alle Ausführungen in den folgenden Kapiteln auf das nicht-kleinzellige Lungenkarzinom (NSCLC).

1.1.2.1 Histologie

Das NSCLC wird histologisch in Adenokarzinome (bis zu 60%), Plattenepithelkarzinome (bis zu 30%) sowie großzellige Karzinome und sonstige Subtypen wie undifferenzierte Karzinome unterteilt (1). Insbesondere bei den Adenokarzinomen werden zudem weitere Subtypen und Wachstumsmuster unterschieden (Tabelle 1), deren aktuelle Klassifikation aus dem Jahre 2011 dem Konsensus der International Association for the Study of Lung Cancer (IASLC) / American Thoracic Society (ATS) / European Respiratory Society (ERS) folgt (3).

Invasives Adenokarzinom <ul style="list-style-type: none"> • Lepidisch prädominant (G1) • Azinär prädominant (G2) • Papillär prädominant (G2) • Mikropapillär prädominant (G3) • Solide prädominant mit Schleimbildung (G3) 	Präinvasiv <ul style="list-style-type: none"> • Atypische adenomatöse Hyperplasie • Adenocarcinoma in situ
	Minimalinvasives Adenokarzinom
	Varianten des invasiven Adenokarzinoms

Tabelle 1: Adenokarzinom-Klassifikation gemäß IASLC/ATS/ERS Konsensus (2011). Nach (3, 4).

1.1.2.2 Metastasierung und Prognose

Die durchschnittliche Überlebensprognose beim NSCLC ist ungünstig, ist jedoch deutlich von der Tumorgröße und -ausbreitung bzw. Metastasierung abhängig, die zusammen das klinische Tumorstadium gemäß TNM-Klassifikation und Union for International Cancer Control (UICC)-Stadium definieren (Tabelle 2).

T/M-Deskriptor	N0	N1	N2	N3
T1a	IA1	IIB	IIIA	IIIB
T1b	IA2	IIB	IIIA	IIIB
T1c	IA3	IIB	IIIA	IIIB
T2a	IB	IIB	IIIA	IIIB
T2b	IIA	IIB	IIIA	IIIB
T3	IIB	IIIA	IIIB	IIIC
T4	IIIA	IIIA	IIIB	IIIC
M1a	IVA	IVA	IVA	IVA
M1b	IVA	IVA	IVA	IVA
M1c	IVB	IVB	IVB	IVB

Tabelle 2: TNM-Klassifikation und UICC-Stadien (8. Edition). Nach (5).

Eine lymphatische Metastasierung wird bei resektablen Tumoren in über 50% der Fälle beobachtet (6). Die anatomische Klassifikation der thorakalen Lymphknotenstationen folgt hierbei dem endobronchialen Ultraschall (EBUS). Unabhängig von der Lokalisation des Primärtumors betrifft die Metastasierung am häufigsten hiläre Lymphknotenstationen. Die Häufigkeit von Metastasen in den mediastinalen Lymphknotenstationen variiert hingegen je nach Primärtumorlokalisation (Abbildung 1) (6, 7). Bei ca. 40% der Patienten mit NSCLC liegt bei Erstdiagnose eine Fernmetastasierung vor (UICC-Stadium IV) (8).

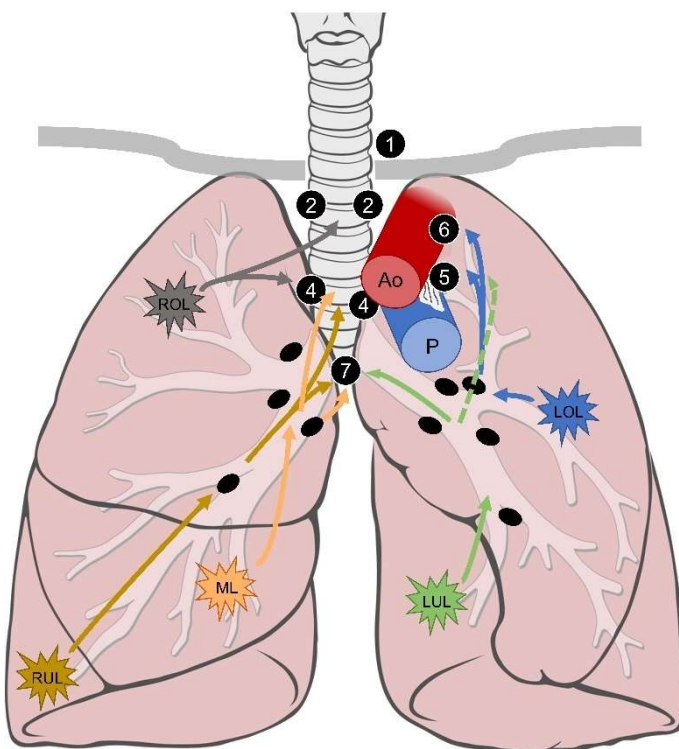


Abbildung 1: Primärtumorlokalisation und mediastinale Lymphknotenmetastasen
Lymphknotenmetastasen sind bei Tumoren des rechten Oberlappens (ROL) im oberen Mediastinum (EBUS 2 und 4) am häufigsten. Primärtumoren im Mittellappen (ML) und rechten Unterlappen (RUL) metastasieren bevorzugt in subkarinale Lymphknoten (EBUS 7) und in das obere Mediastinum (EBUS 2 und 4). Tumoren des linken Oberlappens (LOL) zeigen v.a. Metastasen im aortopulmonalen Fenster (EBUS 5) und paraaortal (EBUS 6) bzw. im vorderen Mediastinum. Metastasen aus dem linken Unterlappen (LUL) treten bevorzugt in subkarinalen Lymphknoten (EBUS 7) auf, gefolgt von EBUS 5 und 6. Abbildung basierend auf (9).

Bei isoliertem, kleinem Primärtumor im UICC-Stadium I beträgt die 5-Jahres-Überlebensrate 80-90%. Bei Auftreten einer lymphatischen Metastasierung (ab UICC-Stadium IIB) sinkt die Überlebensrate auf 50%, im Stadium IV auf $\leq 10\%$ (5). Deshalb ist bei Erstdiagnose eines NSCLC die Ausbreitungsdiagnostik zur Definition des klinischen Tumorstadiums aktuell von entscheidender Bedeutung für die Prognoseabschätzung und Therapieempfehlung. Bei Vorliegen einer Fernmetastasierung ist die Therapie aktuell in der Regel nur in palliativer Intention möglich.

1.1.3 Patienten mit potenziell kurativer Therapieintention

Patienten mit NSCLC im Stadium I-III kommen prinzipiell für eine Therapie in kurativer Intention infrage, vorausgesetzt, dass eine definitive Therapie aller Manifestationen von Primärtumor und eventuellen Lymphknotenmetastasen möglich ist und der klinische Allgemeinzustand und die kardiopulmonale Fitness diese Therapie zulassen (10). Bei Patienten im UICC-Stadium IA kommt regelhaft die alleinige Operation als anatomische Segmentresektion (T1a) bzw. Lobektomie (T1b/c) zum Einsatz. Ab dem Stadium II wird eine systematische hiläre und mediastinale Lymphknotendissektion empfohlen. Die operative Resektion wird ab dem Stadium IB (II) mit einer adjuvanten Chemotherapie kombiniert, welche eine platinhaltige Zweierkombination über 3-4 Zyklen beinhalten sollte (in Deutschland am häufigsten Cisplatin + Vinorelbin) (10). Die adjuvante Chemotherapie mit Cisplatin + Vinorelbin verbessert bei Patienten im Stadium IB und II (11) bzw. Stadium IB-IIIA (12) die Überlebensrate um ca. 11% nach 5 Jahren verglichen mit der Beobachtungsgruppe. In der JBR.10-Studie (11) und einer späteren gepoolten Analyse mehrerer randomisierter kontrollierter Studien (13) war dieser Überlebensvorteil jedoch in der Subgruppe der Patienten im Stadium IB nicht signifikant. In der ANITA-Studie wurde ebenfalls kein Überlebensvorteil für Patienten mit N0 (d.h. Stadium IIA) beobachtet (12). Darauf basierend empfiehlt die S3-Leitlinie (2018) eine adjuvante Chemotherapie im Stadium IB nur als „individuelle Therapieentscheidung unter Berücksichtigung der Komorbidität, des Alters und der kardiopulmonalen Funktion“, im Stadium II und IIIA hingegen grundsätzlich bei „gutem Allgemeinzustand“ (d.h. Eastern Cooperative Oncology Group [ECOG] Status 0 oder 1) (14). Im Stadium IIIA mit Single-Level-N2 wird die Chemotherapie häufig durch eine sequenzielle Radiotherapie ergänzt, bei Multi-Level-N2 oder T4-Tumoren werden hingegen individualisierte trimodale Therapiekonzepte eingesetzt. Im inoperablen Stadium IIIA und im Stadium IIIB wird i.d.R. eine primäre, simultane Radiochemotherapie empfohlen (10, 15).

Da der Überlebensvorteil durch die adjuvante Chemotherapie im Stadium IB-IIIA nicht einheitlich ist, wären prädiktive Faktoren wünschenswert, die helfen können, ein Ansprechen oder Nichtansprechen auf die adjuvante Chemotherapie vorherzusagen. Sie könnten damit – über das klinische Tumorstadium hinausgehend – eine Entscheidungshilfe für/wider eine Chemotherapie bieten. Solche Faktoren sind jedoch in der klinischen Versorgung nicht etabliert. In einer Post-hoc-Analyse von 253 Patienten im Stadium IB und II aus der JBR.10-Studie konnte ein prädiktiver Wert einer p53-Überexpression gezeigt werden. Nur Patienten mit p53-Überexpression im Tumorgewebe zeigten einen signifikanten Überlebensvorteil durch die adjuvante Chemotherapie verglichen mit alleiniger Operation (16). Eine gepoolte Analyse mit 1209 Patienten mehrerer randomisierter kontrollierter Studien, darunter auch JBR.10, konnte zudem einen prädiktiven Wert des p53-Mutationsstatus

(Wildtyp vs. mutiert) bestätigen (17). Eine Assoziation zwischen einer p53-Überexpression bzw. p53-Mutation mit einer Resistenz gegenüber Cisplatin-basierter Chemotherapie beim NSCLC konnte am Zellmodell und in klinischen Studien gezeigt werden (18-20). Einzug in klinische Entscheidungen haben diese Erkenntnisse bislang jedoch nicht gehalten.

Als prognostisch relevante Parameter, jedoch vorrangig ohne tatsächliche prädiktive Bedeutung, bei Patienten mit NSCLC im Stadium I-IIIa wurden v.a. der ECOG-Status (13), das Geschlecht und Alter (21) identifiziert, wobei letztere beiden Faktoren keine konsistente prognostische Bedeutung zeigen (13). Bezüglich einer unabhängigen prognostischen Bedeutung des histologischen Subtyps sind die Berichte in der Literatur sehr variabel (13, 21). Eine ältere Meta-Analyse von Brundage *et al.* aus 2002 konnte keinen unabhängigen prognostischen Stellenwert des Subtyps nachweisen (22). Eine aktuellere Analyse basierend auf der unter Punkt 1.1.2.1 dargestellten Klassifikation aus 2011 (3) konnte innerhalb der Gruppe der Adenokarzinome die günstigste Prognose für die minimalinvasive und die lepidisch prädominante Variante zeigen, hingegen die schlechteste Prognose für die mikropapillär prädominante und die solide prädominante Variante mit Schleimbildung (23). Tsao *et al.* zeigten in der o.g. gepoolten Analyse randomisierter kontrollierter Studien zur adjuvanten Chemotherapie im Stadium I-III ein signifikant kürzeres krankheitsfreies Überleben (disease-free survival [DFS]) des mikropapillären bzw. solide prädominanten Adenokarzinomsubtyps, jedoch keinen Unterschied hinsichtlich des Gesamtüberlebens (OS). Ein Überlebensvorteil der adjuvanten Chemotherapie gegenüber alleiniger Operation im Sinne eines prädiktiven Effektes war nur in der Subgruppe der mikropapillär bzw. solide prädominanten Adenokarzinome nachweisbar und nur für das DFS (24). Somit bleibt auch diese aktuelle Subklassifikation in ihrer Relevanz für klinische Therapieentscheidungen bislang undefiniert und ist aktuell kein Bestandteil von Leitlinienempfehlungen für/wider eine adjuvante Chemotherapie (10).

Bedeutung für die eigenen Arbeiten:

Bei Patienten mit NSCLC und potenziell kurativer Therapieintention können interindividuelle Unterschiede im Ansprechen auf die Chemotherapie und im Überleben aktuell nur unzureichend durch etablierte klinische und histologische Faktoren erklärt werden. Bildgebende Parameter, die über die Definition des klinischen Tumorstadiums hinausgehen, sind dringend erforderlich, um das Outcome der Patienten durch individualisiertere Therapieentscheidungen zu verbessern.

1.2 FDG-PET/CT Diagnostik

1.2.1 Prinzip

Die Positronen-Emissions-Tomographie / Computertomographie (PET/CT) verbindet die morphologische Bildgebung der CT-Diagnostik mit der Darstellung von Stoffwechselprozessen oder Rezeptor-Liganden-Interaktionen mittels PET. Die verwendeten PET-Radiopharmaka (Tracer) emittieren Positronen, welche mit Elektronen im Zielgewebe im Sinne einer Annihilation interagieren und dabei 511 keV-Photonen im nahezu 180°-Winkel zueinander emittieren. Die Linie, auf der beide Photonen sich voneinander entfernen und zum Detektor gelangen, wird als Koinzidenzlinie

bezeichnet. Die Photonen werden durch ringförmig konfigurierte Detektorelemente registriert und diese Signale durch nachfolgende Bilddatenrekonstruktion in ein dreidimensionales Abbild der Tracerverteilung im untersuchten Körperabschnitt umgewandelt. Die CT-Bilddaten werden dabei zur Schwächungskorrektur genutzt und dienen zudem der anatomischen Befundkorrelation (25).

Der weitaus am häufigsten klinisch eingesetzte Tracer, auch für die Bildgebung beim NSCLC, ist die [¹⁸F]Fluordesoxyglucose (FDG). FDG wird durch Glukosetransporter (GLUT) in die Zelle aufgenommen und bleibt dort nach Phosphorylierung gefangen („metabolic trapping“), d.h. im Zeitverlauf nach der Tracerinjektion wird typischerweise eine stetige Zunahme der zellulären Traceranreicherung beobachtet (25). Aufgrund einer vermehrten Expression von GLUT1 und GLUT3 in neoplastischen Zellen weisen diese typischerweise eine gesteigerte FDG-Aufnahme (Uptake) auf. Eine diagnostisch nutzbare tumorale Traceranreicherung ist über eine Vielzahl von Tumorentitäten nachweisbar, *per se* somit aber entitätsunspezifisch (26-29). Zudem kann der erhöhte Glukosestoffwechsel bestimmter inflammatorischer Zellen die Differenzierung zwischen malignen und benignen (inflammatorischen) Prozessen beeinträchtigen (30, 31). Dies ist insbesondere für die Lymphknotendiagnostik relevant (mögliche falsch positive Befunde). Eine physiologisch gesteigerte FDG-Aufnahme wird zudem in verschiedenen Organsystemen, v.a. Gehirn, Leber, Herz und Nieren, beobachtet (32) und kann dort die Detektion maligner Läsionen erschweren (33, 34).

Zudem variiert die GLUT-Expression und damit die Höhe des FDG-Uptakes (35) deutlich zwischen den histologischen Subtypen des NSCLC. Der im Durchschnitt höchste FDG-Uptake wird beim Plattenepithelkarzinom (SCC) und großzelligen Karzinom beobachtet (36). Innerhalb der Gruppe der Adenokarzinome (ADC) ist die Intensität der Traceranreicherung hingegen sehr variabel und v.a. vom Wachstumsmuster abhängig. Lepidisch prädominante ADC weisen im Durchschnitt den geringsten FDG-Uptake auf, gefolgt von papillär und azinär prädominanten Formen. Der Uptake solide prädominanter Subtypen ist am höchsten und steigt mit dem Anteil der soliden Komponente (35). Beim NSCLC im Allgemeinen, und dem ADC im Speziellen, korreliert der FDG-Uptake außerdem positiv mit der Wachstumsrate und mit dem Differenzierungsgrad G1 bis G3 (36, 37). Zudem ist der Traceruptake bei muzinösen gegenüber nicht-muzinösen Subtypen verringert (38). Die geringste Traceranreicherung kann somit bei gut differenzierten ADC (minimalinvasiv, ADC in situ, lepidisch prädominant) sowie muzinösen invasiven ADC beobachtet werden, der höchste Uptake bei solide prädominanten ADC, bei SCC und großzelligen Lungenkarzinomen (36, 38). Dies hat Implikationen für die klinische Diagnostik, da ein geringer FDG-Uptake die Abgrenzung gegenüber benignen, unspezifischen FDG-Anreicherungen erschweren und die Sensitivität in der Primärtumor- und Ausbreitungsdiagnostik teils deutlich einschränken kann (39, 40).

1.2.2 Bildqualität: Räumliche Auflösung und Signal-Rausch-Abstand (SNR)

Eine optimale Bildqualität in der PET-Diagnostik kann vereinfacht als Kombination aus einer möglichst akkuraten Abbildung fokaler Aktivitätsanreicherungen (sog. Recovery) bei günstigem Rauschverhalten, d.h. einem hohen Signal-Rausch-Abstand (signal-to-noise ratio; SNR), betrachtet werden. Die Recovery wird dabei maßgeblich durch die endliche räumliche Auflösung der PET-Bildgebung limitiert, die SNR durch Unsicherheiten in der Unterscheidung wahrer Zählereignisse von gestreuten Ereignissen (Scatter) und zufälligen Koinzidenzen (Randoms) (41).

Bei der Charakterisierung der räumlichen Auflösung der PET-Bildgebung muss zwischen der intrinsischen räumlichen Auflösung, d.h. der Auflösung, die allein durch physikalische Eigenschaften der Positronen und Photonen sowie deren Interaktion mit dem Detektor limitiert wird, und der effektiven Auflösung, die zusätzlich durch Eigenschaften des abgebildeten Objektes und der Bildverarbeitung begrenzt wird, unterschieden werden. Effekte, die die intrinsische Auflösung des PET-Systems limitieren, sind die physikalische Reichweite von Positronen im Gewebe bis zum eigentlichen Annihilationsereignis (durchschnittlich 0,6 mm in Wasser (42)) und mögliche Abweichungen vom idealen 180° -Winkel der Annihilationsphotonen ($\pm 0,25^\circ$) (43). Zudem sind die physikalischen Eigenschaften des Detektormaterials in der Interaktion mit den Annihilationsphotonen, das Detektordesign (Größe der Elemente und Verschaltung einzelner Detektorelemente zu Detektorblöcken) (43) und Fehlerquellen bei der Signaldetektion und -verarbeitung im System von Bedeutung (z.B. Detektorrauschen und Cross-talk zwischen Detektorelementen (44)).

Die effektive, tatsächlich für die Abbildungsqualität maßgebliche räumliche Auflösung, wird insbesondere von der Bilddatenrekonstruktion beeinflusst. Heutzutage ist die iterative Bildrekonstruktion klinischer Standard. Eine Verbesserung der effektiven räumlichen Auflösung wird durch Integration der systemspezifischen Punktabbildungsfunktion (point spread function; PSF) in die Rekonstruktion erreicht. Diese ermöglicht es, Verzerrungen in der Abbildung von Objekten zu kompensieren, die abseits des Zentrums des Gesichtsfeldes lokalisiert sind und deshalb nicht optimal von den ringförmig angeordneten Detektorelementen abgebildet werden können (45). Diese Kompensation erlaubt eine Verbesserung der effektiven räumlichen Auflösung um bis zu >1 mm Halbwertsbreite (full width at half maximum [FWHM]) (46) und kann die Detektion v.a. von kleinen Läsionen verbessern (47). Die PSF-Kompensation kann jedoch selbst zu Bildartefakten beitragen, die durch Überschwinger/Randanhebungen an der Oberfläche von Läsionen visualisiert werden können (sog. Gibbs-Artefakte) (46). Aufgrund dieser Vielzahl von Einflussgrößen und des stetigen technischen Fortschritts in der Detektortechnologie und Bildrekonstruktion variiert die effektive räumliche Auflösung von aktuell im klinischen Einsatz befindlichen PET-Systemen deutlich, etwa zwischen 4 und 8 mm FWHM ((46), unpublizierte eigene Ergebnisse).

Das Bildrauschen bzw. die SNR wird maßgeblich durch Scatter und Randoms bestimmt. Scatter bezeichnet aufgrund von Streuung im Gewebe von der Koinzidenzlinie abgelenkte und somit nicht bzw. im „falschen“ Detektorelement detektierte Photonen. Randoms sind Zählereignisse, die durch Photonen entstehen, die aufgrund von Streuung nur zufällig gleichzeitig auf gegenüberliegenden Detektorelementen registriert werden. Somit ist die SNR v.a. von der Objekt- bzw. Patientengeometrie (mehr Streuung bei größerem Patientenumfang) und Abbildungseigenschaften des PET-Systems abhängig. Letztere umfasst die zeitliche Auflösung des Detektormaterials und der Detektorelektronik und die Genauigkeit der Scatter- und Schwächungskorrektur im Rahmen der Bildrekonstruktion (41). Da Detektorelemente in modernen PET-Systemen eine hohe zeitliche Auflösung erreichen, kann durch Analyse des Zeitabstandes zwischen den Detektionen beider Annihilationsphotonen (time of flight; TOF) näherungsweise auf die Lokalisation des Annihilationsereignisses entlang der Koinzidenzlinie rückgeschlossen werden. Dadurch können Scatter und Randoms besser von wahren

Signalen differenziert werden und durch Integration dieser Information in die Bildrekonstruktion die SNR verbessert werden (48, 49).

Bedeutung für die eigenen Arbeiten:

Die rekonstruierte räumliche Auflösung, Recovery und SNR sind zentrale Parameter der Bildqualität in der PET. Damit sind sie wichtige Kenngrößen für die Charakterisierung der Bildrekonstruktion und der finalen Bildeigenschaften im Hinblick auf visuelle und quantitative Auswertungen.

1.2.3 Visuelle Bildbefundung

Die Befundung der FDG-PET/CT Diagnostik in der klinischen Routine erfolgt primär visuell, d.h. durch Beurteilung pathologischer Tracermehr- oder -minderanreicherungen. Wie unter 1.2.1 dargestellt, muss dabei beachtet werden, dass eine intensive FDG-Anreicherung von Primärtumor und Metastasen in Abhängigkeit von der Tumorphistologie und der Läsionslokalisation nicht unter allen Umständen gegeben ist und somit die Sensitivität eingeschränkt wird.

Zudem ist aufgrund der limitierten räumlichen Auflösung (siehe 1.2.2) die fokale Abgrenzbarkeit einer wahren Traceranreicherung, und damit die diagnostische Sensitivität der FDG-PET/CT, maßgeblich von der Läsionsgröße abhängig. Die limitierte räumliche Auflösung der PET-Bildgebung führt mit abnehmender Läsionsgröße zu einer zunehmenden Unterschätzung der Aktivitätsanreicherung (sog. Partialvolumeneffekt). Auch mit modernen PET-Systemen beträgt deshalb die Recovery z.B. in Läsionen mit 10 mm Durchmesser nur ca. 33%, d.h. nur ein Drittel der wahren Aktivität wird reproduziert (50). Demgegenüber kann auch eine ungünstige SNR die visuelle Befundung beeinträchtigen, da Bildrauschen fälschlicherweise als wahre Traceranreicherung fehlinterpretiert werden kann. Dies kann zu falsch positiven Befunden führen.

1.2.4 Quantifizierung und quantitative Bildparameter

1.2.4.1 Prinzip der Quantifizierung

Die PET-Bildgebung erlaubt die genaue Quantifizierung der Aktivitätskonzentration im Gewebe, z.B. in kBq/ml. Die Genauigkeit und Stabilität wird durch regelmäßige Kalibrierung an einem Aktivimeter mittels standardisierter Phantommessungen gewährleistet (51). Um die Höhe und Bedeutung einer Aktivitätskonzentration im Zielgewebe zu bewerten, wird diese typischerweise in ein Verhältnis zur injizierten Aktivität und durchschnittlichen Aktivitätskonzentration im Körper gesetzt. Dieses Verhältnis wird als standardized uptake value (SUV) bezeichnet und ist der gängigste semiquantitative Parameter in der klinischen PET-Diagnostik. Der SUV ist einheitslos und wird i.d.R. auf das Körpergewicht in kg bezogen (52):

$$\text{SUV} = \frac{\text{Aktivitätskonzentration (kBq/ml)}}{\text{applizierte Aktivität (MBq) / Körpergewicht (kg)}} \quad (1)$$

Je nachdem, welche Voxel in der Zielläsion betrachtet werden, wird entweder der maximale SUV (SUVmax) oder ein durchschnittlicher SUV mehrerer Voxel (SUVmean) bestimmt. Ein Spezialfall des SUVmean ist der SUVpeak, der als höchster durchschnittlicher SUV definiert wird, der in einem

Volumen von ca. 1 ml (= Kugel mit 1,2 cm Durchmesser) innerhalb der Läsion gemessen werden kann. Diese Standardisierung der Definition zielt auf eine höhere Reproduzierbarkeit des SUVpeak verglichen mit dem SUVmean ab (50). Parameter, die, wie der SUV, die Voxelwerte selbst (Grauwert bzw. Intensität der FDG-Anreicherung) oder deren Histogramm abbilden, werden gemäß Nomenklatur von Sollini *et al.* zu Parametern 1. Ordnung („first-order“) gezählt (53).

1.2.4.2 Das metabolische Tumervolumen (MTV)

Zudem kann das Volumen der abgegrenzten Voxel in ml berechnet werden, um das Volumen der traceranreichernden bzw. metabolisch aktiven Läsionsanteile zu bestimmen (sog. metabolic tumor volume; MTV). Dabei existiert keine einheitliche oder optimale Definition, welche Voxel in das MTV der Läsion eingeschlossen werden und welche Voxel zum Hintergrund gerechnet werden, um die Läsionsgrenzen akkurat zu identifizieren. Ein simpler und weitverbreiteter Ansatz ist die schwellwertbasierte Abgrenzung aller Voxel innerhalb einer Zielregion, die einen vordefinierten SUV-Wert überschreiten, z.B. alle Voxel mit mindestens einem absoluten SUV von 2,5 (54, 55) oder mit mindestens 41% des SUVmax (55, 56). Je starrer und simpler diese Schwellwertdefinition ist, desto reproduzierbarer wird die MTV-Definition, aber umso größer wird auch der potenzielle Fehler in der Abgrenzung der Läsion, wenn das Läsion-zu-Hintergrund-Verhältnis (TBR) gering ist oder die Hintergrundaktivität sehr heterogen (46, 57). Um diese Probleme zu minimieren, kann ein relativer Schwellwert eine Hintergrundanpassung beinhalten, um auch bei geringer TBR eine sinnvolle Läsionsabgrenzung zu ermöglichen. Dies erfordert jedoch zusätzlich zur Schwellwertdefinition die Definition der Hintergrundaktivität und deren Standardisierung (58). Iterative (59) und stochastische Verfahren (57) sowie Machine-Learning-Ansätze (60, 61) mit dem Ziel einer automatisierten Ermittlung der Läsionsgrenzen wurden vorgeschlagen. Neben der Voxelintensität als Entscheidungsgrundlage kann auch die räumliche Verteilung und die Homogenität bzw. Heterogenität benachbarter Voxel für deren Klassifizierung genutzt werden. Beispiele sind Region-Growing-Verfahren (62), graphenbasierte (63) und gradientenbasierte Methoden (64). Alle Verfahren bleiben jedoch in unterschiedlicher Weise limitiert durch die Einschränkungen in der PET-Bildgebung (räumliche Auflösung und Rauschverhalten) und in der Reproduzierbarkeit der Abgrenzungsmethodik (65).

1.2.4.3 Texturparameter und Bildparameter höherer Ordnung

Gemäß Sollini *et al.* zählt das MTV zu den Form- und Größenparametern („shape and size“) (53), ebenso wie Maße der Form und räumlichen Irregularität des MTV (z.B. asphericity, solidity oder eccentricity) (66, 67). Aus der Verteilung der Anreicherungsintensitäten innerhalb des MTV können zudem Texturparameter abgeleitet werden, die die räumliche Verteilung der Grauwerte/ Mehranreicherungen in der Grauwertematrix (grey-level co-occurrence matrix; GLCM) beschreiben („second-order“ und „high-order“). Sie sind somit Maße der voxelweisen Heterogenität (z.B. energy, contrast, local homogeneity oder entropy) (67).

Formparameter und second-order-Texturparameter wurden erstmalig 2009 von El Naqa *et al.* systematisch im Kontext der PET-Bildgebung beschrieben. In einer kleinen Fallserie von 9 Patienten mit Kopf-Hals-Karzinom zeigte u.a. die solidity eine höhere area under the curve (AUC) als der

SUVmax und SUVmean in der Vorhersage der Fälle mit ungünstigem Verlauf (67). Als weiteren Formparameter entwickelte unsere Arbeitsgruppe die asphericity (ASP), die die prozentuale Abweichung der Oberfläche des MTV von der Oberfläche einer isovolumetrischen Kugel abbildet (66):

$$ASP (\%) = \sqrt[3]{H} - 1 \quad \text{mit} \quad H = \frac{1}{36\pi} \frac{S^3}{V^2} \quad (2)$$

Bei 36 Patienten mit Erstdiagnose eines Kopf-Hals-Karzinoms erlaubte die ASP in der multivariablen Cox-Regression eine unabhängige Prädiktion des progressionsfreien Überlebens (PFS) und OS, hingegen war das MTV lediglich für das PFS unabhängig prädiktiv, der SUVmax für keinen der Endpunkte (66). Der unabhängige Stellenwert der ASP wurde an einem separaten Kollektiv von 37 Patienten mit Kopf-Hals-Karzinom unter Verwendung eines anderen PET-Systems bestätigt (68).

Alle o.g. Parameter folgen, auch wenn sie automatisiert aus Bilddaten extrahiert werden, einer vorherigen menschlichen Definition, mathematischen Formel bzw. statistischen Größe. Dem gegenüber stehen Parameter, die aus Prozessen des maschinellen Lernens und des Deep Learnings entstehen, somit nicht vordefiniert sind und ggf. nur bedingt erklärbar und klassifizierbar sind (69, 70).

1.2.4.4 Limitationen der Quantifizierung

a) Reproduzierbarkeit homogener Aktivitätskonzentrationen

Die Aktivitätskonzentration im Gewebe kann bei homogener Anreicherung in einem ausreichend großen Zielvolumen exakt bestimmt und reproduziert werden; dies ist Grundlage der SUV-Kalibrierung von PET-Systemen mithilfe homogener befüllter Phantome. Damit ist die Grundvoraussetzung für die Vergleichbarkeit von Aktivitätskonzentrationen bzw. SUV-Werten zwischen Untersuchungen desselben Patienten und zwischen verschiedenen Patienten gegeben. Nur dadurch kommen quantitative Parameter grundsätzlich für differenzialdiagnostische, prognostische oder prädiktive Fragestellungen infrage, insbesondere, wenn in multizentrischen Studien die Verwendung verschiedener PET-Systeme notwendig wird.

Die exakte Abbildung einer durchschnittlichen Aktivitätskonzentration in einem homogenen Volumen wird – im Gegensatz zu einer fokalen Anreicherung – kaum durch klinisch übliche Variationen in den unter 1.2.2 genannten Faktoren beeinflusst (räumliche Auflösung, Bildrekonstruktion (71) und Bildrauschen (72)). Bei Patientenuntersuchungen kann dies mithilfe von Wiederholungsuntersuchungen durch die hohe Reproduzierbarkeit von Aktivitätskonzentrationen und SUVmean-Werten in physiologischem Lebergewebe nachgewiesen werden. Die Variabilität des Leber-SUVmean innerhalb desselben Patienten beträgt dann in der Regel unter 5% (-10%) (71, 73-76). In einer Studie von Boktor *et al.* betrug die Variabilität des Blut-SUVmean in der thorakalen Aorta bei 132 Patienten zwischen zwei PET/CT-Untersuchungen innerhalb von 12 Monaten sogar nur ca. 2%. Das standardisierte Untersuchungsprotokoll gewährleistete, dass die mittlere injizierte Aktivität pro kg Körpergewicht und die Uptakezeit zwischen Injektion und Bildgebungsstart zwischen beiden Zeitpunkten vergleichbar waren (75). Eine solche Standardisierung der Patientenvorbereitung (v.a. Nüchternheit und Blutzucker), injizierten Aktivität (möglichst standardisierte, gewichtsabhängige Aktivität) und Untersuchungsprotokolle (v.a. standardisierte Uptakezeit) ist für reproduzierbare SUV-

Werte wichtig (51). Anderenfalls kann die SUV-Wert-Variabilität auch bei demselben Patienten unter Verwendung desselben PET-Systems und der identischen Bildrekonstruktion zunehmen (77, 78).

b) Reproduzierbarkeit fokaler Aktivitätskonzentrationen

Deutlich komplexer ist die Reproduktion von fokalen Aktivitätskonzentrationen und Läsions-SUV. Bereits unter vergleichbaren Untersuchungs- und Bildrekonstruktionsbedingungen kann der Läsions-SUV bei Wiederholungsmessungen auch innerhalb weniger Tage deutlich variieren. Eine Meta-Analyse von de Langen *et al.* zu 86 Patienten mit seriellen PET-Untersuchungen zeigte, dass der SUVmax der Tumorkläsionen eine Variabilität von bis zu 30% aufwies, der SUVmean von bis zu 20% (79). Der Läsions-SUV unterliegt zudem Variation durch die Bildrekonstruktion. Der Einfluss der TOF-Rekonstruktion auf den SUV wird als vergleichsweise gering beschrieben, die PSF-Kompensation kann jedoch durch o.g. Gibbs-Artefakte zu SUV-Erhöhungen von >30% führen (80, 81). Eigene Untersuchungen bei Patienten mit Lebermetastasen eines kolorektalen Karzinoms haben zudem gezeigt, dass die TOF-Rekonstruktion bei geringer TBR ebenfalls zu SUV-Erhöhungen führt. Die Differenzen waren bei kleinen Läsionen besonders deutlich (82).

Houdu *et al.* konnten den Einfluss einer standardisierten Bildrekonstruktion auf die prognostische Wertigkeit des SUVmax bei Patienten mit NSCLC in einer monozentrischen Auswertung zeigen, in der die PET-Daten in 4 Subgruppen der Patienten jeweils mit PSF-Rekonstruktion, aber unterschiedlicher Halbwertsbreite des Post-Processing-Filters rekonstruiert wurden (83). Diese unterschiedlichen Filter (3-10 mm FWHM) wurden mit einer harmonisierten PSF-Rekonstruktion (6,3 mm FWHM) gemäß EANM Research Ltd. (EARL)-Akkreditierung (84, 85) verglichen. Das EARL-Programm hat zum Ziel, durch standardisierte Phantommessungen mit einem Zylinderphantom mit Kugeleinsätzen für jeden akkreditierten PET-Scanner eine spezifische Rekonstruktion zu identifizieren, die eine Recovery, und damit einen SUVmax, innerhalb eines definierten Toleranzbereichs erreicht. Damit wird eine bessere Vergleichbarkeit unterschiedlicher PET-Systeme und Rekonstruktionsalgorithmen v.a. bei multizentrischen Studien angestrebt. Houdu *et al.* zeigten, dass der unabhängige prognostische Wert des SUVmax der Patienten nur erhalten blieb, wenn die Daten aller 139 Patienten identisch (EARL-konform) rekonstruiert wurden (83).

c) Reproduzierbarkeit des MTV: Bildrekonstruktion

Lasnon *et al.* untersuchten die Vergleichbarkeit von MTV bei 18 Patienten mit NSCLC im Stadium IIIA-III B. Sie verglichen eine EARL-konforme, herkömmliche iterative Rekonstruktion (ordered subset expectation maximization; OSEM) mit einer ebenfalls EARL-konformen PSF-Rekonstruktion (Filter mit 7 mm FWHM) und einer nicht EARL-konformen PSF-Rekonstruktion (kein Filter). Die Übereinstimmung der MTV-Volumina zwischen den beiden EARL-konformen Rekonstruktionen war signifikant besser als zwischen EARL-konformer OSEM-Rekonstruktion und nicht konformer PSF-Rekonstruktion, sowohl für die 40%-Schwellwertmethode als auch die gradientenbasierte Abgrenzungsmethode (86).

Zhuang *et al.* berichteten über die Retest-Reliabilität des MTV und von Texturparametern 2. Ordnung (u.a. entropy) bei 10 Patienten mit NSCLC und PET-Akquisition zu zwei verschiedenen Zeitpunkten nach Tracerinjektion (60 und 90 min). Die Autoren untersuchten die Retest-Reliabilität unter dem Einfluss des Zeitpunktes nach Injektion, der MTV-Abgrenzungsmethode und von 4

verschiedenen Bildrekonstruktionen (u.a. eine EARL-konforme PSF-Rekonstruktion). Das MTV und die Texturparameter waren signifikant von der Bildrekonstruktion abhängig. Sofern für alle 10 Patienten die identische Bildrekonstruktion genutzt wurde, war die Retest-Reliabilität aller Parameter jedoch weitgehend unabhängig vom Bildrekonstruktionsverfahren. Die Reliabilität von MTV und Texturparametern blieb aber systematisch von der Segmentierungsmethode, dem Läsions-SUVmax und dem Untersuchungszeitpunkt abhängig. Die Autoren schlussfolgerten, dass eine einheitliche Bildrekonstruktion die Grundlage einer Vergleichbarkeit von PET-Bildparametern bleibt, jedoch selbst nach dieser Harmonisierung spezifische Szenarien (u.a. Bildgebungszeitpunkt nach Injektion und Wahl des MTV-Segmentationsverfahrens) jeweils eine Anpassung, Optimierung und Validierung von Bildparametern erfordern (87).

d) Reproduzierbarkeit des MTV: Segmentationsverfahren

Die Vielzahl möglicher MTV-Segmentierungsverfahren (siehe 1.2.4.1) bewirkt, dass auch bei exakt reproduzierbarer Bilderstellung (Akquisition und Rekonstruktion) die Ausprägung des MTV einer deutlichen Variation unterliegen kann, wenn die Segmentierung nicht vereinheitlicht ist. Gemäß Nestle *et al.* kann diese MTV-Variabilität in Extremfällen ein Mehrfaches des Läsionsvolumens betragen. Bei 25 Patienten mit fortgeschrittenem NSCLC vor Strahlentherapie war das vom Experten definierte MTV durchschnittlich 158 ml groß, verschiedene schwellwertbasierte Abgrenzungsmethoden ergaben durchschnittliche MTV von 54 ml (40%-Methode) bis 165 ml (SUV 2,5-Methode) (54). Diese Variation im MTV wird sich in unterschiedlichem Maß auch in jeglichen daraus abgeleiteten Parametern fortsetzen (SUVmean, Textur- und Formparameter) (88). Bilden diese Parameter, wie Texturparameter, vorrangig die Heterogenität der Aktivitätsverteilung innerhalb des MTV ab und nur nachrangig die Form oder Größe des MTV, kann jedoch der Einfluss der ursprünglichen Bilderstellung auf die Variabilität höher sein als die Bedeutung der MTV-Abgrenzung (87). Umgekehrtes ist für Formparameter (z.B. ASP, solidity) anzunehmen.

e) Reproduzierbarkeit des MTV: Observer-Variabilität

Eine abschließende Ursache für Variabilität in MTV, SUV und weiteren Bildparametern ist die Intra- und Interobserver-Variabilität, d.h. der Einfluss menschlicher Erfahrung und Vorannahmen sowie manueller Interaktionen und Korrekturen während der MTV-Abgrenzung. Die rein manuelle Segmentierung kann bei erfahrenen Befundern eine gute Übereinstimmung mit Größenmessungen in der morphologischen Bildgebung oder in histopathologischen Präparaten ergeben (89), weist jedoch die ungünstigste Intra- und Interobserver-Variabilität auf (90, 91) und kann deshalb für Studienanwendungen ggf. unbrauchbar sein. Die Variabilität sinkt deutlich, wenn halbautomatische Verfahren zum Einsatz kommen und die Notwendigkeit manueller Nachkorrekturen gering ist (90, 91). Dies können z.B. schwellwertbasierte Verfahren mit zuvor festgelegtem Schwellwert oder nicht schwellwertbasierte Algorithmen mit fest definierten Parametern sein (88, 90, 91). Jedoch sind Korrektheit und Ausmaß erforderlicher Nachkorrekturen deutlich von der intraläsionalen Heterogenität und der Variabilität der TBR in der Stichprobe abhängig, insbesondere wenn keine Hintergrundanpassung erfolgt (54, 89, 92). Um die Generalisierbarkeit von MTV-Abgrenzungsalgorithmen zu erhöhen, wurden majority vote-Algorithmen vorgeschlagen, die nur

solche Voxel zum MTV rechnen, die durch eine Mehrheit aus definierten Abgrenzungsmethoden als MTV zugehörig klassifiziert wurden (93). Validiert sind diese Verfahren bislang nicht.

Bedeutung für die eigenen Arbeiten:

Die klinische Evaluation des Stellenwertes von PET-Bildparametern sollte immer auch von einer methodischen Validierung begleitet werden, die insbesondere die Reproduzierbarkeit der Parameter unter dem Einfluss der Bilddatenrekonstruktion und MTV-Abgrenzungsmethodik abbilden sollte.

1.3 FDG-PET/CT Diagnostik beim NSCLC

1.3.1 Prätherapeutische Lymphknotendiagnostik

Aufgrund der prognostischen Bedeutung einer Lymphknotenmetastasierung und ihres Stellenwertes für therapeutische Entscheidungen bei Patienten mit NSCLC ist eine korrekte prätherapeutische Beurteilung der Lymphknoten von entscheidender Wichtigkeit für eine adäquate Therapieempfehlung. Diese Diagnostik erfolgt gemäß Leitlinie (10, 14) zunächst nicht-invasiv mittels kontrastmittelgestützter CT des Thorax und FDG-PET/CT Bildgebung. Falls diese kombinierte Bildgebung den Verdacht auf mediastinale Lymphknotenmetastasen ergibt, sollte dieser Verdacht mittels invasiver Verfahren abgeklärt werden, heutzutage in erster Linie mit der EBUS-gestützten transbronchialen Nadelaspiration/-biopsie (EBUS-TBNA). Sollte diese Diagnostik einen negativen Befund ergeben, wird aufgrund des hohen positiven prädiktiven Wertes der FDG-PET/CT Diagnostik von ca. 80% (14) zusätzlich eine Mediastinoskopie oder videoassistierte Thorakoskopie (VATS) empfohlen. Sollte die initiale Bildgebung alleinig hiläre Lymphknotenmetastasen aufzeigen, wird eine Abklärung auch der mediastinalen Lymphknoten mit der EBUS-TBNA empfohlen (nachrangig Mediastinoskopie bzw. VATS) (10). Ergibt die Bildgebung hingegen keinen Verdacht auf Lymphknotenmetastasen, kann in der Regel ohne zusätzliche invasive Diagnostik die Indikation zur operativen Resektion gestellt werden (10, 14).

Aus diesem zentralen Stellenwert der FDG-PET/CT Diagnostik folgt, dass eine optimale Sensitivität die Rate unnötiger operativer Resektionen (aufgrund unerkannter mediastinaler Lymphknotenmetastasen) minimieren kann (94), und eine optimale Spezifität die Anzahl unnötiger invasiver diagnostischer Verfahren reduziert (95). Entsprechend den etablierten Therapiealgorithmen (siehe Punkt 1.1.3) ist neben der korrekten mediastinalen Lymphknotendiagnostik bereits eine präoperative Befunddifferenzierung zwischen Single-Level-N2 und Multi-Level-N2 wichtig, da bei Single-Level-N2 eher zu einer primären Resektion geraten wird (15).

Die Sensitivität und Spezifität der visuellen Befundung der FDG-PET/CT in der prätherapeutischen Lymphknotendiagnostik bei Patienten mit NSCLC ist von der Definition der Befundungskriterien abhängig. Gemäß der größten Meta-Analyse zu dieser Fragestellung von Schmidt-Hansen *et al.* (2014) wurde ein PET-positiver Befund in den Einzelstudien am häufigsten anhand eines erhöhten FDG-Uptakes gegenüber dem mediastinalen Blutpool-Hintergrund definiert. Dies ergab eine gepoolte Sensitivität von 77% und Spezifität von 90% (40).

a) Ursachen falsch negativer Befunde

Wichtigste Ursache für falsch negative PET-Befunde ist eine geringe Größe der Metastase bzw. des metastatischen Lymphknotens aufgrund der limitierten räumlichen Auflösung und des Partialvolumeneffektes (siehe Kapitel 1.2.2). Insbesondere Metastasen <5 mm Durchmesser werden häufig fälschlicherweise PET-negativ klassifiziert (insbesondere bei einfacher OSEM-Rekonstruktion) (96-98). Vergleichbar mit der Primärtumordiagnostik kann eine geringe FDG-Anreicherung von Adenokarzinomen auch gehäufte falsch negative Lymphknotenbefunde bei diesen Patienten ergeben (99). Außerdem wurde eine verminderte Sensitivität bei zentral wachsenden Tumoren (99) und solchen mit >3 cm Durchmesser berichtet (99, 100); und sogar – entgegen der Intuition – bei intensiver FDG-Anreicherung des Primärtumors (99, 100). Deshalb wird gemäß ESMO-Leitlinie bei zentralen Tumoren, Tumoren >3 cm und insbesondere Adenokarzinomen mit intensivem FDG-Uptake auch bei negativem Lymphknotenbefund in der Bildgebung grundsätzlich eine Abklärung mit EBUS-TBNA empfohlen (10). Nekrotische oder eingeblutete Metastasen sind weitere Ursachen falsch negativer Befunde (98).

b) Ursachen falsch positiver Befunde

Die weitaus häufigste Ursache für falsch positive thorakale Lymphknotenbefunde sind unterschiedliche inflammatorische Veränderungen (98, 101) aufgrund der gesteigerten GLUT-Expression in Entzündungszellen wie Makrophagen (30, 31). Histologische Beispiele sind eine unspezifische follikuläre Hyperplasie, Sinushistiozytose (102), Anthrakose und Anthrakosilikose, granulomatöse Veränderungen oder eine Sarkoidose. Dies umfasst ebenso begleitende entzündliche Geschehen bei Lungengerüstveränderungen wie Bronchiektasen, Atelektasen oder Emphysem (103) sowie bei akuten und chronischen infektiösen Lungenerkrankungen, beispielsweise bei Tuberkulose (104). Erhöhte Raten falsch positiver Befunde wurden außerdem bei Primärtumoren mit guter Differenzierung beschrieben (100).

Es wurde vorgeschlagen, die Rate falsch positiver Befunde zu verringern, indem die Dynamik der Aktivitätsanreicherung nach der Tracerinjektion als zusätzliches Kriterium genutzt wird. In präklinischen Untersuchungen beobachteten Zhuang *et al.* in Tumorzellen eine über die Zeit deutlich zunehmende FDG-Akkumulation zwischen 45 und 90 min nach Injektion, bei mononukleären Zellen (Leukozyten) hingegen eine etwas abnehmende Anreicherung. Sie postulierten, dass die Dynamik der Traceranreicherung von einem frühen zu einem zusätzlichen späteren Akquisitionszeitpunkt (sog. Dual Time Point Imaging [DTPI]) helfen kann, diese Zelltypen zu differenzieren (105). Hu *et al.* berichteten bei 102 Patienten mit NSCLC eine signifikant verbesserte Spezifität, positiven prädiktiven Wert und Genauigkeit in der thorakalen Lymphknotenbefundung mittels DTPI (60 und 120 min p.i.), jedoch nur in der Subgruppe der Patienten mit pulmonalen Komorbiditäten (106). In einer Meta-Analyse von Shen *et al.*, die neben dem NSCLC auch Studien zu anderen Tumorentitäten beinhaltete, konnte auf Patientenebene keine Verbesserung der diagnostischen Performance nachgewiesen werden, in der läsionsbasierten Auswertung jedoch eine Verbesserung der Spezifität von 82 auf 88% bei weitgehend unveränderter Sensitivität (82 vs. 80%). Die Befundung in den Einzelstudien erfolgte überwiegend SUV-basiert (107).

c) Einfluss der Bilddatenrekonstruktion

In den Studien, die der o.g. Meta-Analyse von Schmidt-Hansen *et al.* zugrunde lagen, wurden ausnahmslos nicht TOF-fähige PET-Systeme eingesetzt und die Rekonstruktionen ohne PSF-Kompensation durchgeführt (40). Gemäß Akamatsu *et al.* und Armstrong *et al.* ist der Einfluss der TOF-Rekonstruktion auf den SUVmax von Lymphknoten in der FDG-PET/CT im Durchschnitt gering, mit mittleren SUVmax-Steigerungen von <10% (80, 81). Systematische Untersuchungen zum Einfluss der TOF-Rekonstruktion auf die Genauigkeit der Lymphknotendiagnostik liegen bislang nicht vor, eine relevante Steigerung der Sensitivität ist angesichts des moderaten Einflusses auf den SUVmax jedoch unwahrscheinlich. Die Steigerung der SNR (49) könnte prinzipiell eine verbesserte Spezifität ermöglichen. Hingegen wurden für die PSF-Rekonstruktion SUVmax-Steigerungen in Lymphknoten und anderen kleinen Läsionen von >30% berichtet (80, 81, 108-110). Lasnon *et al.* zeigten, dass mit PSF die Sensitivität in der visuellen Befundung thorakaler Lymphknoten beim NSCLC gegenüber der einfachen OSEM-Rekonstruktion steigt (97% vs. 78%), jedoch die Spezifität abnimmt (58% vs. 71%). In Übereinstimmung mit eigenen Daten (82) war die SUVmax-Zunahme durch PSF in kleinen Läsionen (<10 mm Durchmesser) besonders deutlich (108, 111). Diese Ergebnisse implizieren, dass sowohl bei der visuellen Befundung als auch bei der Anwendung von SUV-Schwellwerten eine Anpassung an diese Besonderheiten der PSF-Rekonstruktion erforderlich ist.

d) Einfluss der Befundungskriterien

Das Fehlen standardisierter bzw. universell anwendbarer Befundungskriterien in der Lymphknotendiagnostik beim NSCLC und der potentielle Einfluss der Bildrekonstruktion auf die diagnostische Genauigkeit begünstigen die hohe Variabilität der Sensitivität (47-100%) und Spezifität (73-100%) in den Einzelstudien der o.g. Meta-Analyse (40). Neben dem Kriterium einer „erhöhten Aktivität gegenüber dem Blutpool-Hintergrund“ (112) wurde „6 Stufen von eindeutig negativ bis eindeutig positiv verglichen mit dem mediastinalen Blutpool“ (113) oder „4 Stufen von einer Intensität unterhalb des mediastinalen Blutpools bis zu einer deutlich höheren Intensität als der mediastinale Blutpool“ (114) vorgeschlagen. Doch die Interobserver-Variabilität bleibt hoch (115, 116), und Ansätze für eine Standardisierung der visuellen Befundung wurden bislang nicht publiziert. Insbesondere für unerfahrene Befunder bedeutet dies ein erhöhtes Maß an Befundungsunsicherheit und lange Lernprozesse. Erschwert wird dies durch stetige Fortschritte in der PET-Detektortechnologie und Bildrekonstruktion, die zu Änderungen in der Bildqualität und im visuellen Erscheinungsbild von Lymphknoten führen. Deshalb gab es wiederholte Bestrebungen, diagnostische SUV-Schwellwerte für Lymphknoten zu etablieren, die gegenüber der visuellen Befundung eine Objektivierung bieten sollten. Doch bereits in Kollektiven, die vergleichsweise einheitlich mit PET-Systemen ohne TOF- und PSF-Rekonstruktion untersucht worden sind, wurden optimale SUVmax-Schwellwerte zwischen 2,5 und 5,3 publiziert (40, 117, 118) und somit erneut keine Vereinheitlichung erreicht.

Bedeutung für die eigenen Arbeiten:

Die visuelle Befundung thorakaler Lymphknoten bei Patienten mit NSCLC bleibt klinischer Standard. Um die Zuverlässigkeit und Reproduzierbarkeit der Befundung zu verbessern, bedarf es jedoch einer standardisierten Befundungsmethodik. Ob auch das DTPI eine relevante Verbesserung der visuellen Befundung ermöglicht, ist nicht abschließend geklärt.

1.3.2 Quantifizierung, Risiko- und Prognoseabschätzung

Wie im Kapitel 1.1.3 erläutert, basiert die Therapieempfehlung bei Patienten mit kurativer Therapieintention maßgeblich auf dem klinischen Tumorstadium (10, 15). Dennoch ist das Outcome individueller Patienten auch innerhalb gleicher Tumorstadien keineswegs einheitlich. Diese Heterogenität könnte anteilig durch klinische Variablen wie ECOG-Status, Raucherstatus, Geschlecht und Alter erklärt werden, die zumindest bei Auswertungen sehr umfangreicher Patientenkollektive wie der Krebsregisterauswertung von Kawaguchi *et al.* mit fast 27.000 Lungenkarzinompatienten unabhängig vom Tumorstadium I-IV einen prognostischen Stellenwert aufwiesen (119). Weitere prognostische Bedeutung bei Patienten mit NSCLC im Stadium I-III konnte für tumorbiologische Faktoren wie dem histologischen Differenzierungsgrad (120) und eine Hämangiosis carcinomatosa (V1) (120, 121) gezeigt werden. Gleichzeitig besteht zunehmende Evidenz, dass bildgebende Charakteristika des Primärtumors und des Tumormilieus in der FDG-PET/CT Routinediagnostik über die visuelle Befundung des klinischen Tumorstadiums hinaus prognostische und prädiktive Informationen bei Patienten mit NSCLC bieten.

1.3.2.1 Bildparameter 1. Ordnung („first-order“)

Eine Meta-Analyse von Paesmans *et al.* (2015) untersuchte den prognostischen Wert des Primärtumor-SUVmax bei Patienten mit NSCLC im Stadium I-III bzw. I-IV in einer gepoolten Analyse individueller Studienergebnisse mit >1.500 Patienten. Der jeweilige mediane SUVmax der einzelnen Kohorten (Spannweite, 2,3 bis 14,3) wurde ohne weitere Schwellwertoptimierung zur Klassifikation der Patienten in hohes vs. niedriges Risiko genutzt. In der multivariablen Cox-Regression war der SUVmax (Hazard Ratio ca. 1,5) prädiktiv für das OS unabhängig von Stadium, Alter und Primärtumordurchmesser (in Stadium I-III und auch Stadium I-IV) (122). Damit konnte belegt werden, dass ein metabolischer Parameter der FDG-PET Bildgebung einen signifikanten Zusatznutzen für die prognostische Klassifikation der Patienten ermöglicht. Ein vereinheitlichter SUVmax-Schwellwert wurde jedoch nicht untersucht; somit bleibt unklar, ob trotz der sehr unterschiedlichen medianen SUVmax-Werte in den Einzelstudien ein übergreifend gültiger Schwellwert definierbar wäre, und insbesondere wäre die Übertragbarkeit auf aktuelle PET-Systeme und Bildrekonstruktionen fraglich. Nair *et al.* zeigten, dass bei Patienten mit NSCLC (>95% im Stadium I- II) eine signifikante Assoziation zwischen verschiedenen SUV-Parametern und SUV-Histogramm-Variablen mit der Expression unterschiedlicher Gene des Zellzyklus, der Proliferation und extrazellulären Matrix besteht (123).

1.3.2.2 Form- und Größenparameter („shape and size“)

Eine Meta-Analyse von Liu *et al.* (2016) zeigte eine unabhängige prognostische Bedeutung des Primärtumor-MTV und des Produkts aus MTV*SUVmean (tumor lesion glycolysis; TLG) hinsichtlich des DFS und OS bei >1.500 Patienten mit NSCLC im Stadium I-III A (<5% der Patienten im Stadium IIIB und IV). Die MTV-Segmentierung war in 6/7 Einzelstudien schwellwertabhängig erfolgt, meist mit einem festen absoluten SUV-Schwellwert von 2,5 oder einem festen relativen Schwellwert zwischen 42 und 50% des SUVmax (124).

Hatt *et al.* konnten bei 87 Patienten mit NSCLC im Stadium IIA-IIIB zeigen, dass das Primärtumor-MTV und der Formparameter sphericity jeweils mit dem OS assoziiert waren und insbesondere die Kombination beider Parameter eine signifikante Differenzierung von Patienten mit günstigem vs. ungünstigem Outcome ermöglichte (Kaplan-Meier-Analyse ohne Adjustierung für sonstige Risikofaktoren). Der optimale Schwellwert für MTV und sphericity unterschied sich deutlich zwischen den 4 untersuchten Abgrenzungsmethoden (125).

Leijenaar *et al.* berichteten, dass verschiedene first-order-, „shape and size“-, second-order- und high-order-Parameter bei Patienten mit NSCLC überwiegend eine hohe Retest-Variabilität und Interobserver-Variabilität erlaubten (Intraklassen-Korrelationskoeffizient [ICC] $\geq 0,8$). Die Tumoren wurden jeweils durch mehrere Befunder manuell abgegrenzt (= Interobserver-Variabilität) oder in zwei wiederholten FDG-PET/CT Untersuchungen mit demselben Verfahren schwellwertbasiert abgegrenzt (50% des SUVmax; = Retest-Variabilität). Im Durchschnitt war die Variabilität für first-order und „shape and size“-Parameter (u.a. sphericity) am geringsten (126).

1.3.2.3 Bildparameter 2. und höherer Ordnung (“second-order” und “high-order”)

Ahn *et al.* untersuchten die prognostische Aussagekraft von >20 first-order-Bildparametern, „shape and size“-Parametern und high-order-Parametern aus der FDG-PET bei 93 Patienten mit NSCLC im Stadium I-III hinsichtlich des Rezidivstatus nach 3 Jahren. Die MTV-Abgrenzung erfolgte mithilfe einer gradientenbasierten Methode. Das Machine-Learning-Verfahren Random Forest war in der Vorhersage des Rezidivstatus am zuverlässigsten (höchste AUC von 0,96), gefolgt von einem neuronalen Netz (AUC, 0,87). Der high-order-Parameter contrast (aus der neighbourhood grey-tone difference matrix; NGTDM) wurde als aussagekräftigster Einzelparameter identifiziert und zeigte auch in der multivariablen Cox-Regression zusätzlich zum UICC-Stadium III vs. I/II einen signifikanten prognostischen Wert (69). In einer Auswertung von Yan *et al.* von 20 Patienten mit pulmonalen Läsionen (17 davon NSCLC) war die Variabilität dieses Parameters contrast zwischen unterschiedlichen Bildrekonstruktionen (OSEM/PSF/TOF, Anzahl der Iterationen, Breite des Post-Processing-Filters, Matrixgröße) jedoch vergleichsweise hoch (Variationskoeffizient [COV] meist >10%) (127). Ein second-order-Parameter mit überwiegend geringer Variabilität (COV meist $\leq 5\%$) war demnach die entropy aus der GLCM (127). Hatt *et al.* und Desseroit *et al.* konnten für die entropy einen signifikanten Zusammenhang mit dem OS bei Patienten mit NSCLC im Stadium I-III zeigen (128, 129), unsere eigene Arbeitsgruppe bei 83 Patienten im Stadium I-IV hingegen nicht (130).

Für keinen der o.g. Texturparameter aus der FDG-PET/CT Bildgebung liegen ausreichende klinische und methodische Validierungsstudien vor, die eine klinische Anwendung für Prognostik und Prädiktion bei Patienten mit NSCLC erlauben.

Bedeutung für die eigenen Arbeiten:

Ein Zusatznutzen für prognostische Abschätzungen bei Patienten mit NSCLC konnte mittlerweile für verschiedene PET-Bildparameter gezeigt werden, darunter auch Formparameter. Die Überlegenheit einzelner Parameter unter Berücksichtigung sowohl der prognostischen Wertigkeit als auch der Reproduzierbarkeit ist hingegen nicht ausreichend belegt. Validierungen an unabhängigen Kollektiven, möglichst auch für unterschiedliche PET-Systeme bzw. Rekonstruktionsverfahren sind unerlässlich für die Beurteilung, ob und unter welchen Bedingungen ein Bildparameter für die klinische Anwendung geeignet ist.

1.4 Eigene Ziele und Fragestellungen

Ziel der Arbeiten, die dieser Habilitationsschrift zugrunde liegen, war die Entwicklung und Evaluation verschiedener Methoden für die Optimierung der prätherapeutischen Lymphknotendiagnostik und Risikostratifizierung bei Patienten mit NSCLC unter Einsatz der FDG-PET/CT Diagnostik. Dies beinhaltet folgende Fragestellungen im Detail:

1.4.1 Prätherapeutische Lymphknotendiagnostik

- Kann die diagnostische Genauigkeit der Lymphknotendiagnostik durch SUV-Parameter verbessert werden?
- Hilft das Dual Time Point Imaging (DTPI) in der Differenzierung maligner und benigner Lymphknoten?
- Kann durch eine Standardisierung der visuellen Lymphknotenbefundung eine hohe diagnostische Genauigkeit und gleichzeitig eine hohe Interobserver-Übereinstimmung erreicht werden?
- Verbessert ein neuartiges Rekonstruktionsverfahren (GE Q.Clear) für die FDG-PET/CT die Bildqualität gegenüber herkömmlichen iterativen Rekonstruktionsalgorithmen?

1.4.2 Quantifizierung, Risiko- und Prognoseabschätzung

- Bietet der Formparameter asphericity (ASP) in der prätherapeutischen FDG-PET/CT einen unabhängigen prognostischen Stellenwert bei Patienten mit NSCLC?
- Kann dieser prognostische Stellenwert auch in einem unabhängigen Kollektiv und trotz Verwendung eines anderen PET-Systems bestätigt werden?
- Wie groß ist die Variabilität der ASP bei Verwendung konventioneller und neuartiger PET-Bildrekonstruktionsverfahren? Wie klinisch relevant ist diese Variabilität? Ist eine Harmonisierung der PET-Bildrekonstruktionsparameter möglich, um diese Variabilität zu minimieren und multizentrische Studien zu ermöglichen?

2. Eigene Arbeiten

2.1 Prätherapeutische Lymphknotendiagnostik

2.1.1 Dual Time Point Imaging für die prätherapeutische Lymphknotendiagnostik

Bislang existieren keine standardisierten Befundungskriterien für die thorakale Lymphknotendiagnostik bei Patienten mit NSCLC, weder visuell noch SUV-basiert noch für das DTPI. Uesaka *et al.* beobachteten eine ähnliche relative SUV-Änderung (sog. Retentionsindex; RI) von Primärtumoren und Metastasen des NSCLC vom frühen zum späten DTPI-Zeitpunkt. Wurde ein Verhältnis der RI von Lymphknoten zu Primärtumor zwischen 0,5 und 2,0 als positiv gewertet, betrug die Sensitivität und Spezifität 100% bzw. 98% (131). Eine externe Validierung dieser Methodik steht bislang aus.

Die nachfolgende Publikation diente dem systematischen Vergleich der diagnostischen Genauigkeit der prätherapeutischen thorakalen Lymphknotenbefundung mit SUV-Parametern, einer eigens entwickelten standardisierten visuellen Befundung und der Methodik von Uesaka *et al.* Für alle Kriterien wurde zudem der diagnostische Mehrwert des DTPI untersucht.

Der nachfolgende Text entspricht dem originalsprachigen Abstrakt der Publikation

Rogasch JMM, Steffen IG, Riedel S, *et al.* Dual time point imaging for F18-FDG-PET/CT does not improve the accuracy of nodal staging in non-small cell lung cancer patients. *Eur Radiol.* 2016 Aug;26(8):2808-18. <https://doi.org/10.1007/s00330-015-4093-5>

“Objectives: To analyze the diagnostic performance of dual time point imaging (DTPI) for pre-therapeutic lymph node (LN) staging in non-small cell lung cancer (NSCLC).

Methods: This was a retrospective analysis of 47 patients with NSCLC who had undergone DTPI by PET (early + delayed) using F18-fluorodeoxyglucose (FDG). PET raw data were reconstructed iteratively (point spread function + time-of-flight). LN uptake in PET was assessed visually (four-step score) and semi-quantitatively (SUVmax, SUVmean, ratios LN/primary, LN/liver, and LN/mediastinal blood pool). DTPI analyses included retention indices (RIs), Δ -ratios and changes in visual score. Histology or cytology served as standards of reference. Accuracy was determined based on ROC analyses.

Results: Thirty-six of 155 LNs were malignant. DTPI accuracy was low for all measures (visual assessment, 24.5%; RI SUVmax, 68.4%; RI SUVmean, 65.8%; Δ -ratios, 63.9-76.1%) and significantly inferior to early PET. Accuracies of early (range, 86.5-92.9%) and delayed PET (range, 85.2-92.9%) were comparable. At early PET, accuracy of the visual score (92.9%) was similar or superior to semi-quantitative analyses (range, 86.5-92.3%).

Conclusions: Using a modern PET/CT device and novel image reconstruction, neither additional delayed PET nor DTPI analyses improved the accuracy of PET-based LN staging. Dedicated visual assessment criteria performed very well.”

2.1.2 Interobserver-Übereinstimmung der standardisierten visuellen Befundung

In obenstehender Publikation war das DTPI ohne Zusatznutzen für die Lymphknotenbefundung bei Patienten mit NSCLC. Hingegen wurde mit der eigens entwickelten standardisierten visuellen Befundung basierend auf einer vereinheitlichten Bildfensterung und einer definierten 4-stufigen Skala eine sehr hohe diagnostische Genauigkeit erreicht. Ungeklärt blieb zunächst jedoch, ob die vereinheitlichte Methodik auch eine hohe Reproduzierbarkeit in der visuellen Befundung ermöglicht.

Das Ziel der nachfolgenden Publikation war deshalb die Untersuchung der Interobserver-Übereinstimmung bei Anwendung der entwickelten Befundungsmethodik. Hierzu wurden Befunder mit sehr unterschiedlichem Erfahrungsstand (Anfänger bis Experte) verglichen. Somit konnte zusätzlich evaluiert werden, ob die standardisierte Methodik ermöglicht, die Befundungsqualität bei unerfahrenen Befundern unmittelbar zu verbessern.

Der nachfolgende Text entspricht dem originalsprachigen Abstrakt der Publikation

Rogasch JMM, Apostolova I, Steffen IG, *et al.* Standardized visual reading of F18-FDG-PET in patients with non-small cell lung cancer scheduled for preoperative thoracic lymph node staging. *Eur J Radiol.* 2016 Aug;85(8):1345-50. <https://doi.org/10.1016/j.ejrad.2016.05.004>

“Objectives: Routine visual assessment of positron emission tomography (PET) for thoracic lymph node (LN) staging in patients with non-small cell lung cancer (NSCLC) is limited by a lack of reliable assessment criteria. This study evaluates the accuracy and inter-rater agreement of a standardized approach with unified windowing and a PET-based visual score.

Materials and methods: This retrospective analysis included pretherapeutic FDG-PET data of 86 patients with NSCLC. After standardized windowing (threshold: $2 \times \text{liver SUV}_{\text{mean}}$) the LN uptake was assessed visually by three independent readers with varying levels of experience using a 4-step score (1, LN uptake \leq mediastinal blood pool structures (MBPS); 2, MBPS < LN < liver; 3, liver \leq LN < 'black'; 4, LN appears 'black'). ROC analyses and respective areas under the curve (AUC) based on histology/cytology as standard of reference. Agreement was analyzed with Cohen's kappa (κ , pairwise) and Fleiss' κ (overall). Subgroup analyses separated between hilar vs. mediastinal LNs, adenocarcinoma vs. squamous cell carcinoma and grading G1/2 vs. G3/4.

Results: Fifty-four of the 278 LNs (19.4%) were malignant (optimal cut-off to differentiate benign vs. malignant, score >3). The inexperienced (n=1), advanced (n=1), and expert readers (n=1) achieved similar accuracies of 93.5%, 91.4% and 92.1%, respectively ($P > 0.05$ each). Cohen's κ ranged from 0.92 to 0.96 and Fleiss' κ was 0.93. ROC-analyses showed no significant differences between attendant readers within any subgroup (AUC, 0.92-0.96).

Conclusion: Applying unified windowing, the introduced PET-score achieved highly accurate and robust LN assessment. This approach may shorten learning curves of inexperienced readers, facilitate multicenter trials, and improve comparability of future studies.”

2.1.3 Einfluss neuartiger PET-Bildrekonstruktionsverfahren auf die Bildqualität

In den vorangehenden Arbeiten wurde ein standardisierter Workflow für eine zuverlässige, reproduzierbare Lymphknotenbefundung evaluiert. Eine zuverlässige visuelle Befundung bedarf jedoch einer hochwertigen Bildqualität.

Bildqualität in der FDG-PET Bildgebung ist eine Balance zwischen akkurater Recovery wahrer fokaler Aktivitäten und minimalem Bildrauschen (siehe Kapitel 1.2.2). Bei konventionellen iterativen Rekonstruktionsverfahren wie dem OSEM-Verfahren erfolgt eine fest vordefinierte Anzahl an Iterationsschritten, mit denen die Recovery von fokalen Mehranreicherungen stetig zunimmt. Da die Anzahl der Iterationen jedoch undifferenziert für den gesamten PET-Datensatz festgelegt wird, nimmt in ähnlicher Weise die Varianz im Bildhintergrund zu, d.h. auch in eigentlich homogen anreichernden Strukturen (132). Eine reduzierte Zahl an Iterationen oder nachträgliche Glättung der Bilddaten (Post-Processing-Filter) mit dem Ziel eines geringeren Bildrauschens verringert wiederum die Recovery. Die Einführung von TOF und PSF kann das Verhältnis zwischen Recovery und Bildrauschen systematisch verbessern (45, 48), der grundlegend notwendige Kompromiss bleibt jedoch bestehen.

Mit dem Anspruch, dieses Dilemma der OSEM-Rekonstruktion zu beheben, wurden die sogenannten Bayesian penalized likelihood-Rekonstruktionsverfahren eingeführt (z.B. „Q.Clear“ der Firma GE Healthcare). Q.Clear ist ebenfalls iterativ, jedoch wird basierend auf einem benutzerabhängig vordefinierten Regulationsfaktor (penalization factor) β der Abbruch der Iterationsschritte während der Rekonstruktion dynamisch reguliert. Sobald die relative Änderung der Voxelwerte ein definiertes Maß unterschreitet, werden die Iterationen beendet, um die Entwicklung von Bildrauschen zu unterdrücken. Der Zeitpunkt dieses Iterationsstopps wird voxelweise angepasst, sodass fokale Aktivitätsmaxima unabhängig von Voxeln im (homogenen) Hintergrund ausgebildet werden können. Dies kann das Verhältnis von Recovery und SNR gegenüber OSEM verbessern (133). Der Zugewinn an Recovery ist jedoch nicht uniform, sondern ebenso wie das Ausmaß des Bildrauschens nun von der Höhe des Regulationsfaktors β abhängig (134, 135).

Die nachfolgenden Auswertungen umfassten deshalb vergleichende Untersuchungen der TOF-, PSF- und Q.Clear-Rekonstruktion hinsichtlich verschiedener Bildqualitätsparameter. Das Ziel war, den potenziellen Bildqualitätszugewinn der Q.Clear-Rekonstruktion in Abhängigkeit vom Regulationsfaktor β zu identifizieren und damit den optimalen β -Wert für die visuelle Befundung in der klinischen Routine mit dem eigenen PET-System zu definieren.

Der nachfolgende Text entspricht dem originalsprachigen Abstrakt der Publikation

Rogasch JM, Suleiman S, Hofheinz F, *et al.* Reconstructed spatial resolution and contrast recovery with Bayesian penalized likelihood reconstruction (Q.Clear) for FDG-PET compared to time-of-flight (TOF) with point spread function (PSF). EJNMMI Phys. 2020 Dec;7:2. <https://doi.org/10.1186/s40658-020-0270-y>

“Background: Bayesian penalized likelihood reconstruction for PET (e.g., GE Q.Clear) aims at improving convergence of lesion activity while ensuring sufficient signal-to-noise ratio (SNR). This study evaluated reconstructed spatial resolution, maximum/peak contrast recovery (CR_{max}/CR_{peak})

and SNR of Q.Clear compared to time-of-flight (TOF) OSEM with and without point spread function (PSF) modeling.

Methods: The NEMA IEC Body phantom was scanned five times (3 min scan duration, 30 min between scans, background, 1.5–3.9 kBq/ml F18) with a GE Discovery MI PET/CT (3-ring detector) with spheres filled with 8-, 4-, or 2-fold the background activity concentration (SBR 8:1, 4:1, 2:1). Reconstruction included Q.Clear (beta, 150/300/450), “PSF+TOF_{4/16}” (iterations, 4; subsets, 16; in-plane filter, 2.0 mm), “OSEM+TOF_{4/16}” (identical parameters), “PSF+TOF_{2/17}” (2 it, 17 ss, 2.0 mm filter), “OSEM+TOF_{2/17}” (identical), “PSF+TOF_{4/8}” (4 it, 8 ss, 6.4 mm), and “OSEM+TOF_{2/8}” (2 it, 8 ss, 6.4 mm). Spatial resolution was derived from 3D sphere activity profiles. RC as (sphere activity concentration [AC]/true AC). SNR as (background mean AC/background AC standard deviation).

Results: Spatial resolution of Q.Clear₁₅₀ was significantly better than all conventional algorithms at SBR 8:1 and 4:1 (Wilcoxon, each $p < 0.05$). At SBR 4:1 and 2:1, the spatial resolution of Q.Clear_{300/450} was similar or inferior to PSF+TOF_{4/16} and OSEM+TOF_{4/16}. Small sphere CRpeak generally underestimated true AC, and it was similar for Q.Clear_{150/300/450} as with PSF+TOF_{4/16} or PSF+TOF_{2/17} (i.e., relative differences < 10%). Q.Clear provided similar or higher CRpeak as OSEM+TOF_{4/16} and OSEM+TOF_{2/17} resulting in a consistently better tradeoff between CRpeak and SNR with Q.Clear. Compared to PSF+TOF_{4/8}/OSEM+TOF_{2/8}, Q.Clear_{150/300/450} showed lower SNR but higher CRpeak.

Conclusions: Q.Clear consistently improved reconstructed spatial resolution at high and medium SBR compared to PSF+TOF and OSEM+TOF, but only with beta = 150. However, this is at the cost of inferior SNR with Q.Clear₁₅₀ compared to Q.Clear_{300/450} and PSF+TOF_{4/16}/PSF+TOF_{2/17} while CRpeak for the small spheres did not improve considerably. This suggests that Q.Clear_{300/450} may be advantageous for the 3-ring detector configuration because the tradeoff between CR and SNR with Q.Clear_{300/450} was superior to PSF+TOF_{4/16}, OSEM+TOF_{4/16}, and OSEM+TOF_{2/17}. However, it requires validation by systematic evaluation in patients at different activity and acquisition protocols.”

2.2 Quantifizierung, Risiko- und Prognoseabschätzung

2.2.1 Prognostischer Stellenwert der ASP beim NSCLC: Explorative Studie

Bei Patienten mit Kopf-Hals-Karzinom war die asphericity (ASP) des Primärtumor-MTV ein unabhängiger Prädiktor des PFS und OS in zwei unabhängigen Kohorten (66, 68) (siehe Kapitel 1.2.4.1). Ziel der nachfolgenden explorativen Studie war die Untersuchung, ob ein solcher prognostischer Stellenwert auch bei Patienten mit NSCLC nachgewiesen werden kann. Die ASP wurde ergänzend zur solidity (einem weiteren „shape and size“-Parameter (67)), dem Primärtumor-SUVmax, dem MTV-Volumen und der TLG erhoben. Die MTV-Abgrenzung erfolgte mittels hintergrundadaptierten, SUV-basierten Schwellwerts (58).

Der nachfolgende Text entspricht dem originalsprachigen Abstrakt der Publikation

Apostolova I, **Rogasch J**, Buchert R, *et al.* Quantitative assessment of the asphericity of pretherapeutic FDG uptake as an independent predictor of outcome in NSCLC. BMC Cancer. 2014 Dec 1;14:896. <https://doi.org/10.1186/1471-2407-14-896>

“Background: The aim of the present study was to evaluate the predictive value of a novel quantitative measure for the spatial heterogeneity of FDG uptake, the asphericity (ASP) in patients with non-small cell lung cancer (NSCLC).

Methods: FDG-PET/CT had been performed in 60 patients (15 women, 45 men; median age, 65.5 years) with newly diagnosed NSCLC prior to therapy. The FDG-PET image of the primary tumor was segmented using the ROVER 3D segmentation tool based on thresholding at the volume-reproducing intensity threshold after subtraction of local background. ASP was defined as the relative deviation of the tumor’s shape from a sphere. Univariate and multivariate Cox regression as well as Kaplan-Meier (KM) analysis and log-rank test with respect to overall (OAS) and progression-free survival (PFS) were performed for clinical variables, SUVmax/mean, metabolically active tumor volume (MTV), total lesion glycolysis (TLG), ASP and “solidity”, another measure of shape irregularity.

Results: ASP, solidity and “primary surgical treatment” were significant independent predictors of PFS in multivariate Cox regression with binarized parameters (HR, 3.66; $p < 0.001$, HR, 2.11; $p = 0.05$ and HR, 2.09; $p = 0.05$), ASP and “primary surgical treatment” of OAS (HR, 3.19; $p = 0.02$ and HR, 3.78; $p = 0.01$, respectively). None of the other semi-quantitative PET parameters showed significant predictive value with respect to OAS or PFS. Kaplan-Meier analysis revealed a probability of 2-year PFS of 52% in patients with low ASP compared to 12% in patients with high ASP ($p < 0.001$). Furthermore, it showed a higher OAS rate in the case of low versus high ASP (1-year-OAS, 91% vs. 67%; $p = 0.02$).

Conclusions: The novel parameter asphericity of pretherapeutic FDG uptake seems to provide better prognostic value for PFS and OAS in NSCLC compared to SUV, metabolic tumor volume, total lesion glycolysis and solidity.”

2.2.2 Prognostischer Stellenwert der ASP beim NSCLC: Konfirmatorische Studie

Eine Grundvoraussetzung für eine mögliche zukünftige klinische Anwendung quantitativer Parameter aus der FDG-PET/CT Bildgebung für die Prognoseabschätzung bei Patienten mit NSCLC ist eine Übertragbarkeit des prognostischen Stellenwertes auf weitere Zentren und weitere PET-Systeme. Zudem ist eine klinische Validierung in einem unabhängigen Patientenkollektiv erforderlich. Die initiale, explorative Studie zur ASP bei Patienten mit NSCLC umfasste zudem ein klinisch heterogenes Kollektiv von Patienten mit Stadium I-IV.

Das Ziel der konfirmatorischen Studie war somit die Bestätigung der unabhängigen prognostischen Wertigkeit der ASP in einem größeren, klinisch homogeneren Kollektiv (nur Patienten mit kurativ intendierter Therapie inkl. Operation) unter Verwendung eines anderen PET-Systems und anderer Bildrekonstruktionsverfahren (PSF+TOF anstelle von TOF alleine). Die höhere Fallzahl sollte zudem Subanalysen ermöglichen, um eine differenziertere Aussage über den prognostischen Wert der ASP in Abhängigkeit von klinischen Faktoren (v.a. UICC-Stadium) zu erlauben. Die Interobserver-Variabilität von MTV und ASP wurde in einer repräsentativen Subgruppe evaluiert (*Supplementary Material*).

Der nachfolgende Text entspricht dem originalsprachigen Abstract der Publikation

Rogasch JMM*, Furth C*, Chibolela C, *et al.* Validation of Independent Prognostic Value of Asphericity of 18 F-Fluorodeoxyglucose Uptake in Non-Small-Cell Lung Cancer Patients Undergoing Treatment With Curative Intent. Clin Lung Cancer. 2020 May;21(3):264-272.e6. <https://doi.org/10.1016/j.clc.2019.10.001>

* *geteilte Erstautorenschaft*

“Background: In patients with non-small-cell lung cancer (NSCLC), asphericity (ASP) of the primary tumor’s metabolic tumor volume (MTV) has shown prognostic significance. This study aimed at validation in an independent and sufficiently large cohort.

Patients and Methods: A retrospective study was performed of 311 NSCLC patients undergoing ¹⁸F-fluorodeoxyglucose positron emission tomography / computed tomography (¹⁸F-FDG PET/CT) before curatively intended treatment (always including surgery). A total of 140 patients had International Union Against Cancer (UICC) stage I disease, 78 had stage II disease, and 93 had stage III disease (adenocarcinoma, n = 153; squamous-cell carcinoma, n = 141). Primary tumor MTV was delineated with semiautomated background-adapted threshold relative to the standardized maximum uptake value (SUVmax). Cox regression (progression-free survival [PFS] and overall survival [OS]) analysis for positron emission tomography (MTV, ASP, SUVmax) as well as for clinical (T/N descriptor, UICC stages), histologic, and treatment variables (Rx/1 vs. R0 resection, chemotherapy/radiotherapy yes/no) were performed.

Results: Events (progression and relapse) occurred in 167 of 311 patients; 137 died (median survivor follow-up, 37 months). In multivariable Cox regression for OS, ASP > 33.3% (hazard ratio, 1.58 [1.04-2.39]), male sex (1.84), age (1.04 per year), Eastern Cooperative Oncology Group performance status ≥ 2 versus 0/1 (2.68), stage II versus I (1.96), and Rx/1 versus R0 resection (2.1) were significant. Among separate UICC stages, ASP only predicted OS in stage II (optimal, > 19.5%; median OS, 33 vs. 59 months). Regarding PFS, ASP > 21.2%, male sex, Eastern Cooperative Oncology Group performance status ≥ 2 , stage II versus I disease, and Rx/1 resection were prognostic. ASP remained prognostic for stage II disease (optimal, > 19.5%; PFS, 12 vs. 47 months). Log-rank test for ASP was significant at any cutoff $\geq 18\%$ (OS) or from 9% to 59% (PFS).

Conclusion: ASP was validated as prognostic factor for PFS and OS in patients with NSCLC and curative treatment intent, especially stage II. High ASP in stage II could imply intensified treatment or intensified follow-up.”

2.2.3 Variabilität der ASP unter dem Einfluss der Bildrekonstruktion

Wie im Kapitel 1.2.4.2 erläutert, unterliegen sämtliche quantitative PET-Bildparameter dem Einfluss der Bilddatenrekonstruktion, somit auch die ASP. Yan *et al.* zeigten am Beispiel von SUV, second-order- und high-order-Parametern, dass diese einer unterschiedlichen Variabilität aufgrund des Rekonstruktionsverfahrens, der Breite des Post-Processing-Filters, der Iterationszahl und der Voxelgröße unterliegen. Die Variation der Bildparameter war im Durchschnitt am höchsten für verschiedene Filterbreiten und Voxelgrößen (127).

Da identische Bildrekonstruktionen zwischen verschiedenen Zentren kaum gewährleistet werden können und auch identische Rekonstruktionsparameter bei unterschiedlichen Geräten nicht zu identischen Bilddaten führen würden, bedarf es einer Harmonisierung basierend auf Charakteristika der finalen Bilddaten. Wie ebenfalls unter 1.2.4.2 dargestellt, zielen Harmonisierungsansätze des EARL-Konsortiums primär auf eine vergleichbare Recovery ab. Vorrangiges Ziel ist somit eine Vergleichbarkeit von SUV-Werten (84, 85). Hingegen ist die ASP als Formparameter *a priori* unabhängig von der Aktivitätsverteilung innerhalb des abgegrenzten MTV, sondern abhängig von der Abbildungsfähigkeit der Irregularität der MTV-Oberfläche. Deshalb postulierten wir, dass die rekonstruierte räumliche Auflösung ein zuverlässigeres Surrogat für die Vergleichbarkeit von Bildrekonstruktionen hinsichtlich der ASP ist.

Das Ziel der nachfolgenden Auswertungen war deshalb, verschiedene klinisch übliche Bildrekonstruktionsverfahren und Kombinationen aus Iterationen und Subsets hinsichtlich ihres Einflusses auf die ASP zu untersuchen. Die rekonstruierte räumliche Auflösung wurde dafür mittels radialer Aktivitätsprofile der füllbaren Kugeleinsätze am Phantom bestimmt (46, 58). Die klinische Relevanz der ASP-Variabilität bei Patienten mit NSCLC wurde anhand der Rate von Fehlklassifikationen in Gruppen hoher vs. niedriger ASP untersucht. Abschließend wurde untersucht, ob eine nachträgliche Filterung von Bilddaten mit dem Ziel einer einheitlichen rekonstruierten räumlichen Auflösung eine Harmonisierung im Rahmen multizentrischer Auswertungen ermöglicht. Der Einfluss einer variablen Voxelgröße konnte nicht untersucht werden, da diese während der Bildrekonstruktion softwareseitig nicht frei wählbar war.

Der nachfolgende Text entspricht dem originalsprachigen Abstrakt der Publikation

Rogasch JMM, Furth C, Bluemel S, *et al.* Asphericity of tumor FDG uptake in non-small cell lung cancer: Reproducibility and implications for harmonization in multicenter studies. *EJNMMI Res.* 2020 Nov 2;10(1):134. <https://doi.org/10.1186/s13550-020-00725-y>

“Background: Asphericity (ASP) of the primary tumor’s metabolic tumor volume (MTV) in FDG-PET/CT is independently predictive for survival in patients with non-small cell lung cancer (NSCLC). However, comparability between PET systems may be limited. Therefore, reproducibility of ASP was evaluated at varying image reconstruction and acquisition times to assess feasibility of ASP assessment in multicenter studies.

Methods: Retrospective study of 50 patients with NSCLC (female, 20; median age, 69 years) undergoing pretherapeutic FDG-PET/CT (median, 3.7 MBq/kg; 180 s/bed position). Reconstruction used OSEM with TOF_{4/16} (iterations, 4; subsets, 16; in-plane filter, 2.0, 6.4 or 9.5 mm), TOF_{4/8} (4 it; 8 ss; filter, 2.0/6.0/9.5 mm), PSF+TOF_{2/17} (2 it; 17 ss; filter, 2.0/7.0/10.0 mm) or Bayesian penalized likelihood (Q.Clear; beta, 600/1750/4000). Resulting reconstructed spatial resolution (FWHM) was determined from hot sphere inserts of a NEMA IEC phantom. Data with approx. 5 mm FWHM were retrospectively smoothed to achieve 7 mm FWHM. List mode data were rebinned for acquisition times of 120/90/60s. Threshold-based delineation of primary tumor MTV was followed by evaluation of relative ASP/SUVmax/MTV differences between datasets and resulting proportions of discordantly classified cases.

Results: Reconstructed resolution for narrow/medium/wide in-plane filter (or low/medium/high beta) was approx. 5/7/9 mm FWHM. Comparing different pairs of reconstructed resolution between TOF_{4/8}, PSF+TOF_{2/17}, Q.Clear and the reference algorithm TOF_{4/16}, ASP differences were lowest at FWHM of 7 vs. 7 mm. Proportions of discordant cases (ASP >19.5% vs. ≤19.5%) were also lowest at 7 mm (TOF_{4/8}, 2%; PSF+TOF_{2/17}, 4%; Q.Clear, 10%). Smoothing of 5 mm data to 7 mm FWHM significantly reduced discordant cases (TOF_{4/8}, 38% reduced to 2%; PSF+TOF_{2/17}, 12% to 4%; Q.Clear, 10% to 6%) resulting in proportions comparable to original 7 mm data. Shorter acquisition time only increased proportions of discordant cases at <90s.

Conclusions: ASP differences were mainly determined by reconstructed spatial resolution, and multicenter studies should aim at comparable FWHM (e.g., 7 mm; determined by in-plane filter width). This reduces discordant cases (high vs. low ASP) to an acceptable proportion for TOF and PSF+TOF of <5% (Q.Clear: 10%). Data with better resolution (i.e. lower FWHM) could be retrospectively smoothed to the desired FWHM resulting in a comparable number of discordant cases.”

3. Diskussion

3.1 Prätherapeutische Lymphknotendiagnostik

Das Ziel der beiden Arbeiten zur thorakalen Lymphknotendiagnostik bei Patienten mit NSCLC war die Verbesserung der diagnostischen Genauigkeit und Reproduzierbarkeit der FDG-PET/CT Bildgebung. Mithilfe des DTPI konnte die diagnostische Genauigkeit nicht verbessert werden, weder durch den zusätzlichen späten Einzelzeitpunkt noch durch die Bewertung der Dynamik von frühem zu spätem Zeitpunkt. Das DTPI war zudem weder für die diagnostische Korrektheit des Lymphknoten-SUVmax noch für verschiedene SUV-Verhältnisse von Lymphknoten zu Referenzgewebe oder für die visuelle Befundung vorteilhaft. Außerdem konnte mit dem Verhältnis von Retentionsindex (RI) des Lymphknotens zum RI des Primärtumors nicht die von Uesaka *et al.* erreichte, sehr hohe diagnostische Genauigkeit (131) reproduziert werden. Nennenswert ist der erhebliche Unterschied in der verwendeten PET-Technologie. Uesaka *et al.* führten die PET-Messungen mit einem nicht TOF fähigen PET-System ohne CT-Komponente durch, d.h. für die Schwächungskorrektur wurden noch Transmissionsquellen verwendet. Damit sind SUVmax-Werte der Lymphknoten – und somit auch die daraus abgeleiteten Parameter – nicht mit den Messwerten heutiger PET/CT-Systeme vergleichbar. Dennoch ist aus den aktuellen Ergebnissen zu schlussfolgern, dass das DTPI für die Lymphknotendiagnostik bei Patienten mit NSCLC unabhängig von der verwendeten Befundungsmethode keinen diagnostischen Zugewinn bietet, der eine routinemäßige Anwendung rechtfertigt. Dieses Fazit steht im Einklang mit einer kurz zuvor veröffentlichten Meta-Analyse von Shen *et al.* zu Einzelstudien mit SUV-basierter Befundung (107). Insbesondere ist die zusätzliche Strahlenexposition der Patienten (zusätzliche low-dose-CT für den zweiten Zeitpunkt) damit nicht zu rechtfertigen.

In beiden Publikationen zur Lymphknotenbefundung wurde mittels standardisierter visueller Bildbefundung eine sehr hohe diagnostische Genauigkeit von >90% erreicht. Diese Standardisierung beinhaltete neben einem definierten 4-stufigen Score für die Beurteilung der Anreicherungsintensität der Lymphknoten im Abgleich mit Referenzgewebe insbesondere auch eine vereinheitlichte Bildfensterung. Diese ist grundlegend erforderlich und entscheidend, um einen vereinheitlichten, reproduzierbaren Bildeindruck zu gewährleisten. Um der interindividuellen Variabilität der Aktivitätsverteilungen gerecht zu werden, wurde kein fester absoluter Schwellenwert genutzt (z.B. SUV = 3), sondern der SUVmean der Leber des jeweiligen Patienten. Die Bildfensterung definiert u.a. den oberen Grenzwert für die Graustufendarstellung, d.h. den Wert, ab welchem ein Bildpunkt mit dem Maximum der jeweiligen Skala abgebildet wird, ohne dass noch höhere Werte weiter differenziert werden können („schwarz“ in der verwendeten *inverse grey* Darstellung). Somit ist die Bildfensterung im Kontext des 4-stufigen Scores insbesondere für die Differenzierung von Score 3 (Anreicherung > Leberniveau) und Score 4 (Lymphknoten erscheint „schwarz“) wichtig. Werden nur Lymphknoten mit einem Score 4 als positiv gewertet (dies erreicht die höchste diagnostische Genauigkeit), ist die Fensterung damit direkt für die diagnostische Genauigkeit mitentscheidend. Bei einem niedrigeren Schwellenwert wäre die Bedeutung der Bildfensterung geringer, da diese nicht das Anreicherungsverhältnis von Lymphknoten gegenüber dem Referenzgewebe verändert.

Mit dieser vereinheitlichten Befundungsmethodik erreichten 3 Befunder von Anfänger bis Experte eine vergleichbare diagnostische Genauigkeit von jeweils >91%. Die Lernkurve für Anfänger kann somit verkürzt werden, und der Befundungsprozess kann vereinfacht, objektiviert und reproduzierbar werden. Zudem konnte für hiläre und mediastinale Lymphknotenstationen, Adenokarzinome und Plattenepithelkarzinome sowie Tumoren mit unterschiedlichem Differenzierungsgrad eine weitgehend gleichbleibende diagnostische Genauigkeit und Interobserver-Übereinstimmung erreicht werden. Die Übereinstimmung zwischen den Befundern war höher als in einer vorherigen Publikation von Hofman *et al.*, die ebenfalls 3 Befunder verglichen. Die Autoren verwendeten den mediastinalen Blutpool als Schwellenwert für die PET-Positivität von thorakalen Lymphknoten bei 100 Patienten mit NSCLC. Jedoch wurde auch die Erscheinung der Lymphknoten in der CT-Bildgebung und ihr Anreicherungsverhältnis gegenüber dem Primärtumor mitberücksichtigt. Die resultierende Intra- und Interobserver-Übereinstimmung waren mit einem $\kappa < 0,9$ geringer als in der eigenen Auswertung ($\kappa > 0,9$) und variierte deutlich zwischen hilären Lymphknoten und solchen in verschiedenen mediastinalen Lokalisationen ($\kappa = 0,48\text{--}0,88$) (115).

Dass die Sensitivität und Spezifität in beiden eigenen Auswertungen mit jeweils ca. 90% höher war als in einer Meta-Analyse von Schmidt-Hansen *et al.* (77% und 90%) (40), dürfte jedoch nicht nur Folge der standardisierten Befundungsmethodik sein. Insbesondere die vergleichbare Genauigkeit zwischen Patienten mit unterschiedlichem Histotyp oder unterschiedlichen Differenzierungsgraden ist nennenswert. Im Durchschnitt früherer Publikationen ist die Sensitivität der FDG-PET geringer für Lymphknotenmetastasen von Adenokarzinomen (40) und für Tumoren mit guter Differenzierung (36). Aufgrund der retrospektiven Durchführung der eigenen Auswertung kann ein Selektionsbias prinzipiell nicht ausgeschlossen werden, obgleich eine konsekutive Auswertung der Patienten erfolgte. Zudem diente als Referenzstandard grundsätzlich die Histologie des OP-Präparates oder – bei malignen Lymphknoten – auch das Ergebnis der EBUS-TBNA. Um falsch negative Befunde der EBUS-TBNA auszuschließen (gemäß Literatur ca. 10% (136)), wurden nur positive Resultate der EBUS-TBNA bewertet. Dies hat mutmaßlich zu einer falsch hohen Spezifität der FDG-PET/CT beigetragen. Die Sensitivität, auch für Metastasen von Adenokarzinomen, blieb dadurch jedoch unberührt. Auch der Einsatz eines modernen PET-Scanners einschließlich der Bilddatenrekonstruktion mit PSF-Kompensation und TOF-Information dürfte die hohe diagnostische Genauigkeit unterstützt haben. In allen Studien, die Grundlage der o.g. Meta-Analyse von Schmidt-Hansen *et al.* waren (40), kamen ausschließlich nicht TOF fähige PET-Systeme zum Einsatz. In eigenen früheren Arbeiten konnte gezeigt werden, dass die TOF-Rekonstruktion insbesondere in solchen Läsionen den SUV-Wert steigert, die einen geringen Kontrast zum Hintergrund aufweisen (82). Für die PSF-Rekonstruktion konnte ebenfalls eine gesteigerte Sensitivität für Lymphknotenmetastasen bei Patienten mit NSCLC gezeigt werden (108). Einschränkend gilt für die aktuellen Arbeiten, dass für die Entwicklung und Validierung der visuellen Befundungsmethodik derselbe PET-Scanner und dieselbe Bildrekonstruktion verwendet wurden. Eine Validierung der Methodik mit anderen PET-Systemen bzw. anderen Rekonstruktionsverfahren steht bislang aus. Da die Vergleichbarkeit der Anreicherungsintensität von Lymphknoten auch im visuellen Bildeindruck primär durch die Recovery bestimmt wird (nur

nachrangig durch die SNR), könnte eine Harmonisierung der Recovery entsprechend dem Ansatz des EARL-Konsortiums helfen, eine Übertragbarkeit auf andere PET-Systeme zu gewährleisten (84, 85). Der Einfluss der Bildrekonstruktion auf die Recovery, insbesondere, wenn diese basierend auf dem SUVmax ermittelt wird, wurde in der dritten Arbeit unterstrichen. Die Recovery wurde mithilfe von Kugeleinsätzen in einem standardisierten Phantom bestimmt. Mit allen Rekonstruktionsverfahren (OSEM mit TOF, OSEM mit PSF+TOF und Q.Clear) überschätzte der SUVmax die wahre Aktivitätskonzentration in den größeren Kugeleinsätzen erheblich (bis +120%). Jedoch war das Ausmaß der Aktivitätsüberschätzung deutlicher von der Anzahl der Iterationen/Subsets und der Halbwertsbreite (FWHM) des in-plane-Filters bzw. der Wahl des Regulationsfaktors β abhängig als von der Wahl des Rekonstruktionsverfahrens selbst. Auch die SNR wurde deutlicher durch o.g. Rekonstruktionsparameter beeinflusst als durch den zugrundeliegenden Algorithmus. Diese deutlichen intermethodischen Differenzen im SUVmax illustrieren zudem die Schwierigkeit, über Zentren hinweg einheitliche SUV-Schwellwerte zu identifizieren, z.B. für die Dignitätsbeurteilung von Lymphknoten (40, 117, 118). Der SUVmax ist naturgemäß ein Extremwert, als statistischer Ausreißer in der Regel besonders anfällig für diverse Störeffekte wie das Bildrauschen und deshalb nicht notwendigerweise repräsentativ für die durchschnittliche Aktivitätskonzentration in der Läsion. Die Recovery mittels SUVpeak zeigte sich in dieser Auswertung deutlich vergleichbarer zwischen den Rekonstruktionsverfahren und β -parametern und überschätzte die wahre Aktivitätskonzentration deutlich weniger als der SUVmax, d.h. der Bias war deutlich geringer. Dies ist im Einklang mit Ergebnissen der EARL-Gruppe, die vier aktuelle PET-Systeme mit herkömmlichen iterativen Rekonstruktionsverfahren bzw. Q.Clear hinsichtlich der Recovery verglichen. Vor und nach Harmonisierung war die Variation der Recovery des SUVpeak deutlich geringer als die Variabilität der Recovery basierend auf dem SUVmax (50). In der klinischen Routine ist der SUVpeak bislang jedoch nicht etabliert.

Die Q.Clear-Rekonstruktion bot bei einem β von 300 bzw. 450 eine überlegene Balance zwischen der Recovery der kleinen Kugeleinsätze und der SNR. Für diese β -Werte bietet Q.Clear somit eine überlegene Bildqualität gegenüber den konventionellen Rekonstruktionsverfahren (OSEM mit TOF bzw. PSF und TOF). Dies gilt jedoch zunächst nur für das eingesetzte PET-System (GE Discovery MI) mit der spezifischen 3-Ring-Detektorkonfiguration. Die vorteilhafte Bildqualität mit Q.Clear ergibt sich jedoch maßgeblich aus einer höheren SNR; die Recovery war vergleichbar, sodass keine deutliche Überlegenheit in der Detektion bzw. Beurteilung kleiner Läsionen zu erwarten ist. Howard *et al.* berichteten, dass die subjektive Bewertung der Auffälligkeit von kleinen pulmonalen Läsionen zwischen Q.Clear mit einem $\beta > 150$ und OSEM (ohne TOF und PSF) nicht unterschiedlich war (134). Ebenso gibt es Berichte über eine überlegene Erkennbarkeit von pulmonalen Läsionen mit Q.Clear, auch bei höheren β -Werten. Jedoch wurde jeweils nur eine spezifische Parameterkonstellation (Iterationen, Subsets, FWHM des in-plane-Filters) für die konventionellen Rekonstruktionsverfahren gewählt, sodass die Generalisierbarkeit der Ergebnisse ungeklärt bleibt (137, 138). Zudem ist die Detektion von pulmonalen Läsionen aufgrund des meist hohen Läsions-zu-Hintergrund-Verhältnisses ein diagnostischer Spezialfall und eine verbesserte Abgrenzbarkeit nicht gleichzusetzen mit verbesserter diagnostischer Genauigkeit. Somit bleibt es fraglich, ob Q.Clear auch einen relevanten

Beitrag zu einer Optimierung der diagnostischen Genauigkeit im spezifischen Kontext der Lymphknotenbeurteilung bei Patienten mit NSCLC bieten kann. Teoh *et al.* verglichen OSEM mit TOF (2 Iterationen, 24 Subsets, Filter mit 6,4 mm FWHM) und Q.Clear mit $\beta = 400$ in der mediastinalen Lymphknotenbeurteilung bei 47 Patienten mit NSCLC. Weder in der visuellen Befundung noch bei einem SUVmax-Schwellwert von 2,5 oder mit einem optimierten SUV-Schwellwert war Q.Clear dem konventionellen Algorithmus in der diagnostischen Genauigkeit überlegen. Die Sensitivität und Spezifität war für beide Rekonstruktionsverfahren ungewöhnlich niedrig. Zwar war die Sensitivität mit Q.Clear tendenziell höher (visuell: 76 vs. 68%), die Spezifität jedoch geringer als mit OSEM+TOF (visuell: 48 vs. 60%) (139). Eine weitere Steigerung der Recovery für kleine Läsionen könnte mit Q.Clear allenfalls mit einem geringeren β -Wert erreicht werden, doch auch ein β von 150 erreichte in der eigenen Auswertung keine relevante Verbesserung der Recovery. Hingegen führte dies bereits zu einer deutlich unterlegenen und klinisch inakzeptablen SNR. Dies spiegelt sich in einer unterlegenen subjektiven Bildqualität gegenüber höheren β -Werten wider (138, 140).

3.2 Quantifizierung, Risiko- und Prognoseabschätzung

Die beiden klinischen Arbeiten zur prognostischen Wertigkeit der asphericity (ASP) bei Patienten mit NSCLC zeigten deren prognostischen Stellenwert für das PFS und OS unabhängig vom klinischen Tumorstadium und weiteren klinischen und histologischen Variablen. In Übereinstimmung mit der Literatur waren zudem etablierte Faktoren wie das Geschlecht, das Alter, der ECOG-Status und das klinische Tumorstadium signifikant mit dem PFS und OS assoziiert (21, 141, 142). Auch für das MTV in ml konnte in der univariablen Cox-Regression ein Zusammenhang mit dem PFS und OS gezeigt werden; dies stimmt mit Publikationen anderer Arbeitsgruppen überein (124, 125). Die ASP war jedoch in beiden aktuellen Arbeiten dem MTV und dem SUVmax des Primärtumors hinsichtlich der prognostischen Wertigkeit überlegen. Da MTV und ASP eine signifikante Korrelation aufweisen (siehe Publikation unter 2.2.3), wurde angesichts der signifikanteren Assoziation mit dem PFS und OS primär nur die ASP in die multivariablen Analysen aufgenommen. Hatt *et al.* untersuchten bei 87 Patienten mit NSCLC im Stadium II bzw. III die prognostische Wertigkeit des MTV und der sphericity, eines gegensätzlich definierten Parameters zur ASP, aus der FDG-PET/CT. Die Autoren fanden eine überlegene prognostische Aussagekraft hinsichtlich des OS für das MTV gegenüber der sphericity, und die Kombination beider Faktoren bot die beste Differenzierung von Patienten mit günstigem bzw. ungünstigem OS. Zwar nutzten die Autoren denselben PET-Scanner, jedoch sind die Ergebnisse nur bedingt mit den eigenen Arbeiten vergleichbar, da die Approximation der MTV-Oberfläche bei Hatt *et al.* auf einem anderen Ansatz basierte (sog. *marching cubes* (143)) als in den eigenen Analysen (144). Zudem unterscheiden sich die Patientenkollektive hinsichtlich der klinischen Prognosefaktoren (v.a. klinisches Stadium und Therapiekonzepte).

In der zweiten eigenen Publikation mit 311 Patienten war der Zusammenhang zwischen ASP und PFS/OS in der klinisch relevanten Subgruppe von Patienten im Stadium II und Operation plus adjuvante Chemotherapie besonders deutlich. Insbesondere in diesem Stadium könnte die ASP somit einen Zusatznutzen in der Entscheidung über die Durchführung und Umfang einer Chemotherapie bieten. Eine unmittelbare Aussage über solche therapeutischen Implikationen erfordert jedoch den

Nachweis eines prädiktiven Wertes, d.h. des Potenzials, den Zusatznutzen einer bestimmten Therapie für das Outcome oder das Ansprechen auf diese spezifische Therapie vorherzusagen. Dies konnte nicht untersucht werden, da hierzu eine repräsentative Vergleichsgruppe (ohne Chemotherapie) erforderlich ist. Da die adjuvante Chemotherapie im Stadium (IB) II-III bereits Standard ist, muss davon ausgegangen werden, dass Patienten in diesen Stadien, die keine Chemotherapie durchlaufen, einer erheblichen Selektion unterliegen. Doch auch der Nachweis einer prognostischen Bedeutung, insbesondere in Bezug auf das PFS, könnte zukünftig Konsequenzen für Therapieentscheidungen haben. So konnte beispielsweise die ANITA-Studie im Stadium IIA keinen Überlebensvorteil der adjuvanten Chemotherapie gegenüber der alleinigen Operation nachweisen (12) (siehe Kapitel 1.1.3). Patienten im Stadium IIA mit geringer ASP und somit günstigem PFS könnten insofern ggf. Kandidaten für den Verzicht auf eine Chemotherapie sein. Patienten mit hoher ASP könnten hingegen von einer engmaschigeren bildgebenden Nachsorge für eine frühe Erkennung des Rezidivs profitieren. Bei Patienten im Stadium I und III war die ASP kein signifikanter Prädiktor des PFS bzw. OS. Ursächlich ist im Stadium I, d.h. bei häufig vergleichsweise kleinen Primärtumoren, mutmaßlich insbesondere die eingeschränkte rekonstruierte räumliche Auflösung der PET. Diese bewirkt, dass tatsächliche Irregularitäten in der Oberfläche des stoffwechselaktiven Tumorgewebes nur unzureichend in einer Irregularität des MTV abgebildet werden können. Somit ist die ASP als Maß für diese Irregularität bei solch kleinen Läsionen i.d.R. grundsätzlich niedrig und bietet kaum Differenzierungsmöglichkeit. Im Stadium I zeigten nur 6 von 140 Patienten eine ASP oberhalb des Schwellwertes von 33,3%. Im Stadium III sind häufig vergleichsweise große Primärtumoren anzutreffen, jedoch wird die Prognose maßgeblich vom Ausmaß der Lymphknotenmetastasierung bestimmt. Diese wird durch bildbasierte Biomarker des Primärtumors nicht abgebildet.

Der nächste Schritt sollte deshalb eine prospektive, nicht-interventionelle Studie zur abschließenden Validierung des prognostischen Wertes der ASP und für dedizierte Subgruppenanalysen sein. Zwar konnte mit den beiden Arbeiten bereits gezeigt werden, dass der prognostische Stellenwert der ASP prinzipiell auch unter Verwendung von zwei sehr unterschiedlichen PET-Systemen und Bildrekonstruktionen nachweisbar ist (rekonstruierte räumliche Auflösung von ca. 4 mm vs. 8 mm FWHM [(46), unpublizierte eigene Daten]). Dennoch müsste für ein multizentrisches Studiendesign eine optimale Vergleichbarkeit der ASP zwischen unterschiedlichen Zentren gewährleistet werden.

Deshalb schloss sich die dritte Publikation als methodische Vorbereitung einer solchen Studie an. Darin wurde die Variabilität der ASP, des MTV und SUVmax von 50 Lungenkarzinomen in Abhängigkeit von der verwendeten Bildrekonstruktion untersucht. Die rekonstruierte räumliche Auflösung diente als Charakteristikum für die Vergleichbarkeit der resultierenden Bilddaten. Die Variabilität war für die ASP generell höher als für den SUVmax und das MTV, da die ASP sowohl durch eine Variation des maximalen Uptakes in der Zielläsion (schwellwertbasierte MTV-Definition) als auch durch die Ortsauflösung an der Läsionsoberfläche beeinflusst wird. Sofern eine vereinheitlichte rekonstruierte räumliche Auflösung, z.B. 7 mm FWHM, gewährleistet werden kann, bleibt die Rate diskordant klassifizierter Patienten (hohe vs. niedrige ASP) jedoch gering und auf tolerablem Niveau (<5% für konventionelle Rekonstruktionsverfahren). Die rekonstruierte räumliche Auflösung für jeden Scanner kann unkompliziert anhand einer Messung mit einem Phantom mit standardisierten füllbaren

Kugeleinsätzen vor Studienbeginn bestimmt werden (58), und die Rekonstruktionsparameter (Iterationen, Subsets, in-plane-Filter) können dann für die gewünschte Auflösung optimiert werden. Q.Clear sollte aufgrund der systematisch höheren SNR und der höheren Abweichungen zu den konventionellen Verfahren (10% diskordante Fälle trotz vergleichbarer FWHM) nicht für multizentrische Auswertungen eingesetzt werden. Sollte eine Auswertung von Bilddaten unterschiedlicher rekonstruierter räumlicher Auflösungen erforderlich werden (z.B. aus retrospektiven Daten), könnten diese angeglichen werden, indem eine Nachglättung der Bilddaten mit besserer rekonstruierter räumlicher Auflösung erfolgt. Die resultierenden Diskordanzen waren in der aktuellen Auswertung so gering wie bei einer primär vergleichbaren rekonstruierten räumlichen Auflösung. Die bislang umfangreichsten Initiativen zur Harmonisierung und Standardisierung von PET-Untersuchungen, z.B. das EARL-Programm (145), nutzen den SUV-Wert (i.d.R. SUVmax) bzw. die Recovery als Surrogat für die Vergleichbarkeit von PET-Systemen, da in der klinischen PET-Bildgebung bislang nur die SUV-Werte als (semi-)quantitative Parameter etabliert sind. Diese Arbeitsgruppen konnten ebenfalls zeigen, dass das Nachglätten von Daten mit zu hoher Recovery mithilfe eines Gauß-Filters von adaptiver FWHM (bis zu 8 mm) die Möglichkeit bietet, eine weitgehende Vergleichbarkeit zwischen Recovery-Werten verschiedener PET-Systeme zu gewährleisten (146-148).

Ungeklärt bleibt der Einfluss der Voxelgröße auf die ASP und ob eine nachträgliche Konversion der Bilddaten in eine standardisierte Matrixgröße eine Harmonisierung der Bilddaten erlaubt. Deshalb sollte für eine multizentrische Studie eine ähnliche Voxelgröße angestrebt werden. Yan *et al.* zeigten, dass neben der Filterbreite die Variation der Voxelgröße den größten Einfluss auf die Variabilität verschiedener second-order- und high-order-Parameter hat (127). Die Voxel der Bilddaten in der eigenen Arbeit maßen $2,73 \times 2,73 \times 2,78 \text{ mm}^3$. Insbesondere bei älteren PET-Systemen sind größere Voxel üblich, z.B. mit einer Kantenlänge von ca. 4 mm, wie bei den beiden Scannern, die für die beiden klinischen Publikationen zur prognostischen Wertigkeit der ASP zum Einsatz kamen. Eine flexible Ausgestaltung der Voxelgröße bei der Rekonstruktion war für die methodische Publikation jedoch softwaretechnisch nicht möglich.

Eine retrospektive Reduktion der Akquisitionszeit mithilfe der PET-Rohdaten war ohne relevanten Einfluss auf die ASP, solange die Akquisitionszeit von 180s pro Bettposition nicht auf <90s reduziert wurde. Dies könnte bedeuten, dass Variationen in der injizierten Aktivität pro kg oder der Akquisitionszeit zwischen den Zentren nicht von entscheidender Bedeutung für die Vergleichbarkeit der ASP sind. Dennoch sollte im Rahmen einer prospektiven Studienplanung immer auch ein standardisiertes Untersuchungsprotokoll angestrebt werden, nicht zuletzt auch mit Rücksicht auf andere SUV- und Texturparameter (51, 77, 78).

In der zweiten klinischen Arbeit zur prognostischen Aussagekraft der ASP bei Patienten mit NSCLC (Punkt 2.2.2) wurde außerdem die Interobserver-Variabilität von MTV und ASP zwischen drei Befundern in einer repräsentativen Subgruppe von 63 der 311 Patienten untersucht (*Supplementary Material*). Die Übereinstimmung in MTV und ASP zwischen den gegeneinander verblindeten Befundern war nahezu perfekt (Pearson-Korrelationskoeffizient $\geq 0,94$). Eine Vergleichbarkeit insbesondere von Form- und Größenparametern („shape and size“) kann jedoch nur gegeben sein,

wenn die Abgrenzungsmethodik vergleichbar ist. Deshalb wurde von allen Befundern derselbe halbautomatische Abgrenzungsalgorithmus verwendet wie in der Hauptauswertung (relativer, hintergrundadaptierter Schwellwert bezogen auf den SUVmax der Läsion). Entsprechend sollte auch derselbe Abgrenzungsalgorithmus für zukünftige Auswertungen in diesem Kontext verwendet und in der Planung von multizentrischen Studien als wesentlicher methodischer Parameter berücksichtigt werden.

4. Zusammenfassung

4.1 Prätherapeutische Lymphknotendiagnostik

Es erfolgten zwei retrospektive Studien mit 47 bzw. 86 Patienten mit NSCLC mit dem Ziel einer verbesserten und reproduzierbaren thorakalen Lymphknotendiagnostik mittels FDG-PET/CT. Ein zusätzlicher später Aufnahmezeitpunkt (Dual Time Point Imaging; DTPI) war ohne Zusatznutzen. SUV-Schwellwerte erreichten eine hohe diagnostische Genauigkeit, sind jedoch durch Variabilität zwischen verschiedenen PET-Systemen und Rekonstruktionsverfahren limitiert. Eine standardisierte visuelle Lymphknotenbeurteilung mit einer vereinheitlichten Bildfensterung und einem 4-stufigen Score wurde vorgeschlagen, mit der in beiden Studien eine hohe diagnostische Genauigkeit (>90%) erreicht wurde. Die Genauigkeit war unabhängig von der Erfahrung des Befunders, von der Lymphknotenlokalisation und Primärtumorhistologie.

In Phantommessungen wurde eine neuartige PET-Bildrekonstruktion („Q.Clear“) hinsichtlich verschiedener Bildqualitätsparameter mit konventionellen iterativen Verfahren verglichen. Die Bildcharakteristik mit Q.Clear hängt entscheidend vom Regulationsfaktor β ab. Mit einem niedrigen β -Wert kann eine überlegene rekonstruierte räumliche Auflösung erreicht werden, jedoch die Recovery fokaler Aktivitäten nicht signifikant gegenüber konventionellen Algorithmen verbessert werden. Zudem ist eine abnehmende SNR zu beachten. Somit ist anzunehmen, dass Q.Clear keinen systematischen Vorteil in der diagnostischen Genauigkeit der Lymphknotendiagnostik beim NSCLC bietet.

4.2 Quantifizierung, Risiko- und Prognoseabschätzung

Bei Patienten mit NSCLC und potenziell kurativer Therapie basiert die Therapieentscheidung aktuell maßgeblich auf einer einzelnen Variable, dem klinischen Tumorstadium. Das Ziel dieser Projekte war deshalb, Bildparameter aus der prätherapeutischen FDG-PET mit prognostischer Wertigkeit zu identifizieren, um eine individualisiertere Risikostratifizierung der Patienten zu ermöglichen. In zwei retrospektiven Studien mit 60 bzw. 311 Patienten mit NSCLC konnte ein Formparameter aus der FDG-PET, die asphericity (ASP), jeweils einen unabhängigen prognostischen Stellenwert für das PFS und OS zeigen, auch unter Verwendung zwei unterschiedlicher PET-Systeme. Diese prognostische Bedeutung war bei Patienten im Stadium II mit Operation und adjuvanter Chemotherapie besonders deutlich und könnte helfen, Patienten zu identifizieren, die aufgrund ihrer günstigen Prognose (geringe ASP) von einer Therapiede Eskalation profitieren könnten oder, bei hohem Rezidivrisiko (hohe ASP), von einer engmaschigeren Nachsorge. Eine Interobserver-Analyse zeigte eine nahezu perfekte Übereinstimmung von 3 Befundern in der halbautomatischen Abgrenzung des metabolischen Tumolvolumens (MTV) und der daraus resultierenden ASP.

In einer separaten methodischen Arbeit wurde die Variabilität der ASP unter verschiedenen Bildrekonstruktionsverfahren, Post-Processing-Filtern und Akquisitionszeiten bei 50 Patienten mit NSCLC evaluiert. Wird eine harmonisierte rekonstruierte räumliche Auflösung eingehalten (gemessen im Phantom), kann der Anteil diskordant klassifizierter Fälle (hohe vs. niedrige ASP) geringgehalten werden (<5%). Damit können therapeutische Implikationen der ASP zukünftig auch in multizentrischen Auswertungen evaluiert und validiert werden.

5. Literaturangaben

1. Robert-Koch-Institut, Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. Krebs in Deutschland 2015/2016. 2019.
2. Kraywinkel K, Barnes B. Epidemiologie des kleinzelligen Lungenkarzinoms in Deutschland. *Der Onkologe*. 2017;23(5):334-9.
3. Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger KR, Yatabe Y, et al. International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society International Multidisciplinary Classification of Lung Adenocarcinoma. *Journal of Thoracic Oncology*. 2011;6(2):244-85.
4. Petersen I. Morphologische und molekulare Diagnostik des Lungenkarzinoms. *Dtsch Arztebl International*. 2011;108(31-32):525-31.
5. Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WE, et al. The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer*. 2016;11(1):39-51.
6. Wang YX, Li BS, Huang W, Yi Y, Zhang J, Wang Zh T, et al. Pattern of lymph node metastases and its implication in radiotherapeutic clinical target volume in patients with non-small-cell lung cancer: a study of 2062 cases. *The British journal of radiology*. 2015;88(1056):20140288.
7. Lin G, Wang Z, Sun X, Liu J, Xu Y. Lymph node drainage patterns in resectable non-small cell lung cancer and its clinical significance. *Journal of Clinical Oncology*. 2016;34(15_suppl):e20065-e.
8. Morgensztern D, Waqar S, Subramanian J, Gao F, Govindan R. Improving survival for stage IV non-small cell lung cancer: a surveillance, epidemiology, and end results survey from 1990 to 2005. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer*. 2009;4(12):1524-9.
9. Lungs PNG. URL: <https://pngimg.com/download/74873>. Lizenz: Creative Commons 4.0 BY-NC. Letzter Zugriff am 07.12.2020.
10. Postmus PE, Kerr KM, Oudkerk M, Senan S, Waller DA, Vansteenkiste J, et al. Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of oncology : official journal of the European Society for Medical Oncology*. 2017;28(suppl_4):iv1-iv21.
11. Winton T, Livingston R, Johnson D, Rigas J, Johnston M, Butts C, et al. Vinorelbine plus cisplatin vs. observation in resected non-small-cell lung cancer. *The New England journal of medicine*. 2005;352(25):2589-97.
12. Douillard JY, Rosell R, De Lena M, Carpagnano F, Ramlau R, González-Larriba JL, et al. Adjuvant vinorelbine plus cisplatin versus observation in patients with completely resected stage IB-IIIa non-small-cell lung cancer (Adjuvant Navelbine International Trialist Association [ANITA]): a randomised controlled trial. *The Lancet Oncology*. 2006;7(9):719-27.
13. Pignon JP, Tribodet H, Scagliotti GV, Douillard JY, Shepherd FA, Stephens RJ, et al. Lung adjuvant cisplatin evaluation: a pooled analysis by the LACE Collaborative Group. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2008;26(21):3552-9.
14. AWMF-Registernummer: 020/007OL. Leitlinienprogramm Onkologie (Deutsche Krebsgesellschaft, Deutsche Krebshilfe, AWMF): Prävention, Diagnostik, Therapie und Nachsorge des Lungenkarzinoms, Langversion 1.0, 2018. Zugriff am: 04.07.2020. URL: <http://leitlinienprogramm-onkologie.de/Lungenkarzinom.98.0.html>.
15. Ewig S, Behringer DM. 5.3 Therapiestandards für Patienten mit Lungenkrebs. 2015. In: *Lungenkrebs: Individuelle Therapie im zertifizierten Thoraxzentrum* [Internet]. Stuttgart: Georg Thieme Verlag. 2015. URL: <http://www.thieme-connect.de/products/ebooks/lookinside/10.1055/b-0034-101524>.
16. Tsao MS, Aviell-Ronen S, Ding K, Lau D, Liu N, Sakurada A, et al. Prognostic and predictive importance of p53 and RAS for adjuvant chemotherapy in non small-cell lung cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2007;25(33):5240-7.
17. Ma X, Le Teuff G, Lacas B, Tsao MS, Graziano S, Pignon JP, et al. Prognostic and Predictive Effect of TP53 Mutations in Patients with Non-Small Cell Lung Cancer from Adjuvant Cisplatin-Based Therapy Randomized Trials: A LACE-Bio Pooled Analysis. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer*. 2016;11(6):850-61.
18. Zhang Y, Han CY, Duan FG, Fan XX, Yao XJ, Parks RJ, et al. p53 sensitizes chemoresistant non-small cell lung cancer via elevation of reactive oxygen species and suppression of EGFR/PI3K/AKT signaling. *Cancer cell international*. 2019;19:188.
19. Kandioler D, Stamatis G, Eberhardt W, Kappel S, Zöchbauer-Müller S, Kühner I, et al. Growing clinical evidence for the interaction of the p53 genotype and response to induction chemotherapy in advanced non-small cell lung cancer. *The Journal of thoracic and cardiovascular surgery*. 2008;135(5):1036-41.
20. Tung MC, Lin PL, Wang YC, He TY, Lee MC, Yeh SD, et al. Mutant p53 confers chemoresistance in non-small cell lung cancer by upregulating Nrf2. *Oncotarget*. 2015;6(39):41692-705.
21. Strand TE, Rostad H, Møller B, Norstein J. Survival after resection for primary lung cancer: a population based study of 3211 resected patients. *Thorax*. 2006;61(8):710-5.

22. Brundage MD, Davies D, Mackillop WJ. Prognostic factors in non-small cell lung cancer: a decade of progress. *Chest*. 2002;122(3):1037-57.
23. Russell PA, Wainer Z, Wright GM, Daniels M, Conron M, Williams RA. Does lung adenocarcinoma subtype predict patient survival?: A clinicopathologic study based on the new International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society international multidisciplinary lung adenocarcinoma classification. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer*. 2011;6(9):1496-504.
24. Tsao MS, Marguet S, Le Teuff G, Lantuejoul S, Shepherd FA, Seymour L, et al. Subtype Classification of Lung Adenocarcinoma Predicts Benefit From Adjuvant Chemotherapy in Patients Undergoing Complete Resection. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2015;33(30):3439-46.
25. Schicha H, Schober O. *Nuklearmedizin: Basiswissen und klinische Anwendung*: Schattauer; 2013.
26. Brown RS, Wahl RL. Overexpression of Glut-1 glucose transporter in human breast cancer. An immunohistochemical study. *Cancer*. 1993;72(10):2979-85.
27. Paul R, Johansson R, Kellokumpu-Lehtinen PL, Söderström KO, Kangas L. Tumor localization with 18F-2-fluoro-2-deoxy-D-glucose: comparative autoradiography, glucose 6-phosphatase histochemistry, and histology of renally implanted sarcoma of the rat. *Research in experimental medicine Zeitschrift fur die gesamte experimentelle Medizin einschliesslich experimenteller Chirurgie*. 1985;185(2):87-94.
28. Park SG, Lee JH, Lee WA, Han KM. Biologic correlation between glucose transporters, hexokinase-II, Ki-67 and FDG uptake in malignant melanoma. *Nuclear medicine and biology*. 2012;39(8):1167-72.
29. Higashi K, Ueda Y, Sakurai A, Mingwang X, Xu L, Murakami M, et al. Correlation of Glut-1 glucose transporter expression with [(18)F]FDG uptake in non-small cell lung cancer. *European journal of nuclear medicine*. 2000;27(12):1778-85.
30. Kubota R, Yamada S, Kubota K, Ishiwata K, Tamahashi N, Ido T. Intratumoral distribution of fluorine-18-fluorodeoxyglucose in vivo: high accumulation in macrophages and granulation tissues studied by microautoradiography. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*. 1992;33(11):1972-80.
31. Yamada S, Kubota K, Kubota R, Ido T, Tamahashi N. High accumulation of fluorine-18-fluorodeoxyglucose in turpentine-induced inflammatory tissue. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*. 1995;36(7):1301-6.
32. Wood IS, Trayhurn P. Glucose transporters (GLUT and SGLT): expanded families of sugar transport proteins. *The British journal of nutrition*. 2003;89(1):3-9.
33. Deuschl C, Nensa F, Grueneisen J, Poeppel TD, Sawicki LM, Heusch P, et al. Diagnostic impact of integrated 18F-FDG PET/MRI in cerebral staging of patients with non-small cell lung cancer. *Acta radiologica (Stockholm, Sweden : 1987)*. 2017;58(8):991-6.
34. Tsili AC, Alexiou G, Naka C, Argyropoulou MI. Imaging of colorectal cancer liver metastases using contrast-enhanced US, multidetector CT, MRI, and FDG PET/CT: a meta-analysis. *Acta radiologica (Stockholm, Sweden : 1987)*. 2020;284185120925481.
35. Chiu CH, Yeh YC, Lin KH, Wu YC, Lee YC, Chou TY, et al. Histological subtypes of lung adenocarcinoma have differential ¹⁸F-fluorodeoxyglucose uptakes on the positron emission tomography/computed tomography scan. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer*. 2011;6(10):1697-703.
36. Vesselle H, Salskov A, Turcotte E, Wiens L, Schmidt R, Jordan CD, et al. Relationship between non-small cell lung cancer FDG uptake at PET, tumor histology, and Ki-67 proliferation index. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer*. 2008;3(9):971-8.
37. Duan XY, Wang W, Li M, Li Y, Guo YM. Predictive significance of standardized uptake value parameters of FDG-PET in patients with non-small cell lung carcinoma. *Brazilian journal of medical and biological research = Revista brasileira de pesquisas medicas e biologicas*. 2015;48(3):267-72.
38. Suárez-Piñera M, Belda-Sanchis J, Taus A, Sánchez-Font A, Mestre-Fusco A, Jiménez M, et al. FDG PET-CT SUVmax and IASLC/ATS/ERS histologic classification: a new profile of lung adenocarcinoma with prognostic value. *American journal of nuclear medicine and molecular imaging*. 2018;8(2):100-9.
39. Higashi K, Ueda Y, Seki H, Yuasa K, Oguchi M, Noguchi T, et al. Fluorine-18-FDG PET imaging is negative in bronchioloalveolar lung carcinoma. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*. 1998;39(6):1016-20.
40. Schmidt-Hansen M, Baldwin DR, Hasler E, Zamora J, Abairra V, Roqué IFM. PET-CT for assessing mediastinal lymph node involvement in patients with suspected resectable non-small cell lung cancer. *The Cochrane database of systematic reviews*. 2014;2014(11):Cd009519.
41. Townsend DW. Physical principles and technology of clinical PET imaging. *Annals of the Academy of Medicine, Singapore*. 2004;33(2):133-45.
42. Conti M, Eriksson L. Physics of pure and non-pure positron emitters for PET: a review and a discussion. *EJNMMI physics*. 2016;3(1):8.

43. Moses WW. Fundamental Limits of Spatial Resolution in PET. Nuclear instruments & methods in physics research Section A, Accelerators, spectrometers, detectors and associated equipment. 2011;648 Supplement 1:S236-s40.
44. Stickel JR, Cherry SR. High-resolution PET detector design: modelling components of intrinsic spatial resolution. Physics in medicine and biology. 2005;50(2):179-95.
45. Panin VY, Kehren F, Michel C, Casey M. Fully 3-D PET reconstruction with system matrix derived from point source measurements. IEEE transactions on medical imaging. 2006;25(7):907-21.
46. Rogasch JMM, Hofheinz F, Lougovski A, Furth C, Ruf J, Großler OS, et al. The influence of different signal-to-background ratios on spatial resolution and F18-FDG-PET quantification using point spread function and time-of-flight reconstruction. EJNMMI physics. 2014;1(1):12.
47. Kadrmas DJ, Casey ME, Conti M, Jakoby BW, Lois C, Townsend DW. Impact of time-of-flight on PET tumor detection. Journal of nuclear medicine : official publication, Society of Nuclear Medicine. 2009;50(8):1315-23.
48. Lois C, Jakoby BW, Long MJ, Hubner KF, Barker DW, Casey ME, et al. An assessment of the impact of incorporating time-of-flight information into clinical PET/CT imaging. Journal of nuclear medicine : official publication, Society of Nuclear Medicine. 2010;51(2):237-45.
49. Akamatsu G, Ishikawa K, Mitsumoto K, Taniguchi T, Ohya N, Baba S, et al. Improvement in PET/CT image quality with a combination of point-spread function and time-of-flight in relation to reconstruction parameters. Journal of nuclear medicine : official publication, Society of Nuclear Medicine. 2012;53(11):1716-22.
50. Kaalep A, Sera T, Rijnsdorp S, Yaqub M, Talsma A, Lodge MA, et al. Feasibility of state of the art PET/CT systems performance harmonisation. European journal of nuclear medicine and molecular imaging. 2018;45(8):1344-61.
51. Boellaard R, Delgado-Bolton R, Oyen WJ, Giammarile F, Tatsch K, Eschner W, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. European journal of nuclear medicine and molecular imaging. 2015;42(2):328-54.
52. Boellaard R, O'Doherty MJ, Weber WA, Mottaghy FM, Lonsdale MN, Stroobants SG, et al. FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0. European journal of nuclear medicine and molecular imaging. 2010;37(1):181-200.
53. Sollini M, Cozzi L, Antunovic L, Chiti A, Kirienko M. PET Radiomics in NSCLC: state of the art and a proposal for harmonization of methodology. Scientific reports. 2017;7(1):358.
54. Nestle U, Kremp S, Schaefer-Schuler A, Sebastian-Welsch C, Hellwig D, Rube C, et al. Comparison of different methods for delineation of 18F-FDG PET-positive tissue for target volume definition in radiotherapy of patients with non-Small cell lung cancer. Journal of nuclear medicine : official publication, Society of Nuclear Medicine. 2005;46(8):1342-8.
55. Kolinger GD, Vázquez García D, Kramer GM, Frings V, Smit EF, de Langen AJ, et al. Repeatability of [(18)F]FDG PET/CT total metabolic active tumour volume and total tumour burden in NSCLC patients. EJNMMI research. 2019;9(1):14.
56. Erdi YE, Mawlawi O, Larson SM, Imbriaco M, Yeung H, Finn R, et al. Segmentation of lung lesion volume by adaptive positron emission tomography image thresholding. Cancer. 1997;80(12 Suppl):2505-9.
57. Hatt M, Cheze le Rest C, Turzo A, Roux C, Visvikis D. A fuzzy locally adaptive Bayesian segmentation approach for volume determination in PET. IEEE transactions on medical imaging. 2009;28(6):881-93.
58. Hofheinz F, Dittrich S, Pötzsch C, Hoff J. Effects of cold sphere walls in PET phantom measurements on the volume reproducing threshold. Physics in medicine and biology. 2010;55(4):1099-113.
59. Jentzen W, Freudenberg L, Eising EG, Heinze M, Brandau W, Bockisch A. Segmentation of PET volumes by iterative image thresholding. Journal of nuclear medicine : official publication, Society of Nuclear Medicine. 2007;48(1):108-14.
60. Kerhet A, Small C, Quon H, Riauka T, Greiner R, McEwan A, et al., editors. Segmentation of Lung Tumours in Positron Emission Tomography Scans: A Machine Learning Approach 2009; Berlin, Heidelberg: Springer Berlin Heidelberg.
61. Sharif MS, Abbod M, Amira A, Zaidi H. Artificial Neural Network-Based System for PET Volume Segmentation. International journal of biomedical imaging. 2010;2010.
62. Day E, Betler J, Parda D, Reitz B, Kirichenko A, Mohammadi S, et al. A region growing method for tumor volume segmentation on PET images for rectal and anal cancer patients. Medical physics. 2009;36(10):4349-58.
63. Bağcı U, Yao J, Caban J, Turkbey E, Aras O, Mollura DJ. A graph-theoretic approach for segmentation of PET images. Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual Conference. 2011;2011:8479-82.
64. Geets X, Lee JA, Bol A, Lonneux M, Grégoire V. A gradient-based method for segmenting FDG-PET images: methodology and validation. European journal of nuclear medicine and molecular imaging. 2007;34(9):1427-38.
65. Foster B, Bagcı U, Mansoor A, Xu Z, Mollura DJ. A review on segmentation of positron emission tomography images. Computers in biology and medicine. 2014;50:76-96.

66. Apostolova I, Steffen IG, Wedel F, Lougovski A, Marnitz S, Derlin T, et al. Asphericity of pretherapeutic tumour FDG uptake provides independent prognostic value in head-and-neck cancer. *European radiology*. 2014;24(9):2077-87.
67. El Naqa I, Grigsby P, Apte A, Kidd E, Donnelly E, Khullar D, et al. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *Pattern recognition*. 2009;42(6):1162-71.
68. Hofheinz F, Lougovski A, Zöphel K, Hentschel M, Steffen IG, Apostolova I, et al. Increased evidence for the prognostic value of primary tumor asphericity in pretherapeutic FDG PET for risk stratification in patients with head and neck cancer. *European journal of nuclear medicine and molecular imaging*. 2015;42(3):429-37.
69. Ahn HK, Lee H, Kim SG, Hyun SH. Pre-treatment (18)F-FDG PET-based radiomics predict survival in resected non-small cell lung cancer. *Clinical radiology*. 2019;74(6):467-73.
70. Bianconi F, Palumbo I, Spanu A, Nuvoli S, Fravolini m, Palumbo B. PET/CT radiomics in lung cancer: An overview. *Applied Sciences*. 2020;10:1718.
71. Kuhnert G, Boellaard R, Sterzer S, Kahraman D, Scheffler M, Wolf J, et al. Impact of PET/CT image reconstruction methods and liver uptake normalization strategies on quantitative image analysis. *European journal of nuclear medicine and molecular imaging*. 2016;43(2):249-58.
72. Amakusa S, Matsuoka K, Kawano M, Hasegawa K, Ouchida M, Date A, et al. Influence of region-of-interest determination on measurement of signal-to-noise ratio in liver on PET images. *Annals of nuclear medicine*. 2018;32(1):1-6.
73. Kanstrup IL, Klausen TL, Bojsen-Møller J, Magnusson P, Zerahn B. Variability and reproducibility of hepatic FDG uptake measured as SUV as well as tissue-to-blood background ratio using positron emission tomography in healthy humans. *Clinical physiology and functional imaging*. 2009;29(2):108-13.
74. Schwartz J, Humm JL, Gonen M, Kalaigian H, Schoder H, Larson SM, et al. Repeatability of SUV measurements in serial PET. *Medical physics*. 2011;38(5):2629-38.
75. Boktor RR, Walker G, Stacey R, Gledhill S, Pitman AG. Reference range for inpatient variability in blood-pool and liver SUV for 18F-FDG PET. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*. 2013;54(5):677-82.
76. Su M, Wang R, Miller J. Intra-patient SUV repeatability of background on 18F-FDG PET/CT. *Journal of Nuclear Medicine*. 2018;59(supplement 1):662.
77. Hristova I, Boellaard R, Vogel W, Mottaghy F, Marreaud S, Collette S, et al. Retrospective quality control review of FDG scans in the imaging sub-study of PALETTE EORTC 62072/VEG110727: a randomized, double-blind, placebo-controlled phase III trial. *European journal of nuclear medicine and molecular imaging*. 2015;42(6):848-57.
78. Tahari AK, Paidpally V, Chirindel A, Wahl RL, Subramaniam RM. Two-time-point FDG PET/CT: liver SULmean repeatability. *AJR American journal of roentgenology*. 2015;204(2):402-7.
79. de Langen AJ, Vincent A, Velasquez LM, van Tinteren H, Boellaard R, Shankar LK, et al. Repeatability of 18F-FDG uptake measurements in tumors: a metaanalysis. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*. 2012;53(5):701-8.
80. Akamatsu G, Mitsumoto K, Taniguchi T, Tsutsui Y, Baba S, Sasaki M. Influences of point-spread function and time-of-flight reconstructions on standardized uptake value of lymph node metastases in FDG-PET. *European journal of radiology*. 2014;83(1):226-30.
81. Armstrong IS, Kelly MD, Williams HA, Matthews JC. Impact of point spread function modelling and time of flight on FDG uptake measurements in lung lesions using alternative filtering strategies. *EJNMMI physics*. 2014;1(1):99.
82. Rogasch JM, Steffen IG, Hofheinz F, Großer OS, Furth C, Mohnike K, et al. The association of tumor-to-background ratios and SUVmax deviations related to point spread function and time-of-flight F18-FDG-PET/CT reconstruction in colorectal liver metastases. *EJNMMI research*. 2015;5:31.
83. Houdu B, Lasnon C, Licaj I, Thomas G, Do P, Guizard AV, et al. Why harmonization is needed when using FDG PET/CT as a prognosticator: demonstration with EARL-compliant SUV as an independent prognostic factor in lung cancer. *European journal of nuclear medicine and molecular imaging*. 2019;46(2):421-8.
84. Lasnon C, Desmots C, Quak E, Gervais R, Do P, Dubos-Arvis C, et al. Harmonizing SUVs in multicentre trials when using different generation PET systems: prospective validation in non-small cell lung cancer patients. *European journal of nuclear medicine and molecular imaging*. 2013;40(7):985-96.
85. Aide N, Lasnon C, Veit-Haibach P, Sera T, Sattler B, Boellaard R. EANM/EARL harmonization strategies in PET quantification: from daily practice to multicentre oncological studies. *European journal of nuclear medicine and molecular imaging*. 2017;44(Suppl 1):17-31.
86. Lasnon C, Enilorac B, Popotte H, Aide N. Impact of the EARL harmonization program on automatic delineation of metabolic active tumour volumes (MATVs). *EJNMMI research*. 2017;7(1):30.
87. Zhuang M, García DV, Kramer GM, Frings V, Smit EF, Dierckx R, et al. Variability and Repeatability of Quantitative Uptake Metrics in (18)F-FDG PET/CT of Non-Small Cell Lung Cancer: Impact of Segmentation Method, Uptake Interval, and Reconstruction Protocol. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*. 2019;60(5):600-7.

88. Guezennec C, Bourhis D, Orlhac F, Robin P, Corre JB, Delcroix O, et al. Inter-observer and segmentation method variability of textural analysis in pre-therapeutic FDG PET/CT in head and neck cancer. *PloS one*. 2019;14(3):e0214299.
89. Hatt M, Cheze-le Rest C, van Baardwijk A, Lambin P, Pradier O, Visvikis D. Impact of tumor size and tracer uptake heterogeneity in (18)F-FDG PET and CT non-small cell lung cancer tumor delineation. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*. 2011;52(11):1690-7.
90. Hofheinz F, Pöttsch C, Oehme L, Beuthien-Baumann B, Steinbach J, Kotzerke J, et al. Automatic volume delineation in oncological PET. Evaluation of a dedicated software tool and comparison with manual delineation in clinical data sets. *Nuklearmedizin Nuclear medicine*. 2012;51(1):9-16.
91. Pfaehler E, Burggraaff C, Kramer G, Zijlstra J, Hoekstra OS, Jalving M, et al. PET segmentation of bulky tumors: Strategies and workflows to improve inter-observer variability. *PloS one*. 2020;15(3):e0230901.
92. Hatt M, Cheze le Rest C, Descourt P, Dekker A, De Ruyscher D, Oellers M, et al. Accurate automatic delineation of heterogeneous functional volumes in positron emission tomography for oncology applications. *International journal of radiation oncology, biology, physics*. 2010;77(1):301-8.
93. Driessen J, Zwezerijnen GJC, Eertink JJ, Kersten MJ, Hagenbeek A, Hoekstra OS, et al. Baseline Metabolic Tumor Volume in 18FDG-PET-CT Scans in Classical Hodgkin Lymphoma Using Semi-Automatic Segmentation. *Blood*. 2019;134(Supplement_1):4049-.
94. van Tinteren H, Hoekstra OS, Smit EF, van den Bergh JH, Schreurs AJ, Stallaert RA, et al. Effectiveness of positron emission tomography in the preoperative assessment of patients with suspected non-small-cell lung cancer: the PLUS multicentre randomised trial. *Lancet (London, England)*. 2002;359(9315):1388-93.
95. Gunluoglu MZ, Melek H, Medetoglu B, Demir A, Kara HV, Dincer SI. The validity of preoperative lymph node staging guidelines of European Society of Thoracic Surgeons in non-small-cell lung cancer patients. *European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery*. 2011;40(2):287-90.
96. Simo M, Serra M, Bastus R, Cirera L, Garcia JR, Soler M, et al. False-negative and false-positive lymph node staging with FDG-PET scan in patients with proven or suspected non-small-cell lung cancer (NSCLC). *Journal of Nuclear Medicine*. 2006;47(suppl 1):470P.
97. Nomori H, Watanabe K, Ohtsuka T, Naruke T, Suemasu K, Uno K. The size of metastatic foci and lymph nodes yielding false-negative and false-positive lymph node staging with positron emission tomography in patients with lung cancer. *The Journal of thoracic and cardiovascular surgery*. 2004;127(4):1087-92.
98. Takamochi K, Yoshida J, Murakami K, Niho S, Ishii G, Nishimura M, et al. Pitfalls in lymph node staging with positron emission tomography in non-small cell lung cancer patients. *Lung cancer (Amsterdam, Netherlands)*. 2005;47(2):235-42.
99. Kaseda K, Watanabe K, Asakura K, Kazama A, Ozawa Y. Identification of false-negative and false-positive diagnoses of lymph node metastases in non-small cell lung cancer patients staged by integrated (18F-)fluorodeoxyglucose-positron emission tomography/computed tomography: A retrospective cohort study. *Thoracic cancer*. 2016;7(4):473-80.
100. Li S, Zheng Q, Ma Y, Wang Y, Feng Y, Zhao B, et al. Implications of false negative and false positive diagnosis in lymph node staging of NSCLC by means of ¹⁸F-FDG PET/CT. *PloS one*. 2013;8(10):e78552.
101. Konishi J, Yamazaki K, Tsukamoto E, Tamaki N, Onodera Y, Otake T, et al. Mediastinal lymph node staging by FDG-PET in patients with non-small cell lung cancer: analysis of false-positive FDG-PET findings. *Respiration; international review of thoracic diseases*. 2003;70(5):500-6.
102. Belda-Sanchis J, Serra-Mitjans M, Rami-Porta R, Heredia JL, Bastus R, Gonzalez S, et al. P1-057: False-positive and false-negative results of [18F] 2-fluoro-2-deoxyglucose positron emission tomography (FDG-PET) in non-small cell lung cancer (NSCLC). Lessons learnt from clinical practice. *Journal of Thoracic Oncology*. 2007;2(8):S572-S3.
103. Roberts PF, Follette DM, von Haag D, Park JA, Valk PE, Pounds TR, et al. Factors associated with false-positive staging of lung cancer by positron emission tomography. *The Annals of thoracic surgery*. 2000;70(4):1154-9; discussion 9-60.
104. Werutsky G, Hochegger B, Lopes de Figueiredo Pinto JA, Martinez-Mesa J, Zanini ML, Berdichevski EH, et al. PET-CT has low specificity for mediastinal staging of non-small-cell lung cancer in an endemic area for tuberculosis: a diagnostic test study (LACOG 0114). *BMC cancer*. 2019;19(1):5.
105. Zhuang H, Pourdehnad M, Lambright ES, Yamamoto AJ, Lanuti M, Li P, et al. Dual time point 18F-FDG PET imaging for differentiating malignant from inflammatory processes. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*. 2001;42(9):1412-7.
106. Hu M, Han A, Xing L, Yang W, Fu Z, Huang C, et al. Value of dual-time-point FDG PET/CT for mediastinal nodal staging in non-small-cell lung cancer patients with lung comorbidity. *Clinical nuclear medicine*. 2011;36(6):429-33.
107. Shen G, Deng H, Hu S, Jia Z. Potential performance of dual-time-point 18F-FDG PET/CT compared with single-time-point imaging for differential diagnosis of metastatic lymph nodes: a meta-analysis. *Nuclear medicine communications*. 2014;35(10):1003-10.

108. Lasnon C, Hicks RJ, Beaugerard JM, Milner A, Paciencia M, Guizard AV, et al. Impact of point spread function reconstruction on thoracic lymph node staging with 18F-FDG PET/CT in non-small cell lung cancer. *Clinical nuclear medicine*. 2012;37(10):971-6.
109. Koopman D, van Dalen JA, Knollema S, Slump C, Jager P. Impact of a high-resolution reconstruction using state-of-the-art PET/CT on intra-thoracic lymph node staging. *Journal of Nuclear Medicine*. 2015;56(supplement 3):1384.
110. Hotta M, Minamimoto R, Yano H, Gohda Y, Shuno Y. Diagnostic performance of (18)F-FDG PET/CT using point spread function reconstruction on initial staging of rectal cancer: a comparison study with conventional PET/CT and pelvic MRI. *Cancer imaging : the official publication of the International Cancer Imaging Society*. 2018;18(1):4.
111. Leite J, Kubo T. Advanced PET reconstruction techniques: Does it matters in the identification of very small pelvic lymph nodes in a PSMA study? *Journal of Nuclear Medicine*. 2019;60(supplement 1):1586.
112. Yang W, Fu Z, Yu J, Yuan S, Zhang B, Li D, et al. Value of PET/CT versus enhanced CT for locoregional lymph nodes in non-small cell lung cancer. *Lung cancer (Amsterdam, Netherlands)*. 2008;61(1):35-43.
113. Hellwig D, Graeter TP, Ukena D, Groeschel A, Sybrecht GW, Schaeffers HJ, et al. 18F-FDG PET for mediastinal staging of lung cancer: which SUV threshold makes sense? *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*. 2007;48(11):1761-6.
114. Yen RF, Chen KC, Lee JM, Chang YC, Wang J, Cheng MF, et al. 18F-FDG PET for the lymph node staging of non-small cell lung cancer in a tuberculosis-endemic country: is dual time point imaging worth the effort? *European journal of nuclear medicine and molecular imaging*. 2008;35(7):1305-15.
115. Hofman MS, Smeeton NC, Rankin SC, Nunan T, O'Doherty MJ. Observer variation in FDG PET-CT for staging of non-small-cell lung carcinoma. *European journal of nuclear medicine and molecular imaging*. 2009;36(2):194-9.
116. Smulders SA, Gundy CM, van Lingen A, Comans EF, Smeenk FW, Hoekstra OS. Observer variation of 2-deoxy-2-[F-18]fluoro-D-glucose-positron emission tomography in mediastinal staging of non-small cell lung cancer as a function of experience, and its potential clinical impact. *Molecular imaging and biology*. 2007;9(5):318-22.
117. Bryant AS, Cerfolio RJ, Klemm KM, Ojha B. Maximum standard uptake value of mediastinal lymph nodes on integrated FDG-PET-CT predicts pathology in patients with non-small cell lung cancer. *The Annals of thoracic surgery*. 2006;82(2):417-22; discussion 22-3.
118. Hwangbo B, Kim SK, Lee HS, Lee HS, Kim MS, Lee JM, et al. Application of endobronchial ultrasound-guided transbronchial needle aspiration following integrated PET/CT in mediastinal staging of potentially operable non-small cell lung cancer. *Chest*. 2009;135(5):1280-7.
119. Kawaguchi T, Takada M, Kubo A, Matsumura A, Fukai S, Tamura A, et al. Performance status and smoking status are independent favorable prognostic factors for survival in non-small cell lung cancer: a comprehensive analysis of 26,957 patients with NSCLC. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer*. 2010;5(5):620-30.
120. Yasukawa M, Sawabata N, Kawaguchi T, Kawai N, Nakai T, Ohbayashi C, et al. Histological Grade: Analysis of Prognosis of Non-small Cell Lung Cancer After Complete Resection. *In vivo (Athens, Greece)*. 2018;32(6):1505-12.
121. Cagini L, Monacelli M, Giustozzi G, Moggi L, Bellezza G, Sidoni A, et al. Biological prognostic factors for early stage completely resected non-small cell lung cancer. *Journal of surgical oncology*. 2000;74(1):53-60.
122. Paesmans M, Garcia C, Wong CY, Patz EF, Jr., Komaki R, Eschmann S, et al. Primary tumour standardised uptake value is prognostic in nonsmall cell lung cancer: a multivariate pooled analysis of individual data. *The European respiratory journal*. 2015;46(6):1751-61.
123. Nair VS, Gevaert O, Davidzon G, Napel S, Graves EE, Hoang CD, et al. Prognostic PET 18F-FDG uptake imaging features are associated with major oncogenomic alterations in patients with resected non-small cell lung cancer. *Cancer research*. 2012;72(15):3725-34.
124. Liu J, Dong M, Sun X, Li W, Xing L, Yu J. Prognostic Value of 18F-FDG PET/CT in Surgical Non-Small Cell Lung Cancer: A Meta-Analysis. *PloS one*. 2016;11(1):e0146195.
125. Hatt M, Laurent B, Fayad H, Jaouen V, Visvikis D, Le Rest CC. Tumour functional sphericity from PET images: prognostic value in NSCLC and impact of delineation method. *European journal of nuclear medicine and molecular imaging*. 2018;45(4):630-41.
126. Leijenaar RT, Carvalho S, Velazquez ER, van Elmpt WJ, Parmar C, Hoekstra OS, et al. Stability of FDG-PET Radiomics features: an integrated analysis of test-retest and inter-observer variability. *Acta oncologica (Stockholm, Sweden)*. 2013;52(7):1391-7.
127. Yan J, Chu-Shern JL, Loi HY, Khor LK, Sinha AK, Quek ST, et al. Impact of Image Reconstruction Settings on Texture Features in 18F-FDG PET. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*. 2015;56(11):1667-73.
128. Hatt M, Majdoub M, Vallières M, Tixier F, Le Rest CC, Groheux D, et al. 18F-FDG PET uptake characterization through texture analysis: investigating the complementary nature of heterogeneity and functional tumor volume in a multi-cancer site patient cohort. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*. 2015;56(1):38-44.

129. Desseroit MC, Visvikis D, Tixier F, Majdoub M, Perdrisot R, Guillemin R, et al. Development of a nomogram combining clinical staging with (18)F-FDG PET/CT image features in non-small-cell lung cancer stage I-III. *European journal of nuclear medicine and molecular imaging*. 2016;43(8):1477-85.
130. Apostolova I, Ego K, Steffen IG, Buchert R, Wertz H, Achenbach HJ, et al. The asphericity of the metabolic tumour volume in NSCLC: correlation with histopathology and molecular markers. *European journal of nuclear medicine and molecular imaging*. 2016;43(13):2360-73.
131. Uesaka D, Demura Y, Ishizaki T, Ameshima S, Miyamori I, Sasaki M, et al. Evaluation of dual-time-point 18F-FDG PET for staging in patients with lung cancer. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*. 2008;49(10):1606-12.
132. Tong S, Alessio AM, Kinahan PE. Image reconstruction for PET/CT scanners: past achievements and future challenges. *Imaging in medicine*. 2010;2(5):529-45.
133. Reynés-Llompard G, Gámez-Cenzano C, Romero-Zayas I, Rodríguez-Bel L, Vercher-Conejero JL, Martí-Climent JM. Performance Characteristics of the Whole-Body Discovery IQ PET/CT System. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*. 2017;58(7):1155-61.
134. Howard BA, Morgan R, Thorpe MP, Turkington TG, Oldan J, James OG, et al. Comparison of Bayesian penalized likelihood reconstruction versus OS-EM for characterization of small pulmonary nodules in oncologic PET/CT. *Annals of nuclear medicine*. 2017;31(8):623-8.
135. Lindström E, Sundin A, Trampal C, Lindsjö L, Ilan E, Danfors T, et al. Evaluation of Penalized-Likelihood Estimation Reconstruction on a Digital Time-of-Flight PET/CT Scanner for (18)F-FDG Whole-Body Examinations. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*. 2018;59(7):1152-8.
136. Um SW, Kim HK, Jung SH, Han J, Lee KJ, Park HY, et al. Endobronchial ultrasound versus mediastinoscopy for mediastinal nodal staging of non-small-cell lung cancer. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer*. 2015;10(2):331-7.
137. Messerli M, Stolzmann P, Egger-Sigg M, Trinckauf J, D'Aguanno S, Burger IA, et al. Impact of a Bayesian penalized likelihood reconstruction algorithm on image quality in novel digital PET/CT: clinical implications for the assessment of lung tumors. *EJNMMI physics*. 2018;5(1):27.
138. Kurita Y, Ichikawa Y, Nakanishi T, Tomita Y, Hasegawa D, Murashima S, et al. The value of Bayesian penalized likelihood reconstruction for improving lesion conspicuity of malignant lung tumors on (18)F-FDG PET/CT: comparison with ordered subset expectation maximization reconstruction incorporating time-of-flight model and point spread function correction. *Annals of nuclear medicine*. 2020;34(4):272-9.
139. Teoh EJ, McGowan DR, Bradley KM, Belcher E, Black E, Moore A, et al. 18F-FDG PET/CT assessment of histopathologically confirmed mediastinal lymph nodes in non-small cell lung cancer using a penalised likelihood reconstruction. *European radiology*. 2016;26(11):4098-106.
140. Teoh EJ, McGowan DR, Macpherson RE, Bradley KM, Gleeson FV. Phantom and Clinical Evaluation of the Bayesian Penalized Likelihood Reconstruction Algorithm Q.Clear on an LYSO PET/CT System. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*. 2015;56(9):1447-52.
141. Schild SE, Tan AD, Wampfler JA, Ross HJ, Yang P, Sloan JA. A new scoring system for predicting survival in patients with non-small cell lung cancer. *Cancer Med*. 2015;4(9):1334-43.
142. Nakamura H, Ando K, Shinmyo T, Morita K, Mochizuki A, Kurimoto N, et al. Female gender is an independent prognostic factor in non-small-cell lung cancer: a meta-analysis. *Ann Thorac Cardiovasc Surg*. 2011;17(5):469-80.
143. Lorensen W, Cline H. Marching Cubes: A High Resolution 3D Surface Construction Algorithm. *ACM SIGGRAPH Computer Graphics*. 1987;21:163.
144. Mullikin JC, Verbeek PW. Surface area estimation of digitized planes. *Bioimaging*. 1993;1(1):6-16.
145. Kaalep A, Sera T, Oyen W, Krause BJ, Chiti A, Liu Y, et al. EANM/EARL FDG-PET/CT accreditation - summary results from the first 200 accredited imaging systems. *European journal of nuclear medicine and molecular imaging*. 2018;45(3):412-22.
146. Kaalep A, Burggraaff CN, Piepenbosch S, Verwer EE, Sera T, Zijlstra J, et al. Quantitative implications of the updated EARL 2019 PET-CT performance standards. *EJNMMI physics*. 2019;6(1):28.
147. Tsutsui Y, Awamoto S, Himuro K, Umezumi Y, Baba S, Sasaki M. Characteristics of Smoothing Filters to Achieve the Guideline Recommended Positron Emission Tomography Image without Harmonization. *Asia Ocean J Nucl Med Biol*. 2018;6(1):15-23.
148. Tsutsui Y, Daisaki H, Akamatsu G, Umeda T, Ogawa M, Kajiwara H, et al. Multicentre analysis of PET SUV using vendor-neutral software: the Japanese Harmonization Technology (J-Hart) study. *EJNMMI research*. 2018;8(1):83.

Danksagung

Ich möchte allen Personen sehr danken, die es mir ermöglicht haben, diese Habilitationsschrift zu verfassen.

Ich danke meiner Partnerin, meinen Geschwistern und Eltern, die mich in den Jahren der klinisch-wissenschaftlichen Tätigkeit und diverser arbeitsreicher Abende und Wochenenden unterstützt haben und mir die dafür notwendigen zeitlichen Freiräume ermöglicht haben.

Ich danke herzlich meinen langjährigen Mentoren, Herrn Prof. Dr. Holger Amthauer und Herrn PD Dr. Christian Furth, für die stets warmherzige und zeitintensive Betreuung, für das frühe Heranführen an eine selbständige wissenschaftliche Arbeit und die vielen gemeinsamen Projekte, die wir seitdem bestreiten konnten.

Ein besonderer Dank gilt auch Herrn Dr. Frank Hofheinz vom Helmholtz-Zentrum Dresden-Rossendorf für seinen Rat und seine Expertise zu medizinphysikalischen und methodischen Problemen sowie für die vielen Telefonate, um Ergebnisse und neue Ideen gemeinsam zu besprechen.

Ich möchte zudem Herrn Dr. Nikolaj Frost, Frau PD Dr. Ivayla Apostolova, Herrn Dr. Oliver Großer, Herrn Dr. Ingo Steffen, Frau Catharina Lange, Herrn Uwe Heimann und Herrn Mathias Lukas für ihre fachliche Unterstützung und hilfreiche Kritik bei verschiedenen wissenschaftlichen Fragen danken.

Erklärung

§ 4 Abs. 3 (k) der HabOMed der Charité

Hiermit erkläre ich, dass

- weder früher noch gleichzeitig ein Habilitationsverfahren durchgeführt oder angemeldet wurde,
- die vorgelegte Habilitationsschrift ohne fremde Hilfe verfasst, die beschriebenen Ergebnisse selbst gewonnen sowie die verwendeten Hilfsmittel, die Zusammenarbeit mit anderen Wissenschaftlern/Wissenschaftlerinnen und mit technischen Hilfskräften sowie die verwendete Literatur vollständig in der Habilitationsschrift angegeben wurden,
- mir die geltende Habilitationsordnung bekannt ist.

Ich erkläre ferner, dass mir die Satzung der Charité – Universitätsmedizin Berlin zur Sicherung Guter Wissenschaftlicher Praxis bekannt ist und ich mich zur Einhaltung dieser Satzung verpflichte.

Datum

Unterschrift