

Aus der Klinik für Radiologie und Nuklearmedizin,
Bereich Nuklearmedizin
der Medizinischen Fakultät
der Otto-von-Guericke-Universität Magdeburg

**Optimierte diagnostische Genauigkeit der F18-FDG-PET zur
prätherapeutischen Dignitätsbeurteilung thorakaler Lymphknoten beim nicht-
kleinzelligen Lungenkarzinom (NSCLC): Einfluss einer Dual Time Point-
Diagnostik und eines visuellen Befundungsscores.**

DISSERTATION

zur Erlangung des Doktorgrades

Dr. med.
(doctor medicinae)

an der Medizinischen Fakultät
der Otto-von-Guericke Universität Magdeburg

vorgelegt von	Ferdinand Ludwig Gunnar Alvaro Steinkrüger
aus	Bonn
Magdeburg	2019

1. Dokumentationsblatt

Bibliographische Beschreibung

Steinkröger, Ferdinand Ludwig Gunnar Alvaro:

Optimierte diagnostische Genauigkeit der F18-FDG-PET zur prätherapeutischen Dignitätsbeurteilung thorakaler Lymphknoten bei nicht-kleinzelligen Lungenkarzinom (NSCLC): Einfluss einer Dual Time Point-Diagnostik und eines visuellen Befundungsscores.

– 2019. - 63 Bl., 11 Abb., 12 Tab.

Kurzreferat

In der vorliegenden Arbeit wurden verschiedene Verfahren zur Optimierung der diagnostischen Genauigkeit der FDG-PET für das prätherapeutische Staging thorakaler Lymphknoten (LK) beim NSCLC untersucht.

Zunächst wurden retrospektiv 47 Patienten mit 155 LKs mit Dual Time Point-PET-Bildgebung (DTPI) untersucht (PET-Akquisition nach ca. 60 und 120 min). Der Uptake der LKs wurde semiquantitativ (SUV und SUV-Ratios) und visuell (4-stufiger Score) bestimmt. Die Genauigkeit der DTPI-Auswertung war dabei für alle Messungen gering und signifikant geringer als die routinemäßige frühe PET allein. Die Genauigkeit der frühen und späten PET war vergleichbar. Späte PET und DTPI erbrachten insgesamt keinen Vorteil. Die frühe PET erreichte für die visuelle Auswertung eine hohe Genauigkeit von 92,9 % und war der semiquantitativen Auswertung ähnlich oder überlegen (Spannweite, 86,5 bis 92,3%).

Der zweite Abschnitt diente der Validierung des visuellen Scores und zur Untersuchung des Interrater-Agreements. Dazu wurden 86 Patienten mit 278 LKs unter standardisierten Bedingungen (Bildfensterung und Scoring) von 3 Readern mit unterschiedlichem Erfahrungsgrad befundet. Die Reader „Unerfahren“, „Erfahren“ und „Experte“ erreichten ähnliche Genauigkeiten von 93,5%, 91,4% und 92,1%. Cohen's κ betrug zwischen 0,92 und 0,96, und Fleiss' κ lag bei 0,93 und zeigte damit keine signifikanten Unterschiede zwischen den Readern.

Die DTPI und späte PET erbrachten keinen Vorteil beim thorakalen LK-Staging bei NSCLC. Der visuelle Score mit seinen standardisierten Kriterien erreichte eine hohe Genauigkeit und robuste LK-Befundung.

Schlüsselwörter:

FDG-PET, NSCLC, LK-Staging, DTPI, SUV, visuelle Beurteilungskriterien, Interrater-Agreement

Inhaltsverzeichnis

1. Dokumentationsblatt	2
2. Abkürzungsverzeichnis.....	5
3. Einführung.....	7
3.1 Lungenkarzinom	7
3.1.1 Epidemiologie, Ätiologie, Pathogenese und Histologie	7
3.1.2 Diagnostik des Lungenkarzinoms	7
3.1.3 Übersicht Therapie.....	11
3.1.4 Prognose und Metastasierung	12
3.2 FDG-PET/CT.....	14
3.2.1 PET/CT	14
3.2.2 CT.....	15
3.2.3 F-18-FDG.....	16
3.2.4 DTPI	17
3.3 LK-Auswertungsmethoden.....	18
3.3.1 Visuelle Auswertung.....	18
3.3.2 Semiquantitative Auswertung (SUVmax, SUVmean)	19
3.4 Zielsetzung der Promotionsschrift	20
4. Patienten, Material und Methoden.....	21
4.1 DTPI-Studie	21
4.1.1 Patienten	21
4.1.2 FDG-PET/CT	22
4.1.3 PET/CT Bilddatenrekonstruktion.....	23
4.1.4 Visuelle Auswertung der PET-Daten.....	23
4.1.5 Semiquantitative Auswertung der PET-Daten.....	25
4.1.6 Referenzstandard (SOR)	26
4.1.7 Statistik.....	26
4.2 Multireader-Analyse.....	27
4.2.1 Patienten	27
4.2.2 FDG-PET/CT	27
4.2.3 Visuelle PET-Beurteilung	27
4.2.4 SOR.....	28
4.2.5 Statistik.....	28
5. Ergebnisse.....	30
5.1 DTPI-Studie	30
5.1.1 Frühe PET	33
5.1.1.1 Visuelle Auswertung.....	33
5.1.1.2 Semiquantitative Auswertung.....	36
5.1.2 Späte PET.....	36
5.1.2.1 Visuelle Auswertung.....	36
5.1.2.2 Semiquantitative Auswertung.....	38
5.1.3 DTPI	38
5.1.3.1 Visuelle Auswertung.....	38

5.1.3.2 Semiquantitative Auswertung	38
5.2 Multireader-Analyse	41
5.2.1 Diagnostische Performance	41
5.2.2 Trennung nach anatomischen oder histologischen Charakteristika	43
5.2.3 Interrater-Agreement	44
6. Diskussion	45
6.1 DTPI-Studie	45
6.1.1 Semiquantitative Auswertung.....	45
6.1.2 Visuelle Auswertung.....	46
6.1.3 Limitationen in Abhängigkeit von der Histologie	47
6.2 Standardisierte visuelle Befundung	48
6.2.1 Validierung des visuellen Scores	48
6.2.2 Interrater-Agreement.....	49
6.2.3 Standardisierung	49
6.3 Limitationen	51
6.4 Schlussfolgerung.....	51
7. Zusammenfassung	52
8. Literaturverzeichnis	53
9. Publikationsverzeichnis	60
10. Danksagung	61
11. Ehrenerklärung	62
12. Lebenslauf	63

2. Abkürzungsverzeichnis

AC	Adenokarzinom
AUC	area under the curve (Fläche unter der Kurve)
AWMF	Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften
BGO	Bismutgermanat
BMI	Body-Mass-Index
bspw.	beispielsweise
bzw.	beziehungsweise
CI	confidence interval (Konfidenzintervall)
CT	Computertomographie
EANM	European Association of Nuclear Medicine
EBUS	endobronchialer Ultraschall
ESTS	European Society of Thoracic Surgeons
EUS	endoscopic ultrasound (Endosonografie)
FDG	Fluordesoxyglukose
FN	falsch negativ
FOV	field of view (Gesichtsfeld)
FP	falsch positiv
FWHM	full width at half maximum (Halbwertsbreite)
ggf.	gegebenenfalls
GLUT	Glukosetransporter
HU	Hounsfield unit (Hounsfield-Einheit)
IQR	interquartile range (Interquartilsabstand)
LK	Lymphknoten
LoA	Limits of Agreement
LOR	line of Response (Koinzidenzlinie)
MBPS	Mediastinale Blutpoolstrukturen
MIP	Maximumintensitätsprojektion
MRT	Magnetresonanztomographie
n	Anzahl
NN	Nebenniere
NOS	nicht anderweitig spezifiziert
NPV	negative predictive value (negativer Vorhersagewert)
NSCLC	nichtkleinzelliges Lungenkarzinom
OSEM	ordered subset expectation maximization

PET	Positronen-Emissions-Tomographie
PPV	positive predictive value (positiver Vorhersagewert)
PSF	point spread function (Punktabbildungsfunktion)
RI	Retentionsindex
RP	richtig positiv
RN	richtig negativ
ROC	receiver operating characteristics
ROI	region of Interest
SCC	squamous-cell carcinoma (Plattenepithelkarzinom)
SCLC	Kleinzelliges Lungenkarzinom
SOR	standard of reference (Referenzstandard)
SUV	standardized uptake value
TBNA	transbronchiale Nadelaspiration
TOF	time-of-flight
TTNA	transthorakale Nadelaspiration
UICC	Union internationale contre le cancer
VATS	videoassistierte Thorakoskopie
VOI	volume of Interest
vs.	versus
ZNS	Zentralnervensystem

3. Einführung

3.1 Lungenkarzinom

3.1.1 Epidemiologie, Ätiologie, Pathogenese und Histologie

Lungenkarzinome machen hierzulande 8% aller weiblichen und 13,7% aller männlichen Krebsneuerkrankungen aus und sind bei Männern die häufigste und bei Frauen die dritthäufigste Todesursache in Deutschland. Die Mortalität hat sich dabei insbesondere bei Frauen seit 1975 verdoppelt (Serke et al. 2013). Ab dem sechsten Lebensjahrzehnt nimmt die Inzidenz zu. Insgesamt liegt der Häufigkeitsgipfel zwischen dem 75.-80. Lebensjahr (Hammerschmidt und Wirtz 2009).

Zigarettenrauchinhalation ist mit 85% die häufigste Ursache aller Lungenkarzinome. Entscheidend ist hier nicht nur die Dauer der Rauchinhalation, sondern auch die Häufigkeit der Exposition sowie das Vorliegen von zusätzlichen Risikofaktoren wie die Exposition gegenüber ionisierender Strahlung, Asbest, Feinstaub sowie allgemein Verkehrs- und Industriegasen (Alberg et al. 2013). Der Einfluss der Risikofaktoren auf die histologischen Subtypen ist dabei unterschiedlich hoch (Samet et al. 2009).

Der Pathogeneseprozess kann mehrstufig eingeteilt werden: Exposition mit einem karzinogenen Stoff führt zu genetischen Schäden des Lungengewebes und nach einer Latenzphase folgt aus der Epitheldysplasie ein Carcinoma in situ.

Das Lungenkarzinom wird hauptsächlich in kleinzelliges (SCLC) mit einem Anteil von 15% und nicht-kleinzelliges Lungenkarzinom (NSCLC) mit einem Anteil von 85% eingeteilt. Das NSCLC wird vorwiegend in ca. 40% Plattenepithelkarzinome (SCC) und ca. 35% Adenokarzinome (AC) eingeteilt; weiterhin lassen sich noch das großzellige Lungenkarzinom (ca. 10%) sowie weitere, seltenere Subtypen unterscheiden.

3.1.2 Diagnostik des Lungenkarzinoms

Als Basisdiagnostik kommt häufig eine Röntgenaufnahme nach dem Auftritt von unspezifischen Symptomen zum Einsatz. Die häufigsten Symptome umfassen dabei Husten (8 bis 75%), Gewichtsverlust (0 bis 68%), Luftnot (3 bis 60%), Brustschmerzen (20 bis 49%), Hämoptysen (6 bis 35%) und Knochenschmerzen (6 bis 25%) (Spiro et al. 2007). Die Veranlassung zum Röntgen hat dabei meist eine andere Ursache, und der Lungenherd tritt oft als Zufallsbefund in Erscheinung.

Laut S3-Leitlinie wird für die Ausbreitungsdiagnostik des Primärtumors eine Computertomographie (CT) mit Kontrastmittel von Thorax und Oberbauch empfohlen.

Sollten in der CT mediastinale LK-Vergrößerungen über 1 cm auffallen und kein Hinweis auf Fernmetastasen vorliegen, sollte der mediastinale LK-Status vor der Behandlung des Primärtumors evaluiert werden (Abbildung 1).

Bei kurativer Behandlungsintention kann im Stadium IA eine F18-Fluorodesoxyglukose-Positronen-Emissions-Tomographie/ Computertomographie (FDG-PET/CT) für das mediastinale und extrathorakale Staging zum Einsatz kommen. Für die klinischen Stadien IB-III B und das metastasierte Stadium IVA mit M1b sollte auch bei negativem klinischem Untersuchungsbefund und kurativer Behandlungsintention eine Untersuchung auf extrathorakale Metastasen mittels Ganzkörper-PET/CT und MRT-Schädel erfolgen. (AWMF 2018).

Gemäß Leitlinienempfehlungen der *European Society of Thoracic Surgeons* (ESTS) beinhaltet das prätherapeutische LK-Staging bei kurativer Intention zudem eine invasive Abklärung; ausgenommen davon sind lediglich kleine periphere Tumoren ohne bildgebenden Hinweis für hiläre LK-Metastasen (äußeres Drittel der Lunge, Größe ≤ 3 cm, cN0) (De Leyn et al. 2014). Zum Einsatz kommen dafür Nadelaspirationsverfahren (endobronchialer Ultraschall [EBUS], Endosonografie [EUS], transbronchiale Nadelaspiration [TBNA], transthorakale Nadelaspiration [TTNA]) und chirurgische Verfahren wie Mediastinoskopie oder videoassistierte Thorakoskopie (VATS). Im Falle von Kontraindikationen stehen weitere diagnostischen Möglichkeiten zur Verfügung.

Nur in den Fällen, dass sich ein bildgebender Befund pathologisch bestätigt, klinisch eindeutig oder radiologisch evident ist, wird von einer möglich kurativen Behandlung abgeraten.

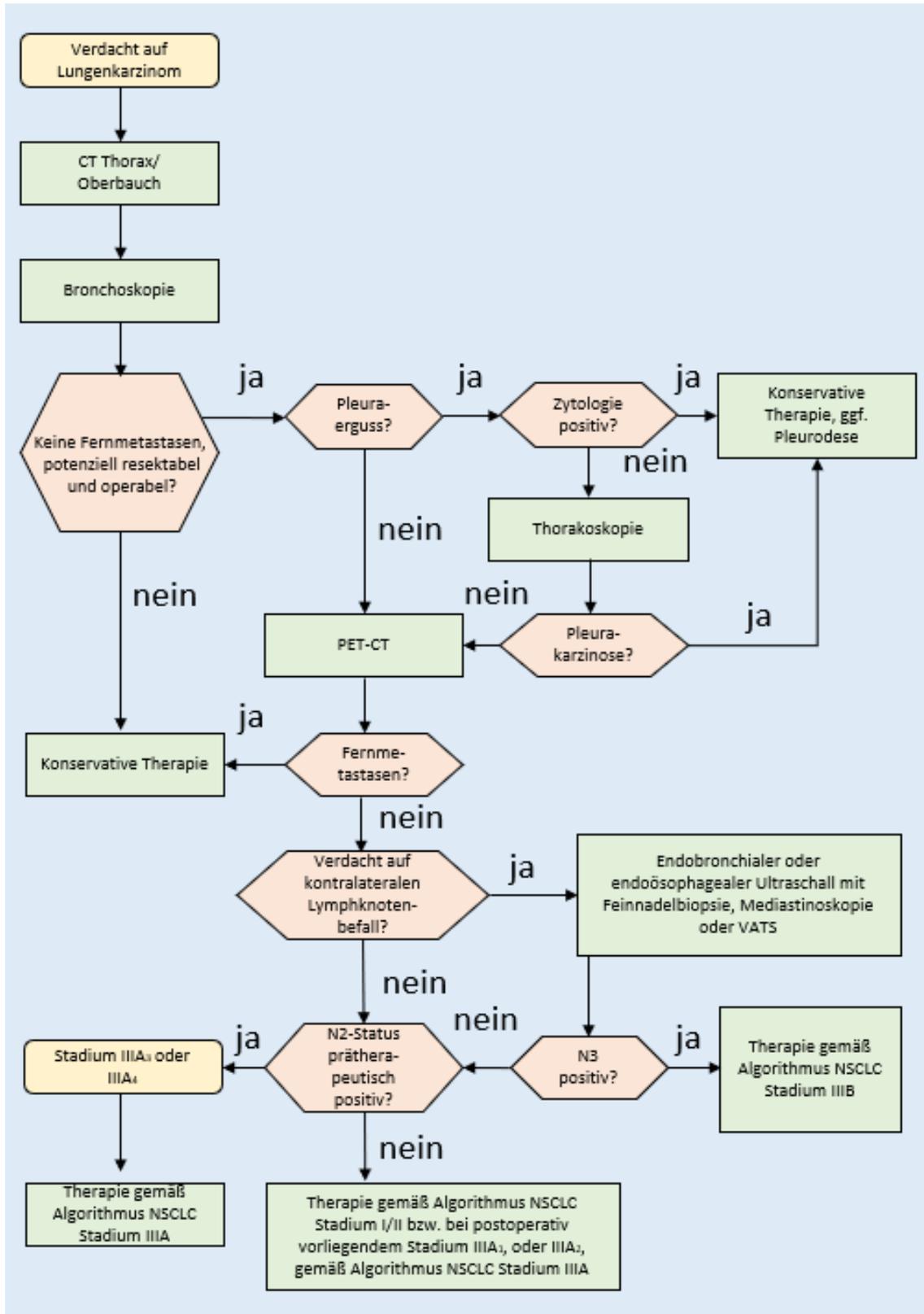


Abbildung 1: Diagnostischer Algorithmus des NSCLC nach S3-Leitlinie (Februar 2018)

Das Ausmaß, die Ausbreitung des Primärtumors und Fernmetastasen werden über die aktuelle 8. TNM-Klassifikation nach IASLC festgehalten und sind seit dem 01.01.2017 gültig (Tabelle 1). Wesentliche Änderungen zur Vorgängerversion sind eine weitere Unterteilung der T- und M-Stadien (Tabelle 2).

Kategorie	Stadium	Kurzbeschreibung
T (Tumor)	Tis	Carcinoma in situ
	T1	größter Durchmesser < 3cm, umgeben von Lungengewebe oder viszeraler Pleura, Hauptbronchus nicht beteiligt
	T1a(mi)	Minimal invasives Adenokarzinom
	T1a	größter Durchmesser ≤ 1cm
	T1b	größter Durchmesser > 1 ≤ 2 cm
	T1c	größter Durchmesser > 2 ≤ 3 cm
	T 2	größter Durchmesser > 3 aber < oder = 5 cm <u>oder</u> Infiltration des Hauptbronchus unabhängig vom Abstand von der Carina aber ohne direkte Invasion der Carina Infiltration der viszeralen Pleura <u>oder</u> tumorbedingte partielle Atelektase oder obstruktive Pneumonie die bis in den Hilus reichen, Teile der Lunge oder die gesamte Lunge umfassen
	T2a T2b	größter Durchmesser > 3 ≤ 4 cm größter Durchmesser > 4 ≤ 5 cm
T3	größter Durchmesser > 5 aber ≤ 7 cm Infiltration von Thoraxwand (inklusive parietale Pleura und Superior Sulcus), N. phrenicus, <u>oder</u> parietales Perikard zusätzlicher Tumorknoten im selben Lungenlappen wie der Primärtumor	
T4	größter Durchmesser > 7cm oder <u>mit</u> direkter Infiltration von Diaphragma, Mediastinum, Herz, großen Gefäßen, Trachea, N. laryngeus recurrens, Ösophagus, Wirbelkörper oder Karina zusätzlicher Tumorknoten in einem anderen ipsilateralen Lungenlappen	
N (Lymphknoten)	N0	keine Lymphknotenmetastasen
	N1	Metastase in ipsilateralen, peribronchialen und / oder ipsilateralen hilären Lymphknoten und / oder intrapulmonalen Lymphknoten oder direkte Invasion dieser Lymphknoten
	N2	Metastase in ipsilateralen mediastinalen und / oder subkarinalen Lymphknoten
	N3	Metastase in kontralateralen mediastinalen, kontralateralen hilären, ipsi- oder kontralateral tief zervikalen, supraklavikulären Lymphknoten
M (Metastase)	M0	Keine Fernmetastasen
	M1	Fernmetastasen
	M1a	separater Tumorknoten in einem kontralateralen Lungenlappen Pleura mit knotigem Befall maligner Pleuraerguss maligner Perikarderguss
	M1b	isolierte Fernmetastase in einem extrathorakalen Organ
	M1c	mehrere Fernmetastasen (>1) in einem oder mehreren Organen

Tabelle 1: TNM-Klassifikation des NSCLC nach IASLC 2017 (Goldstraw et al. 2016)

Stadium	Primärtumor	Lymphknoten	Fernmetastasen
0	Tis	N0	M0
IA1	T1a(mi)	N0	M0
	T1a	N0	M0
IA2	T1b	N0	M0
IA3	T1c	N0	M0
IB	T2a	N0	M0
IIA	T2b	N0	M0
IIB	T1a-c	N1	M0
	T2a	N1	M0
	T2b	N1	M0
	T3	N0	M0
IIIA	T1a-c	N2	M0
	T2a-b	N2	M0
	T3	N1	M0
	T4	N0	M0
	T4	N1	M0
IIIB	T1a-b	N3	M0
	T2 a-b	N3	M0
	T3	N2	M0
	T4	N2	M0
IIIC	T3	N3	M0
	T4	N3	M0
IVA	jedes T	jedes N	M1a
	jedes T	jedes N	M1b
IVB	jedes T	jedes N	M1c

Tabelle 2: Klassifikation der Tumorstadien nach UICC (8. Edition) (Goldstraw et al. 2016)

3.1.3 Übersicht Therapie

Nach den S3-Leitlinien der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF) ist die operative Resektion des Lungenlappens der wichtigste Prognosefaktor für die Stadien IA bis IIB. Nur bei kardiopulmonaler Einschränkung wird eine limitierte Resektion empfohlen. Um Komplikationen gering zu halten, sollte möglichst Parenchym bei der Resektion erhalten werden. Eine Manschettenresektion sollte daher -wenn möglich- der Pneumektomie vorgezogen werden. Brustwandresektionen bei Brustwandinfiltrationen bilden, abhängig von ihrer Radikalität, einen signifikanten prognostischen Faktor (AWMF 2018).

Bei kurativer Therapieintention ist das therapeutische Vorgehen maßgeblich vom N-Stadium abhängig. Laut den ESMO-Leitlinien sollte bei einem N0- bis N1-Stadium chirurgisch, bei einem N2-Stadium und Resektabilität chirurgisch-multimodal und bei einem N3- bzw. einem

nicht resektablen N2-Stadium nicht-chirurgisch-multimodal vorgegangen werden (Postmus et al. 2017).

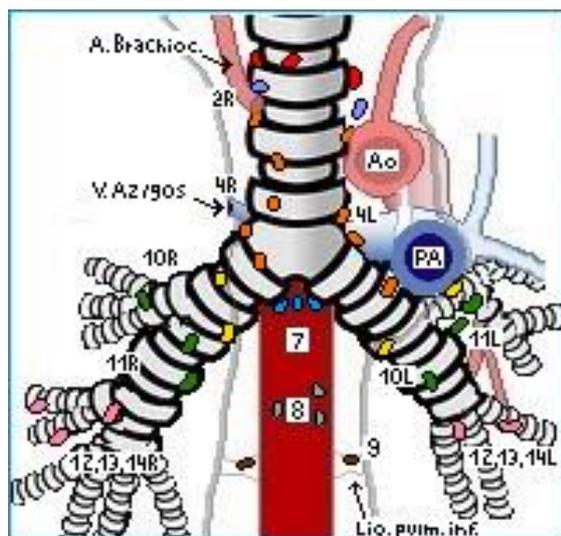
Patienten in gutem Allgemeinzustand nach R0-Resektion und LK-Dissektion sollten im Stadium II bzw. IIIA1/IIIA2 eine adjuvante Chemotherapie erhalten. Es werden adjuvant cisplatinhaltige Kombinationen mit 4 Zyklen empfohlen (AWMF 2018). Für das Stadium IIIB wird eine kombinierte Strahlenchemotherapie angeraten.

Bei der Mehrzahl der Patienten im Stadium IV ist das therapeutische Vorgehen palliativ. Bei Patienten mit solitären Metastasen des Zentralnervensystems (ZNS) oder Nebennieren-Metastasen (NN) und einer lokalen Tumorausdehnung von maximal IIIA3 kann eine optimale Lokalthherapie von Primärtumor und Metastasen, gefolgt von additiver Chemotherapie, erfolgen (DGHO 2017).

3.1.4 Prognose und Metastasierung

Die 5-Jahres-Überlebensrate ist beim NSCLC mit zunehmendem Stadium abnehmend (Tabelle 3). Beträgt sie im Stadium IA noch 75% bis 80%, sinkt sie im Stadium IIIB auf 5 bis 10%; Stadium IV ist in der Mehrzahl palliativ (DGHO 2017). Zum Diagnosezeitpunkt sind 65% der Lungenkarzinome bereits inoperabel, aufgrund von Metastasierung oder unzureichender Lungenfunktion (funktionell inoperabel).

Tumorzellen erreichen auf lymphatischem Wege vor allem intrapulmonale, hiläre und mediastinale LKs sowie Skalenus-LKs und supraklavikuläre LKs (Abbildung 2). Eine Fernmetastasierung erfolgt hingegen bevorzugt hämatogen; betroffen sind vorzugsweise Hirn, Knochen und NN (Popper 2016).



Superiore mediastinale LKs

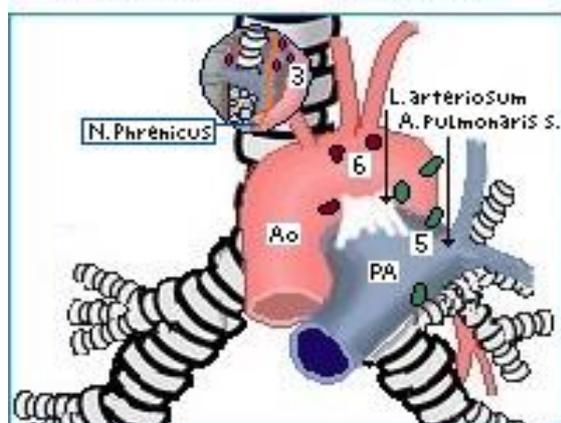
- 1 Oberste mediastinale
- 2 Obere paratracheale
- 3 Prä- und retrotracheale
- 4 Untere paratracheale (inklusive Azygosknoten)

Aortale LKs

- 5 Subaortale (LKs im Aortenfenster)
- 6 Paraaortale (LKs an Aorta ascendens oder phrenische LKs)

Inferiore mediastinale LKs

- 7 Subcarinale
- 8 Paraösophageale (unter Carina)
- 9 LKs im Lig. pulmonale



N_i LKs

- 10 Hiläre
- 11 Interlobäre
- 12 Lobäre
- 13 Segmentäre
- 14 Subsegmentäre

Abbildung 2: Regionale LK-Klassifikation nach Mountain und Dresler (Mountain und Dresler 1997)

Stadium	5-Jahresüberlebensrate
IA1	92%
IA2	83%
IA3	77%
IB	68%
IIA	60%
IIB	53%
IIIA	36%
IIIB	26%
IIIC	13%
IVA	10%
IVB	0%

Tabelle 3: 5-Jahresüberlebensrate nach Goldstraw (Goldstraw et al. 2016)

3.2 FDG-PET/CT

3.2.1 PET/CT

Die Vereinigung der beiden Verfahren CT und PET in den 1990er Jahren durch Townsend und Nutt führte zur Etablierung der PET/CT ab der Jahrtausendwende (Townsend 2008). Durch die Kombination wird die funktionelle Aktivität der PET mit der anatomisch-morphologischen Information der CT gemeinsam dargestellt.

Das Prinzip der PET ist die Darstellung der räumlichen Verteilung eines Radiopharmakons. Die physikalischen Prinzipien dahinter beruhen auf der Paarvernichtung. Trifft das durch den Zerfall eines Positronenstrahlers (Fluor-18 im Falle des FDG) emittierte Positron auf ein Elektron des umgebenden Gewebes, kommt es zur Annihilation beider Teilchen und damit zur Emission von zwei Gammaquanten. Aufgrund der beiden Ruhemassen werden die beiden Quanten mit einer Energie von 511 keV -idealerweise - in einem Winkel von 180° zueinander ausgestrahlt. Gammadetektoren sind dabei in einem geschlossenen Kreis um den Patienten angeordnet. Damit der Koinzidenznachweis erfolgen kann, müssen beide Quanten an entgegengesetzten Detektionskristallen registriert werden und sich auf der Geraden, der sogenannten Line of Response (LOR), befinden (Abbildung 3). Außerdem muss die Registrierung innerhalb eines Zeitfensters (Koinzidenzfenster im Nanosekundenbereich) und Energiefensters (typischerweise 400 bis 650 keV) stattfinden, um in der nachfolgenden Bildrekonstruktion als echtes Bildsignal bewertet zu werden. Unter diesen Voraussetzungen kann dann auf den Ursprungsort der Paarvernichtung geschlossen werden (Mohnike et al. 2011). Die Szintillatorkristalle sind wiederum in Detektorblöcken angeordnet. Zur Verstärkung des Photonensignales sind die Kristalle an Photomultiplier gekoppelt (Dössel 2016). In Sinogrammen werden die Informationen über Entfernung und Drehwinkel einer Koinzidenzlinie in Referenz auf das Zentrum des Detektorringes festgehalten (Fahey 2002). Aus diesen Rohdaten können dann PET-Bilddaten rekonstruiert werden, die ein möglichst genaues Abbild der wahren Aktivitätsverteilung im untersuchten Areal darstellen.

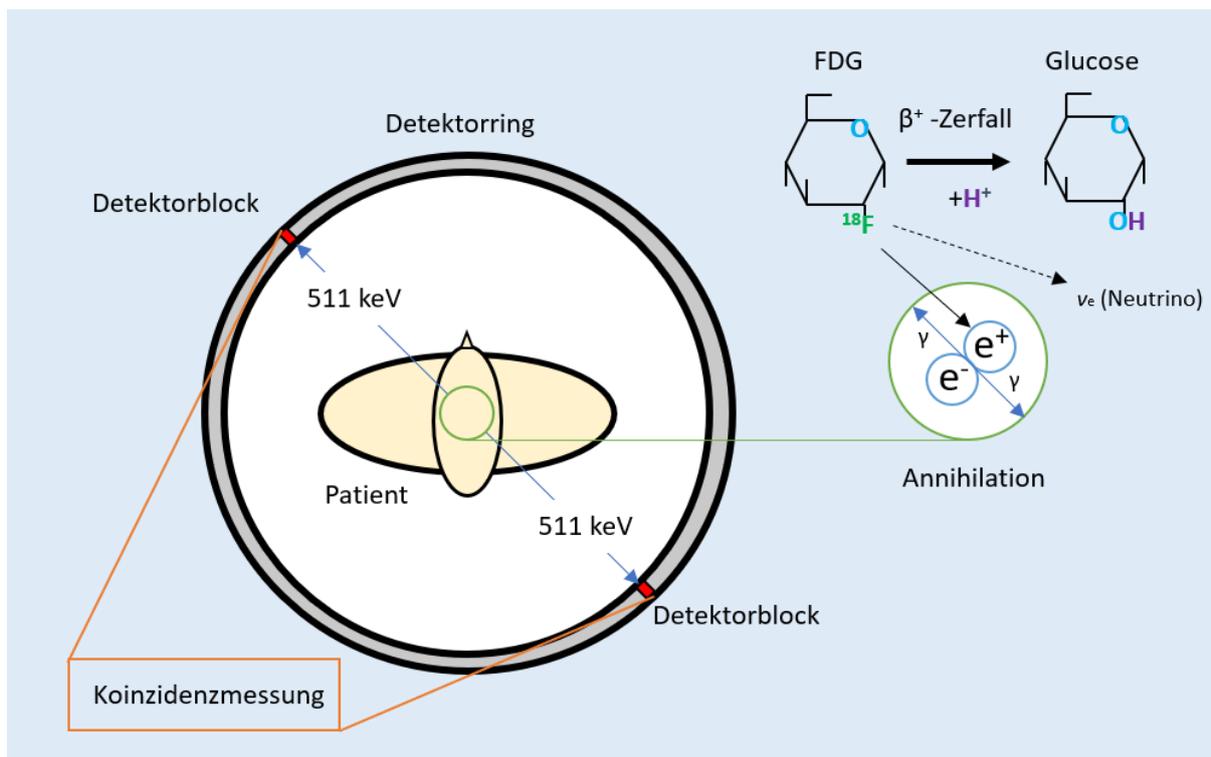


Abbildung 3: Koinzidenzmessung beim PET

Das F-18 des FDG zerfällt durch einen β^+ -Zerfall zu O-18 und emittiert dabei ein Positron, aus FDG wird mit einem freien Wasserstoffatom somit Glucose. Durch Paarvernichtung des Positrons und eines Elektrons entstehen zwei Gammaquanten, die an gegenüberliegenden Detektorblöcken gemessen werden und die Koinzidenzmessung erlauben.

3.2.2 CT

Bei der CT handelt es sich um ein Verfahren, das mit Hilfe von Röntgenstrahlen überlagerungsfreie Schnittbilder erzeugt. Sie ist dadurch in der Lage geringe Dichteunterschiede zwischen verschiedenen Weichteilgewebetypen gut wiederzugeben.

Während des Scanvorganges werden Röntgenstrahlen emittiert, die das zu untersuchende Objekt durchdringen und am gegenüberliegenden Detektor registriert werden. Damit der Scanner aus unterschiedlichen Winkeln Messungen vornehmen kann, wird der Strahler bzw. der Detektor nach jeder Messung um eine Position gedreht und die Messung wiederholt. Die unterschiedliche Dichte der untersuchten Materie spiegelt sich durch unterschiedliche Schwächungswerte am Detektor wider. Die Einheit der Schwächung wird in Hounsfield angegeben (HU). Die geschwächte Strahlung wird in ein elektrisches Signal umgewandelt und zur Rekonstruktion an einen Rechner weitergeleitet. Dort wird eine Matrix von Volumenelementen (Voxeln) bestimmt. Der mittlere Röntgenabsorptionskoeffizient wird in einen Grauwert übersetzt, der im betreffenden Pixel des 2-dimensionalen CT-Bildes dargestellt wird (Alkadhi et al. 2011).

Als Einzelmodalität innerhalb der PET/CT stellt die CT die anatomische Grundlage für die Bildfusion mit den PET-Befunden. Als low-dose-CT kommen CT-Transmissionsbilder für die Schwächungs- und Streukorrektur der Emissionsdaten zum Einsatz. Dadurch kann auf zeitaufwendigen PET-Transmissionsscans verzichtet werden (von Schulhess 2000).

3.2.3 F-18-FDG

F-18-FDG ist ein in der klinischen Routine gebräuchlicher Tracer, der bei stoffwechselbezogenen Fragestellungen in der Neurologie, Kardiologie und Onkologie Verwendung findet. Dabei wird das Radiopharmakon meist intravenös injiziert und über GLUT-Transporter in die Zielzellen (z.B. Tumorzellen) aufgenommen. In der Zelle erfolgt die Phosphorylierung der Glukose durch eine Hexokinase zu FDG-6-Phosphat. Während herkömmliche Glukose bis zur Phosphorylierung die gleichen Schritte vollzieht und anschließend durch die Glucose-6-phosphat-Isomerase zu Fruktose-6-Phosphat isomerisiert wird, kann FDG-6-Phosphat nicht katabolisiert werden, da ein Sauerstoffatom an der C-2 Position fehlt (Kelloff et al. 2005). Da der Tracer nun in der Zelle „gefangen“ (*trapping*) ist (Gallagher et al. 1978), erfolgt eine langsame Dephosphorylierungsreaktion, wobei eine Downregulation der Hexokinase auch eine Rolle spielt (Nelson et al. 1996). FDG-6 akkumuliert proportional zu der Glukoseverstoffwechslung (Abbildung 4). Wie schon von Warburg postuliert, haben neoplastische Zellen einen erhöhten Energiebedarf und damit auf zellulärer Ebene auch mehr GLUT-Transporter; folglich die erhöhte FDG-Aufnahme und Abgrenzung zu gesundem Gewebe. Anzumerken ist dabei, dass der FDG-Umsatz abhängig ist von der Dynamik der Zelle und dem histologischen Typ. Physiologischerweise haben verschiedene Gewebe und Entzündungszellen unterschiedliche GLUT-Transporter (Wood und Trayhurn 2003, Mochizuki et al. 2001), weshalb bei der Bildbefundung auch die physiologische Tracerdistribution im Patienten zu beachten ist.

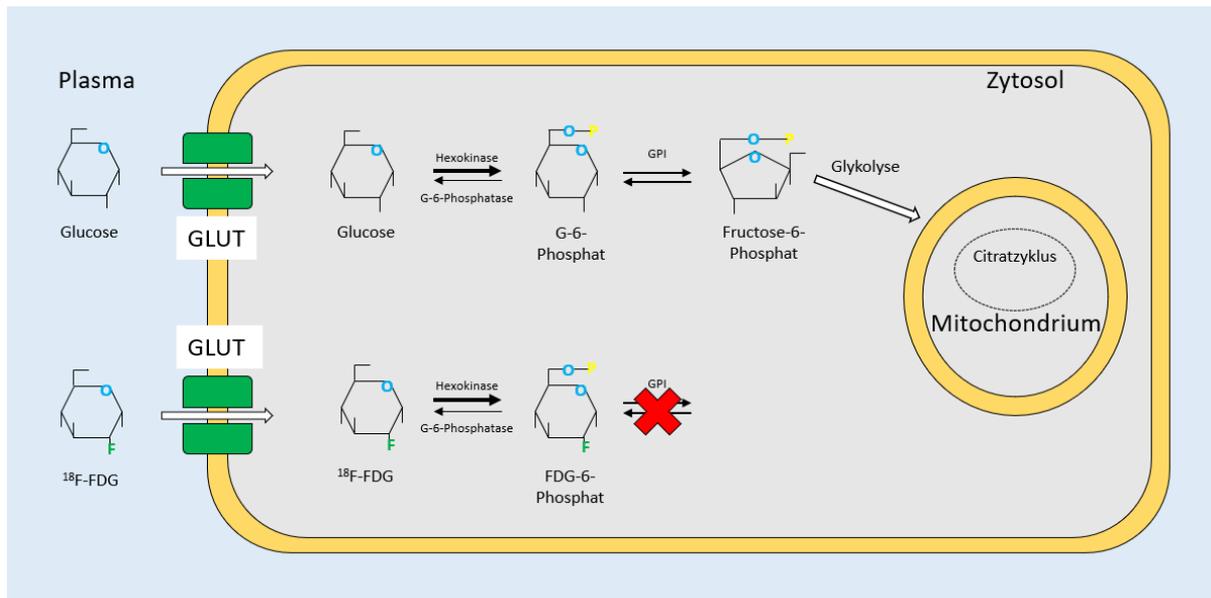


Abbildung 4: FDG-Trapping

Beide Glucose-Moleküle gelangen über einen GLUT-Transporter in die Zelle und werden über eine Hexokinase phosphoryliert. F-18-FDG kann nicht durch die G-6-P-Isomerase isomerisiert werden. Durch die ausbleibende Glykolyse bleibt F-18-FDG in der Zelle „gefangen“ (*trapping*).

3.2.4 DTPI

Die Dual Time Point beschreibt ein Verfahren, bei dem Patienten zu zwei unterschiedlichen Zeitpunkten nach der Verabreichung des Radiopharmakons untersucht werden. Hintergrund ist die bereits oben beschriebene veränderte Pharmakodynamik von FDG in Tumorzellen und in Entzündungszellen. Da auch letztere einen gesteigerten FDG-Metabolismus aufweisen können, können insbesondere semiquantitative Messungen zu einem isolierten Zeitpunkt zu falsch positiven Ergebnissen führen, die Spezifität für die Diagnostik maligner Läsionen sinkt somit (Matthies et al. 2002). Durch eine kontinuierliche Eliminierung von physiologischer bzw. unspezifischer Hintergrundaktivität kann zu einem späteren Messzeitpunkt ein besseres Tumor-zu-Hintergrund-Verhältnis vorgefunden werden und damit eine bessere Unterscheidung von benignen und malignen Läsionen ermöglicht werden (Basu und Alavi 2007). Laut Houshmand et al. könnte der späte Aufnahmezeitpunkt die Beurteilung der Tumorerogenität durch zuverlässigere Trennung von Primärtumor und lokoregionären Metastasen verbessern (Houshmand et al. 2016).

In der Literatur finden sich mehrere Studien, die die DTPI für das thorakale LK-Staging untersucht haben, so konnten Suga et al. bei 67 Patienten, von denen einige an NSCLC litten, zeigen, dass die DTPI (Sensitivität 88,8%) die Diskrimination von malignen und benignen LKs im Vergleich zu den Einzelzeitpunktmessung verbessert (85,8% bzw. 73,1%). Für die

Unterscheidung der LKs wurde ein *standardized uptake value* (SUV) max für die frühe und späte PET sowie ein $\% \Delta \text{SUV}_{\text{max}}$ angewendet (Suga et al. 2009).

Kritischer äußern sich Li M. et al., die 265 LK mittels DTPI untersuchten (39 davon PET-positiv zum frühen Zeitpunkt). Der Retentionsindex (RI), also die relative Änderung des SUV vom frühen zum späten Zeitpunkt, zeigte keinen signifikanten Unterschied zwischen benignen und malignen LK ($P > 0,05$), weshalb die Autoren der DTPI nur einen limitierten Effekt für die Differenzierung von richtig positiven und falsch positiven LKs zusprechen (Li et al. 2012). In einer Metaanalyse bescheinigen Shen et al. der DTPI eine verbesserte diagnostische Aussagekraft (Sensitivität 85%, Spezifität 75%) gegenüber der Einzelzeitpunktmessung (79% bzw. 73%). Allerdings schränken sie ein, dass die Stichprobengröße der gepoolten Studien klein war und Unterschiede zwischen den Studien, z.B. unterschiedlich späte PET-Akquisitionszeitpunkte, Bewertungskriterien oder Referenzstandards zu groß waren, um eine Anwendung der DTPI in der klinischen Routine zu rechtfertigen (Shen et al. 2015).

Aufgrund der kontroversen Studienlage infolge unterschiedlicher Studienbedingungen und Messkriterien, die eine zweifelsfreie Vergleichbarkeit verhindern, ist bisher somit kein ausreichender Beleg vorhanden, der die Anwendung des DTPI rechtfertigen würde.

3.3 LK-Auswertungsmethoden

3.3.1 Visuelle Auswertung

Die visuelle Auswertung von PET-Bildern stellt immer noch den klinischen Standard dar (Boellaard 2013). Sie beruht auf dem Kontrast zwischen Bereichen mit hohem und niedrigem Uptake, die, wie oben beschrieben, die unterschiedlichen FDG-Konzentrationen widerspiegeln. Diese Form der Auswertung ist schnell und unkompliziert durchzuführen, allerdings fehlen einheitliche Kriterien zur Bildeinstellung (Fensterung) und Befundung. Sie ist außerdem vom Erfahrungsgrad des Befunders abhängig, was sie anfällig für Interobserver-Variabilität macht und einer gleichbleibenden Befundungsqualität hinderlich ist (Houshmand et al. 2015). Eine Vergleichbarkeit von verschiedenen Studien und Untersuchungen unterschiedlicher Zentren ist durch diesen Umstand beeinträchtigt.

In der Literatur werden uneinheitliche Kriterien für die PET-Positivität in der visuellen LK-Befundung beim NSCLC beschrieben: Von einem einfachen „fokal prominenten Uptake verglichen mit den umgebenden mediastinalen Strukturen“ (Yang et al. 2008) über einen 4-stufigen Score mit einer „Intensität niedriger als die mediastinalen Blutpoolstrukturen (MBPS)“ bis „viel höher als die MBPS“ (positiv, wenn höher als die MBPS) (Yen et al. 2008) bis hin zu einem 6-stufigen Score von „klar negativ“ bis „klar positiv“ mit der Intensität der MBPS als Referenzstruktur (Hellwig et al. 2007). Keine von diesen diagnostischen Kriterien hat Eingang

in die klinische Routine gefunden. Es besteht daher eine Nachfrage nach standardisierten und allgemein akzeptierten Kriterien für die visuelle PET-Befundung.

3.3.2 Semiquantitative Auswertung (SUVmax, SUVmean)

Der *standardized uptake value* (SUV) ist der meistverwendete semiquantitative Parameter zur Befundcharakterisierung und Dignitätsbeurteilung in der PET-Diagnostik (Basu und Alavi 2007). Da die Genauigkeit der Charakterisierung von Lungenherden mittels FDG-PET von der Aufnahme in den Tumor abhängt, empfehlen die S3-Leitlinien die FDG-Anreicherung als SUV anzugeben (AWMF 2018).

Der SUV lässt sich als Formel folgendermaßen ausdrücken: $SUV = C/(D/W)$. Dabei steht C für die fokale Aktivitätskonzentration (kBq/ml), D für die injizierte Aktivität (kBq) und W für das Körpergewicht (g). Der SUV stellt somit die relative fokale FDG-Anreicherung verglichen mit dem Durchschnitt innerhalb des Körpers dar; er wird meist als dimensionslose Zahl angegeben (Moloney et al 2004). Andere SUV-Definitionen setzen die Aktivitätskonzentration anstelle des Körpergewichts mit der fettfreien Körpermasse („lean body mass“) oder der Körperoberfläche („body surface area“) in Relation. Der meist benutzte Parameter ist dabei der SUVmax, also die maximal berechnete Aktivitätskonzentration innerhalb eines Zielbereichs (sogenannte *region of interest* [ROI]), wohingegen der SUVmean den durchschnittlichen Uptake innerhalb eines solchen Zielbereichs darstellt. Für die Beurteilung von Lungenrundherden kommt die PET/CT bei Malignomverdacht nach den S3-Leitlinien ab einer Größe von 8 mm zum Einsatz, wenn eingeschränkte funktionelle Operabilität besteht (AWMF 2018). Als PET-positiv wird generell angenommen, wenn der Lungenrundherd einen SUV von 2,5 aufweist (Groheux et al. 2016). Zur Dignitätsbeurteilung von LKs wird üblicherweise ebenfalls ein SUV-Schwellenwert von 2,5 für den SUVmax verwendet, allerdings wird dieser Wert von mehreren Autoren in Frage gestellt, da er zwar zur Hilfestellung von Nutzen ist, die Genauigkeit jedoch von mehreren Einflussfaktoren abhängig ist und daher sehr variieren kann (Hellwig et al. 2007, Schmidt-Hansen et al. 2014, Kinahan et Fletcher 2010). Es gibt darüber hinaus Hinweise in der Literatur, dass der SUVmax des Primarius ein Prognosefaktor beim NSCLC sein kann (Tournoy et al. 2007, Takeda et al. 2014).

Einflussfaktoren auf den SUV-Wert sind auf der technischen Seite unterschiedliche Scannertypen und Rekonstruktionsalgorithmen sowie Fehler bei der Kalibrierung von Scanner und Aktivimeter (Akamatsu et al 2014, Adams et al. 2010). Biologische und physikalische Einflussfaktoren sind Patientengröße bzw. -gewicht und die applizierte Aktivität. Eine weitere Rolle spielt die Latenzzeit zwischen Applikation und Bildakquisition, da zum einen der radioaktive Zerfall mit der Zeit zu Signalverlust führt, die FDG-Verteilung im Körper sich zum anderen mit der Zeit ändert (insbesondere „Auswaschen“ der physiologischen

Hintergrundaktivität und ggf. zunehmende Tracerakkumulation im Tumorgewebe). Insbesondere die FDG-Aufnahme in malignen Tumoren steht in komplexer Beziehung zu der Proliferation des Tumorgewebes und der Anzahl der Tumorzellen (Kinahan und Fletcher 2010). Die Korrektheit des SUV-basierten LK-Stagings beim NSCLC kann außerdem durch FDG-Anreicherungen in benignen LK-Veränderungen wie Entzündungen und Anthrakose sowie unterschiedliche Anreicherungsmuster verschiedener histologischer NSCLC-Subtypen beeinträchtigt werden (Park et al. 2013, Li S. et al. 2013).

3.4 Zielsetzung der Promotionsschrift

Das Ziel dieser Arbeit war, den diagnostischen Zusatznutzen der DTPI gegenüber der etablierten Einzelzeitpunktmessung für das prätherapeutische LK-Staging bei Patienten mit NSCLC zu überprüfen („Gibt es einen diagnostischen Mehrwert des DTPI, der den erhöhten Untersuchungsaufwand rechtfertigt?“).

Dazu wurde bei Patienten mit NSCLC und prätherapeutischer FDG-PET/CT der LK-Uptake in der PET zu beiden Zeitpunkten visuell und semiquantitativ ausgewertet.

Da sich die Annahme einer besseren Performance durch die Anwendung der DTPI nicht bestätigte, aber der im Verlauf getestete visuelle Score vielversprechende Resultate zeigte, folgte darauf ein zweiter Studienabschnitt.

Ziel war es hier, die diagnostische Genauigkeit und das Interrater-Agreement unter Verwendung des PET-basierten visuellen Scores unter standardisierten Bedingungen (Bildfensterung und Befundungskriterien) an einem größeren Patientenkollektiv zu evaluieren („Ist der definierte visuelle Score auch reproduzierbar und robust genug für eine Anwendung im klinischen Setting?“).

4. Patienten, Material und Methoden

4.1 DTPI-Studie

4.1.1 Patienten

Der erste Teil dieser retrospektiven Studie beinhaltete 47 Patienten (weiblich, n = 17; männlich, n = 30; medianes Alter, 67 Jahre [Spannweite, 42 bis 84 Jahre]) mit 155 untersuchten LKs. Die mediane Anzahl an LKs lag bei 3 (Spannweite, 1 bis 6).

Die Studie beinhaltete folgende histologischen Subtypen: Adenokarzinom (AC; n = 23), Plattenepithelkarzinom (SCC; n = 19) und nicht anders charakterisierte NSCLC (NOS; n = 5). Patientencharakteristiken sind in Tabelle 4 zusammengefasst.

Patienten konnten für die Auswertung berücksichtigt werden, wenn folgende Kriterien erfüllt waren: 1. histologisch gesichertes NSCLC, 2. DTPI FDG-PET/CT-Untersuchung vor jeglicher antitumoröser Therapie durchgeführt, 3. histologische oder zytologische LK-Proben als *standard of reference* (SOR) vorhanden, 4. Zeitintervall zwischen PET-Untersuchung und Probenentnahme ≤ 8 Wochen, 5. erste PET-Akquisition (frühe PET) ≤ 90 min und Intervall zu später PET ≥ 45 min, 6. Blutglukosespiegel ≤ 150 mg/dl, 7. Primärtumor und LK können in den PET-Bildern eindeutig voneinander abgegrenzt werden.

Alle Vorgänge, die Personen involvierten, waren für beide Studien im Einklang mit den ethischen Standards der Otto-von-Guericke Universität Magdeburg (Interne Studien-ID: RAD271, Votum, 44/15) und den Prinzipien der Deklaration von Helsinki von 1964 und späteren Ergänzungen oder vergleichbaren ethischen Standards. Eine schriftliche Einwilligung zur wissenschaftlichen Datenauswertung lag für alle Patienten vor.

	DTPI-Studie (%)	Multireader-Analyse (%)
Gesamt	47 (100)	86 (100)
Geschlecht		
männlich	30 (64)	62 (72)
weiblich	17 (36)	24 (28)
Histologie		
Adenokarzinom	23 (49)	43 (50)
Plattenepithelkarzinom	19 (40)	37 (43)
NSCLC, NOS	5 (11)	6 (7)
T-Stadium (TNM)		
1	8 (17)	28 (33)
2	29 (62)	39 (45)
3	7 (15)	16 (19)
4	3 (6)	3 (3)
N-Stadium (TNM)		
0	24 (51)	47 (55)
1	10 (21)	15 (17)
2	11 (23)	17 (20)
3	2 (4)	7 (8)
UICC-Stadium (7. Edition)		
I	13 (28)	31 (36)
II	13 (28)	20 (23)
III	16 (34)	27 (31)
IV	5 (11)	8 (9)
Histologisches Grading		
1	0	2 (2)
2	14 (30)	31 (36)
3	28 (60)	46 (53)
4	2 (4)	3 (3)
Nicht näher angegeben	3 (6)	4 (5)

Tabelle 4: Patientencharakteristika

Angaben in ganzen Zahlen angegeben (Prozentwerte in Klammern). Abweichungen von 100% aufgrund von Rundung.

4.1.2 FDG-PET/CT

Für die Akquisition der Bilder wurde ein PET/CT-Gerät der Firma Siemens für beide Studien verwendet (Biograph mCT 64®; Siemens Healthcare, Erlangen, Deutschland). Zum Einsatz kam dabei der Tracer FDG gemäß Handlungsempfehlungen der European Association of Nuclear Medicine (EANM) (Boellaard et al. 2010). Um einen Blutglukosespiegel von unter <8,3 mmol/l sicherzustellen (Median, 5,8 mmol/l; Interquartilsabstand [IQR], 5.2 bis 6,3 mmol/l; Spannweite, 4,1 bis 8,3 mmol/l), war eine Blutuntersuchung für jeden Patienten verpflichtend. Es wurde F18-FDG mit einer medianen Aktivität von 236 MBq intravenös appliziert (IQR, 232 bis 238 MBq; Spannweite, 179 bis 245 MBq). Die frühe PET-Akquisition wurde nach einer

mittleren Uptake-Zeit von 64 min (IQR, 62 bis 68 min; Spannweite, 52 bis 90 min) durchgeführt, gefolgt von einer späten PET-Akquisition 125 min nach der Injektion (IQR, 117 bis 144 min; Spannweite, 102 bis 175 min). Das mediane Intervall zwischen den PET-Scans war 60 min (IQR, 52 bis 73 min; Spannweite, 45 bis 112 min). Die frühe PET erfolgte durch Ganzkörper-Aufnahmen von der Schädelbasis bis zum proximalen Femur innerhalb von 6 bis 8 Bettposition (Emission, 3 min. pro Bettposition) und einer Abdeckung in axialer Richtung von jeweils 216 mm (Siemens TrueV®; Überlappung der Bettpositionen, 89 mm). Die späte PET deckte den Thorax von der oberen Thoraxapertur bis zum Leberhilus innerhalb von zwei bis drei Bettpositionen ab. Ein low-dose-CT wurde für die Schwächungskorrektur und die anatomische Zuordnung verwendet (max. Röhrenstrom, 50 mAs; Röhrenspannung, 120 kV; Rotationszeit, 0,5 s/Rotation; Pitch, 0,8).

4.1.3 PET/CT Bilddatenrekonstruktion

Die PET-Rohdaten wurden durch iterative Rekonstruktion mit systemspezifischer point spread function (PSF) Modellierung und time-of-flight (TOF) Analyse (Siemens TrueX®, „HD-PET®“; Iterationen, 2; Subsets, 21) rekonstruiert. Die Projektionsdaten wurden mit einer Schichtdicke von 5 mm rekonstruiert (Reihen, 512; Spalten, 512; Voxelgröße 1,5 x 1,5 x 5,0 mm). Ein Gaußfilter (full width at half maximum [FWHM], 2 mm) als Postprocessing-Filter wurde nach der Rekonstruktion angewendet. Die Schwächungskorrektur der CT-Rohdaten erfolgte mit einer Schichtdicke von 5 mm und mit einem speziellen Filter für die low-dose-CT (convolution kernel, B19f).

Die Vorgehensweise war bei der Multireader-Analyse identisch.

4.1.4 Visuelle Auswertung der PET-Daten

Die visuelle Auswertung erfolgte durch einen erfahrenen Nuklearmediziner (9 Jahre Erfahrung) unter Verwendung des TrueD Viewers (Version, VE10C04) auf einem Siemens syngo® MutliModality Workplace (Software, syngoMMWP; Version, VE52A).

Der Befunder war vollständig verblindet in Bezug auf patientenspezifische Daten (bspw. Krankheitsgeschichte, körperliche Untersuchungen, Ergebnisse von ergänzenden bildgebenden Verfahren), Ergebnisse vom SOR, die finale Diagnose und das Patienten-Outcome.

Um eine einheitliche visuelle Auswertung zu gewährleisten, wurde die Maximumintensitätsprojektion (MIP) mit einem hohen (T) und einem niedrigen Schwellenwert (B) gefenstert. T wurde dabei als der doppelte SUVmean der Leber definiert (gemessen innerhalb einer dreidimensionalen ROI im rechten Leberlappen), während B auf null gestellt wurde. Anschließend wurde jeder LK zu beiden Zeitpunkten, basierend auf der MIP-

Darstellung, mit einem visuellen Score beurteilt (Abbildung 5). Unter Verwendung einer „inverse grey“ Skala für die Visualisierung der PET-Daten wurde das Punktesystem folgendermaßen definiert (Tabelle 5): (Score 1) LK-Uptake erscheint \leq MBPS, (Score 2) LK-Uptake erscheint $>$ Uptake des MBPS aber $<$ Uptake der Leber, (Score 3) Uptake der LK erscheint \geq Uptake der Leber, aber nicht „schwarz“, (Score 4) Uptake des LK erscheint „schwarz“. Zusätzlich zu letzterem, wenn der Uptake des Primarius selber nicht „schwarz“ bei der vorgeschlagenen Fensterungseinstellung war, sondern eher vergleichbar mit der Leber, wurden die korrespondierenden LKs mit einem Score von 4 gewertet, wenn sie zumindest vergleichbar mit dem Primarius waren, und mit einem Score von 1 bzw. 2, wenn ihr Uptake niedriger war als der der Leber. Die anatomische Korrelation zwischen PET-Befunden und histologischen/zytologischen Proben basierte auf der Klassifikation nach Mountain und Dresler für endobronchialen Ultraschall (EBUS) (Mountain und Dresler 1997).

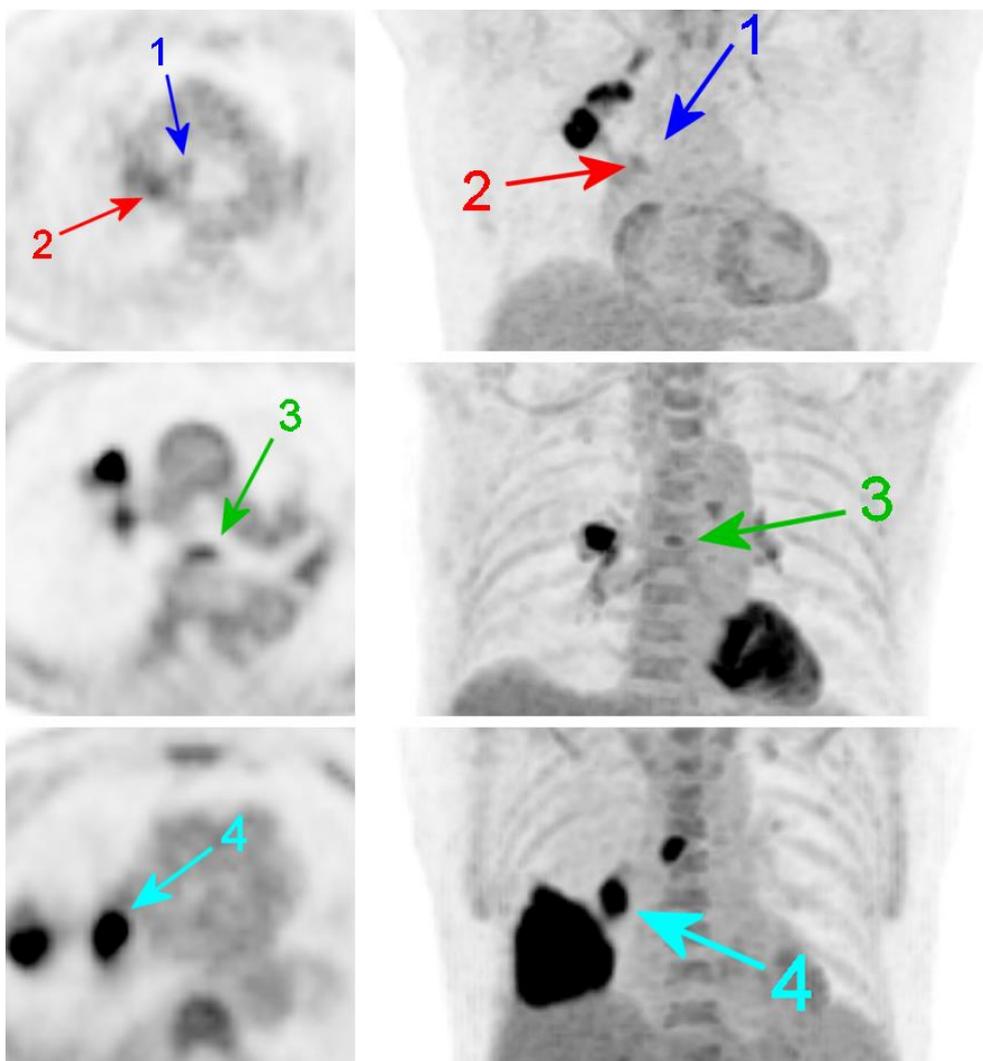


Abbildung 5: Repräsentative Beispiele des visuellen Scores

Axiale FDG-PET und zugehörige MIP-Darstellungen zeigen typische Beispiele für Score 1 und 2 (obere Reihe), Score 3 (mittlere Reihe) und Score 4 (untere Reihe).

Score	Kriterium
1	LK \leq MBPS
2	LK $>$ MBPS aber $<$ Leber
3	LK \geq Leber aber nicht „schwarz“ ^a
4	LK erscheint „schwarz“ ^{a,b}

Tabelle 5: Definition des visuellen Scores

^a „Schwarz“ unter Verwendung der „inverse grey“ Farbskala für die Visualisierung.

^b Wenn der Uptake des Primarius selber nicht „schwarz“ bei der vorgeschlagenen Fensterungseinstellung war, sondern eher vergleichbar mit der Leber, wurden die korrespondierenden LKs mit einem Score von 4 gewertet, wenn sie zumindest vergleichbar mit dem Primarius waren.

4.1.5 Semiquantitative Auswertung der PET-Daten

Alle LKs wurden sowohl in den frühen als auch den zum späten Zeitpunkt akquirierten PET-Datensätzen mit entsprechender Software abgegrenzt (ROVER®, version 2.1.4, ABX advanced biochemical compounds GmbH, Radeberg, Deutschland). Die initiale Abgrenzung war basierend auf einem relativen Schwellenwert bezogen auf die maximale Aktivität (ohne Hintergrundkorrektur) und wurde läsionsweise mithilfe entsprechender CT-Daten auf Plausibilität überprüft. Wenn erforderlich, wurden manuelle Anpassungen ausgeführt. Sowohl SUVmax als auch der SUVmean der dargestellten LKs wurden erfasst. Der RI zwischen früher und später PET wurde nach Shinya (Shinya et al. 2009) berechnet. SUVmax und SUVmean der Leber wurden innerhalb eines standardisierten sphärischen *Volume of Interest* (VOI; Diameter, 50 mm) innerhalb eines physiologischen Gewebeareals gemessen. Der SUVmax der MBPS wurde innerhalb eines standardisierten VOI im rechten Atrium (Diameter, 33 mm) ermittelt. Folgende SUV-Ratios wurden zwischen Läsionen und Referenzgewebe für frühe und späte PET berechnet:

$$\text{LK / Primarius} = \text{SUVmax (LK)} / \text{SUVmax (Primärtumor)}$$

$$\text{LK / Leber max} = \text{SUVmax (LK)} / \text{SUVmax (Leber)}$$

$$\text{LK / Leber mean} = \text{SUVmax (LK)} / \text{SUVmean (Leber)}$$

$$\text{LK / MBPS} = \text{SUVmax (LK)} / \text{SUVmax (MBPS)}$$

Die relative Änderung (Δ) zwischen früher und später PET wurde folgendermaßen wiedergegeben:

$$\Delta\text{Ratio (\%)} = \frac{\text{Ratio}(\text{spät}) - \text{Ratio}(\text{früh})}{\text{Ratio}(\text{früh})} * 100$$

4.1.6 Referenzstandard (SOR)

Histologische (n = 149; inklusive OP, n = 130; und Mediastinoskopie, n = 19) oder zytologische (TBNA; n = 6) Proben der LKs dienten als SOR. Proben von PET-negativen LKs wurden meist chirurgisch gewonnen (107 von 120 LKs; 89,2%) im Gegensatz zur Mediastinoskopie (11 von 120 LKs; 9,2%) oder TBNA (2 von 120 LKs; 1,7 %). Dreiundzwanzig (65,7 %) der 35 PET-positiven LKs wurden chirurgisch, 8 (22,8%) durch Mediastinoskopie und 4 (11,4%) durch TBNA verifiziert. Das mediane Zeitintervall zwischen PET/CT-Untersuchung und Akquisition von histologischen/zytologischen Proben betrug 19 Tage (IQR, 14 bis 28 Tage; Spannweite, 0 bis 50 Tage). Ein interdisziplinäres Gremium (Ärzte oder Experten auf dem Gebiet der Thoraxchirurgie, Pathologie, Nuklearmedizin und Radiologie) glichen im Konsensus prätherapeutische Befunde mit postoperativen und histopathologischen Ergebnissen unter Verwendung der Klassifikation von Mountain und Dresler ab (Mountain und Dresler 1997).

4.1.7 Statistik

Die Datenanalyse wurde mittels SPSS 22 (IBM Corporation, Armonk, NY, USA) vorgenommen. Entsprechend Histogrammen und Q-Q-Plots wurde eine nichtparametrische Verteilung der Daten angenommen. *Receiver-Operating-Characteristic*-Kurven (ROC) und entsprechende areas under the curve (AUC) mit 95%-Konfidenzintervallen (95%-CI) wurden für die visuelle und die semiquantitative Auswertung erstellt. Dieses beinhaltete den visuellen Score, die Änderung des visuellen Scores zwischen früher und später PET (höherer, gleicher oder geringerer Score), SUVmax und SUVmean, RI SUVmax und RI SUVmean (%), alle SUV-Ratios und die Δ -Ratios (%). Die optimalen Cutoff-Werte wurden definiert als der Punkt auf der ROC-Kurve mit minimaler Distanz d zum Punkt (0,1) und folgendermaßen berechnet:

$$d = \sqrt{(1 - \text{Sensitivität})^2 + (1 - \text{Spezifität})^2}$$

Resultierende Sensitivität, Spezifität, positiver prädikativer Wert (PPV), negativer prädikativer Wert (NPV) und Genauigkeit wurden bei diesem Cutoff-Wert mittels Standardformeln ermittelt. McNemar's Test für abhängige nominale Daten wurde durchgeführt, um zwei Parameter hinsichtlich ihrer Genauigkeit zu vergleichen. ROC-Kurven wurden entsprechend dem nichtparametrischen Ansatz von DeLong (DeLong et al. 1988) für korrelierende ROC-Kurven unter Verwendung entsprechender Software verglichen (MedCalc®, Cersion 15.2.2, 64-bit, MedCalc Software bvba, Ostende, Belgien).

In einem weiteren Ansatz wurde basierend auf der DTPI-Analyse eine Variable U folgendermaßen definiert:

$$U = \text{RI SUVmax (LK)} / \text{RI SUVmax (Primarius)}$$

PET-Positivität eines LKs wurde angenommen, wenn U zwischen 0,5 und 2 lag (wie vorgeschlagen von Uesaka (Uesaka et al. 2008)). Ein Streudiagramm diente der Darstellung der Verteilung von U in nicht-malignen und malignen LKs.

Um eine Beurteilung zwischen PET-positiven gegenüber PET-negativen LK zu ermöglichen, wurde das visuelle Bewertungssystem in einer Post-hoc-Analyse unter Verwendung verschiedener Cutoff-Werte dichotomisiert. Für p4 wurden die LKs nur positiv gewertet, wenn der visuelle Score >3 war. Für p3+4 waren LK bereits PET-positiv, wenn der Score >2 betrug. Statistische Signifikanz wurde bei einem P-Wert < 0,05 angenommen.

4.2 Multireader-Analyse

4.2.1 Patienten

Sechshundachtzig Patienten (weiblich, n = 24; männlich, n = 62; medianes Alter, 64 Jahre [Spannweite, 42 bis 84 Jahre]) mit 278 auswertbaren LK wurden analysiert. Die Kriterien, um für die Auswertung berücksichtigt zu werden, entsprachen der DTPI-Studie. Die Studie beinhaltete folgende histologische Subtypen: Adenokarzinome (AC; n = 43), Plattenepithelkarzinome (SCC; n = 37), und nicht anderweitig charakterisierte NSCLCs (NOS; n = 6). Patientencharakteristiken sind in Tabelle 4 zusammengefasst.

4.2.2 FDG-PET/CT

Die mediane Blutglukosekonzentration betrug 5,8 mmol/l (IQR, 5,3 bis 6,5 mmol/l; Spannweite, 4,1 bis 8,2 mmol/l). Verabreicht wurde FDG intravenös mit einer Aktivität von 235 MBq (IQR, 231 bis 238 MBq; Spannweite 179 bis 254 MBq). Die PET-Akquisition wurde nach einer medianen Uptakezeit von 64 min durchgeführt (IQR, 61 bis 68 min; Spannweite, 52 bis 90 min). PET-Datenakquisition und low-dose-CT erfolgten wie bei der DTPI-Studie.

4.2.3 Visuelle PET-Beurteilung

Die visuelle Befundung der PET-Datensätze erfolgte durch drei unabhängige Reader unter Verwendung der gleichen Software wie bei der DTPI-Studie. Die PET-Erfahrungswerte der drei Befunder waren folgende: Reader 1, 9 Jahre Erfahrung; Reader 2, 4 Jahre Erfahrung;

Reader 3, 3 Monate Erfahrung. Alle Reader waren vollständig verblindet gegenüber den Bewertungen der anderen Reader, den patientenspezifischen Daten (bspw. Krankheitsgeschichte, körperliche Untersuchungen, Ergebnisse von ergänzenden bildgebenden Verfahren), den Ergebnissen der Auswertung mittels SOR, der finalen Diagnose und dem Patientenoutcome. Die Datensätze wurden nur visuell analysiert. Die Bewertungen von jedem Reader wurden in einem standardisierten Antwortbogen festgehalten.

Die Fensterung der MIP und die Kriterien der Punkteskala waren die gleichen wie bei der DTPI-Studie und sind unter 4.1.4 beschrieben.

Um alle Reader mit dem zu testenden Score vertraut zu machen, erhielten alle ein Training in der Beurteilung. Das Training beinhaltete Datensätze von 20 Patienten, die durch alle Reader unabhängig voneinander bewertet wurden und nicht Teil der finalen Analyse waren. Den Readern wurde nach Abschluss der Befundung jedes Trainingsdatensatzes Feedback bezüglich der Ergebnisse des SOR gegeben und der vergebene Score jeweils mit den definierten Kriterien abgeglichen.

4.2.4 SOR

Histologische (n = 271; inklusive OP, n = 236; und Mediastinoskopie, n = 35) oder zytologische (n = 7, TBNA) Proben der LKs dienten als SOR. Proben von PET-negativen LKs wurden meist chirurgisch gewonnen (204 von 224 LKs; 91,1%) im Gegensatz zur Mediastinoskopie (19 von 224 LKs; 8,5%) oder TBNA (1 von 224 LKs; 0,4%). Zweiunddreißig (59,3%) der 54 PET-positiven LKs wurden chirurgisch, 16 (29,6%) durch Mediastinoskopie und 6 (11,1%) durch TBNA verifiziert. Das mediane Zeitintervall zwischen PET/CT-Untersuchung und Akquisition von histologischen/zytologischen Proben betrug 20 Tage (IQR, 15 bis 29 Tage; Spannweite, 0 bis 50 Tage). Ein Abgleich der prätherapeutischen Befunde mit den histopathologischen Ergebnissen erfolgte wie in der DTPI-Studie.

4.2.5 Statistik

Die Datenanalyse erfolgte im Wesentlichen wie bei der DTPI-Studie. Zusätzlich kam das R-Plug-in für SPSS (Version, 2.1.5; R Vereinigung für statistisches Rechnen, Wien, Österreich; <http://www.R-project.org>; 2015) zum Einsatz.

McNemar's Test für abhängige nominale Daten wurde durchgeführt, um zwei Reader hinsichtlich ihrer Genauigkeit zu vergleichen. Die Analyse der diagnostischen Genauigkeit beinhaltete außerdem die Einteilung der Patienten in Untergruppen. Letztere umfassten hiläre LKs (Regionen 10-12 [Mountain und Dresler 1997]) vs. mediastinale LKs (Regionen 2, 4, 5, 7 und 8), LKs von AC vs. SCC, oder LKs von Primärtumoren vom Grading G1/G2 vs. G3/G4.

Das Interrater-Agreement von zwei Readern wurde mittels Cohen's Kappa (κ) mit linearen Wichtungen nach Dichotomisierung der Bewertungen in PET-positive (Score >3) vs. PET-negative Befunde (Score ≤ 3) überprüft. Limits of Agreement (LoA) wurden mit den κ -Benchmarks von Landis und Koch (Landis und Koch 1977) bestimmt. Die Übereinstimmung zwischen allen Reader wurde mit Fleiss' κ (ungewichtet) überprüft. Entsprechende 95%-CI wurden unter Annahme einer Standardnormalverteilung errechnet (Anzahl der Fälle >20). Statistische Signifikanz wurde bei einem P-Wert < 0,05 angenommen.

5. Ergebnisse

5.1 DTPI-Studie

Die Verteilung der 155 zu beurteilenden LKs entsprechend dem Stagingssystem von Mountain und Dresler (Mountain und Dresler 1997) war wie folgt: EBUS 2 (n = 3), 4 (n = 37), 5 (n = 1), 7 (n = 36), 10 (n = 40), 11 (n = 26) und 12 (n = 12). Sechsenddreißig der LKs (23,3%) waren maligne (Tabelle 6). Fälle mit nur marginal erhöhtem Uptake des Primarius (weniger als die Leber) wurden nicht beobachtet. Bei 4 Patienten (10 LKs betreffend) war der Uptake des Primarius ähnlich intensiv wie in der Leber. Die restlichen Patienten hatten einen Primarius mit einer höheren Uptakeintensität. Abbildung 6 stellt die ROC-Kurven für die visuelle und semiquantitative Auswertung dar. Abbildung 7 dient als Beispiel für eine korrekte visuelle, aber inkorrekte semiquantitative Beurteilung. Abbildung 8 zeigt ein Beispiel für eine falsch positive visuelle Auswertung.

Region	Gesamt (n = 155)	Maligne (n = 36)	Benigne (n = 119)
2	3	0	3
4	37	9	28
5	1	1	0
7	36	5	31
10	40	9	31
11	26	9	17
12	12	3	9

Tabelle 6: LK-Region nach Mountain und Dresler (Mountain und Dresler 1997)

LK-Regionen aufgeteilt in maligne und nicht-maligne LKs

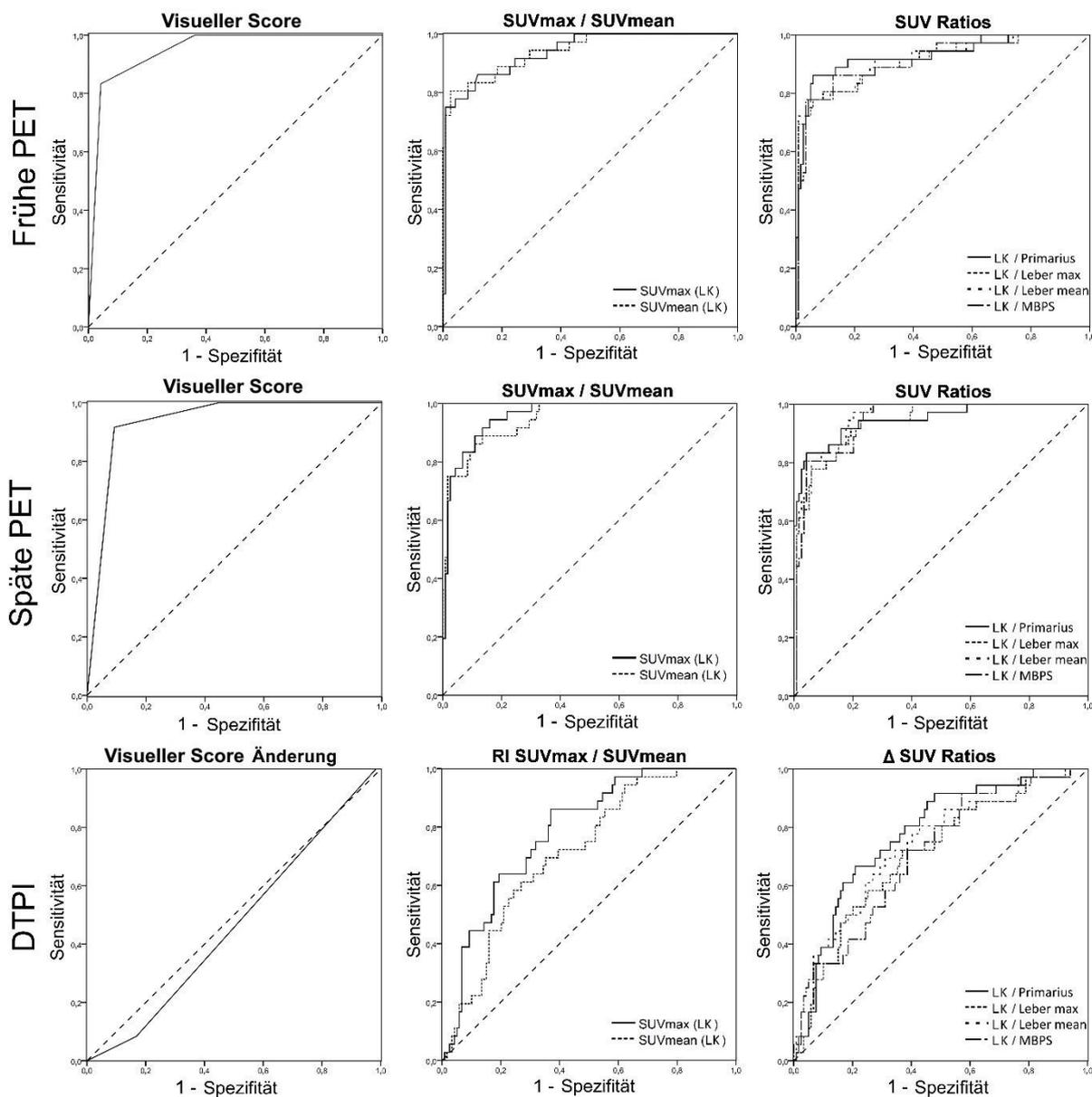


Abbildung 6: ROC-Kurven für die visuelle und semiquantitative Auswertung

ROC-Kurven sind für frühe PET (**obere Reihe**), späte PET (**mittlere Reihe**) und DTPI (**untere Reihe**) bzw. visuelle Auswertung, SUVmax / SUVmean und SUV-Ratios abgebildet.

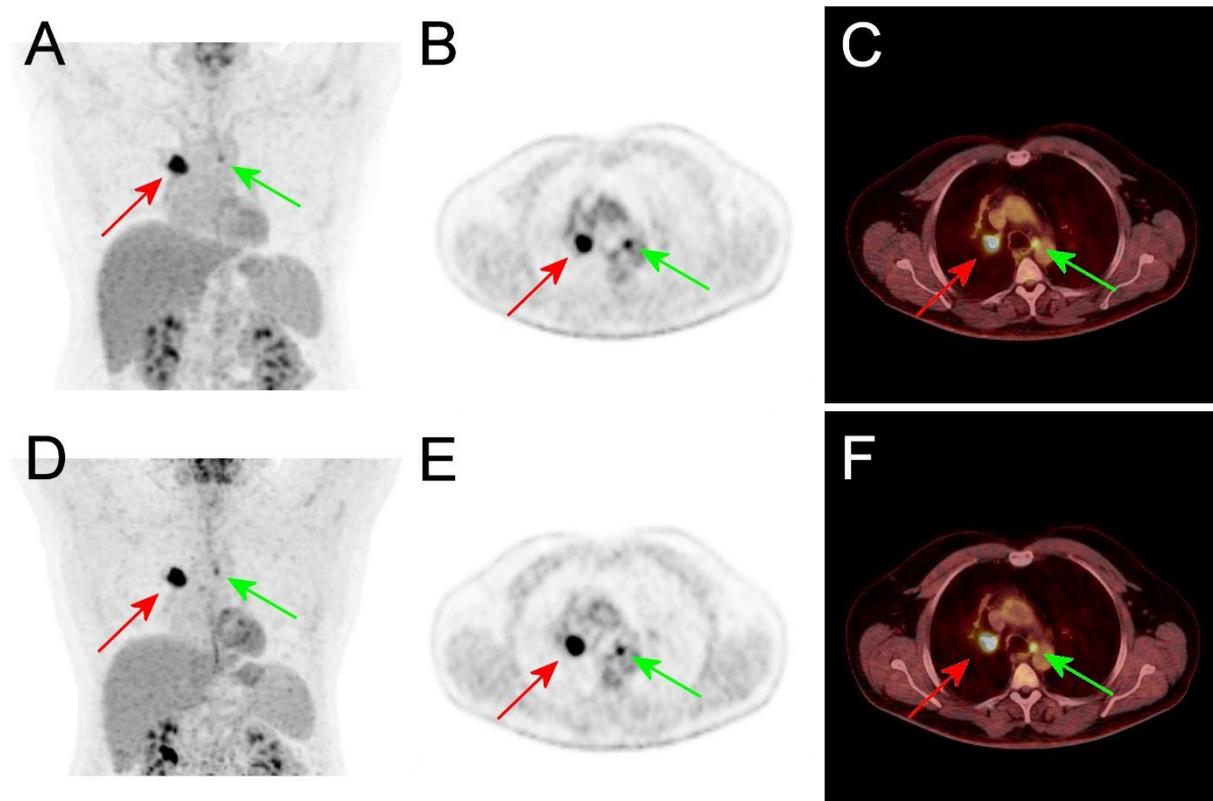


Abbildung 7: Patientenbeispiel einer korrekten visuellen aber inkorrekten semiquantitativen Beurteilung

Diese Abbildung zeigt einen 49 Jahre alten männlichen Patienten, der an einem SCC im oberen rechten Lungenlappen (rote Pfeile) leidet. Für die frühe PET (**A, B, C**) zeigt die MIP-Darstellung (**A**) erhöhten FDG-Uptake eines kontralateralen LKs der Station 4 L nach Mountain und Dresler (Mountain und Dresler 1997) (grüne Pfeile) und wurde mit einem Score von 3 in der visuellen Auswertung bewertet. Bei der späten PET (**D, E, F**) erscheint die Uptakeintensität leicht erhöht, verglichen mit der frühen PET in der MIP-Darstellung (**D**) aufgrund des verringerten Uptake des Hintergrundes, wurde aber weiterhin mit einem Score von 3 bewertet. SUVmax (früh, 6,0; spät, 7,5) und SUVmean (3,8; 5,0) sowie ihre RIs (+24,2% / +31,5%) waren über den korrespondierenden Cutoff-Werten. Die postoperative histologische Aufarbeitung zeigte benigne Veränderungen des LK.

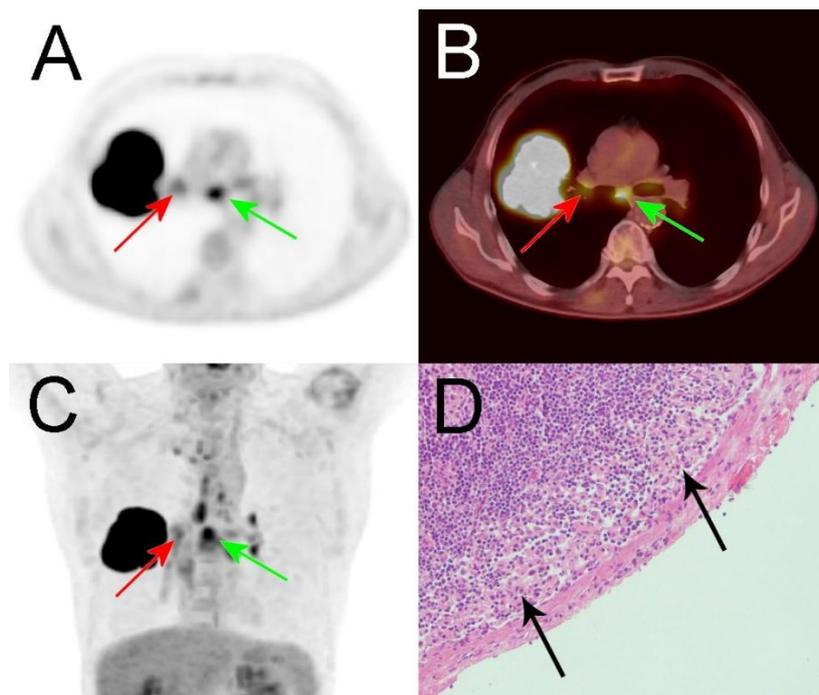


Abbildung 8: Beispiel einer falsch positiven visuellen Auswertung

Axiales PET (A), überlagerte PET/CT-Bilder (B) und MIP-Datei (C) für die frühe PET bei einem 56-jährigen männlichen Patienten, der an einem Lungenkarzinom (NSCLC, NOS) im oberen rechten Lungenlappen leidet. Die PET zeigte einen erhöhten fokalen FDG-Uptake von mehreren ipsilateralen und kontralateralen LKs. Unter anderem wurde LK-Station 7 nach Mountain und Dresler (Mountain und Dresler 1997) (grüne Pfeile) mit Score 4 bewertet bei der Auswertung der MIP-Abbildung (C) während Station 10R (roter Pfeil) mit Score 3 bewertet wurde. Die postoperative histologische Aufarbeitung zeigte für beide LK-Stationen, Station 7 und 10R, benigne LKs mit Sinushistiozytose (D).

5.1.1 Frühe PET

5.1.1.1 Visuelle Auswertung

Für die frühe PET wurde ein Score von 1 für 13 LKs (8,4%) erteilt, ein Score von 2 für 63 LKs (40,6%), Score 3 für 44 LKs (28,4%) und ein Score von 4 wurde an 35 LKs (22,6%) erteilt. Eine ROC-Analyse (Abbildung 6, obere Reihe; Tabelle 8) ergab einen optimalen visuellen Cutoff für einen Score >3 (AUC, 0,95; 95%-CI, 0,91 bis 0,98). Entsprechende Sensitivität, Spezifität, PPV, NPV und Genauigkeit waren 83,3%, 95,9%, 85,7%, 95,0% und 92,9% (Tabelle 7, Abbildung 7).

Frühe PET	RP	RN	FP	FN	Sensitivität		Spezifität		PPV		NPV		Genauigkeit		P
					%	95%-CI	%	95%-CI	%	95%-CI	%	95%-CI	%	95%-CI	
<i>Visuell</i>															
p4	30	114	5	6	83,3	67,2 bis 93,6	95,8	90,5 bis 98,6	85,7	69,7 bis 95,2	95,0	89,4 bis 98,1	92,9	87,7 bis 96,4	
p3+4	36	76	43	0	100	90,3 bis 100	63,9	54,6 bis 72,5	45,6	34,3 bis 57,2	100	95,3 bis 100	72,3	64,5 bis 79,1	<0,01
<i>SUV (LK)</i>															
SUVmax	31	105	14	5	86,1	70,5 bis 95,3	88,2	81,1 bis 93,4	68,9	53,4 bis 81,8	95,5	89,7 bis 98,5	87,7	81,5 bis 92,5	0,02
SUVmean	30	109	10	6	83,3	67,2 bis 93,6	91,6	85,1 bis 95,9	75,0	58,8 bis 87,3	94,8	89,0 bis 98,1	89,7	83,8 bis 94,0	0,18
<i>SUV Ratios</i>															
LK / Primarius	31	112	7	5	86,1	70,5 bis 95,3	94,1	88,3 bis 97,6	81,6	65,7 bis 92,3	95,7	90,3 bis 98,6	92,3	86,9 bis 95,9	1,0
LK / Leber max	29	108	11	7	80,6	64,0 bis 91,8	90,8	84,1 bis 95,3	72,5	56,1 bis 85,4	93,9	87,9 bis 97,5	88,4	82,3 bis 93,0	0,02
LK / Leber mean	29	105	14	7	80,6	64,0 bis 91,8	88,2	81,1 bis 93,4	67,4	51,5 bis 80,9	93,8	87,6 bis 97,5	86,5	80,0 bis 91,4	<0,01
LK / MBPS	31	104	15	5	86,1	70,5 bis 95,3	87,4	80,1 bis 92,8	67,4	52,0 bis 80,5	95,4	89,6 bis 98,5	87,1	80,8 bis 91,9	0,01
RP, richtig positiv; RN, richtig negativ; FP, falsch positiv; FN, falsch negativ															

Tabelle 7 Diagnostische Performance der visuellen und semiquantitativen Auswertung für die frühe PET

Die diagnostische Performance der visuellen und semiquantitativen Auswertung für die frühe PET wird angezeigt. Die Genauigkeit von p4 wird mit den verbleibenden Parametern verglichen (McNemar-Test). Signifikante Ergebnisse fett gedruckt.

Parameter	Frühe PET			Späte PET			DTPI		
	AUC (95%-CI)	Cutoff- Wert	<i>P</i>	AUC (95%-CI)	Cutoff- Wert	<i>P</i>	AUC (95%-CI)	Cutoff- Wert	<i>P</i>
Visuell	0,95 (0,91 bis 0,98)	> 3		0,94 (0,9 bis 0,98)	> 3		0,47 (0,36 bis 0,57)	≥ <i>stabil</i>	
SUV (LK)									
SUVmax	0,94 (0,9 bis 0,98)	> 4,8	0,46	0,96 (0,93 bis 0,99)	> 6,0	0,06	0,79 (0,71 bis 0,86)	> 19,4%	<0,001
SUVmean	0,94 (0,9 bis 0,99)	> 3,6	0,72	0,95 (0,91 bis 0,98)	> 3,5	0,47	0,71 (0,63 bis 0,8)	> 10,6%	<0,01
SUV Ratios									
LK / Primarius	0,93 (0,87 bis 0,99)	> 0,4	0,39	0,95 (0,9 bis 0,99)	> 0,4	0,52	0,78 (0,69 bis 0,86)	> -0,3%	<0,001
LK / Leber max	0,91 (0,85 bis 0,97)	> 1,4	0,08	0,94 (0,9 bis 0,98)	> 1,5	0,98	0,71 (0,61 bis 0,8)	> 21,7%	<0,01
LK / Leber mean	0,92 (0,86 bis 0,97)	> 1,7	0,11	0,95 (0,92 bis 0,98)	> 2,7	0,24	0,74 (0,64 bis 0,83)	> 38,9%	<0,01
LK / MBPS	0,91 (0,86 bis 0,97)	> 1,8	0,06	0,94 (0,91 bis 0,98)	> 3,4	0,61	0,71 (0,62 bis 0,8)	> 38,8%	<0,01

Tabelle 8: ROC-Analyse: AUCs und optimale Cutoff-Werte

Die AUCs mit den zugehörigen 95%-CIs und den optimalen Cutoff-Werten werden für alle untersuchten Parameter für die frühe und späte PET sowie die DTPI angezeigt. Darüber hinaus werden die abgeleiteten *P*-Werte für den Vergleich der ROC-Kurven des visuellen Scores versus jede einzelne semiquantitative Messung angegeben (DeLong et al. 1988). Signifikante Ergebnisse fett gedruckt.

5.1.1.2 Semiquantitative Auswertung

Bei der frühen PET war der SUVmax in malignen LKs (Median 11,4; IQR 6,3 bis 16,3; Spannweite, 3,2 bis 24,8) signifikant höher ($P < 0,001$) als in non-malignen LKs (Median, 3,0; IQR 2,2 bis 3,9; Spannweite, 0,8 bis 19,1). Selbiges wurde für den SUVmean in malignen LKs (Median, 6,5; IQR, 4,1 bis 7,8; Spannweite, 2,3 bis 13,3) und benignen LKs beobachtet (Median, 2,3; IQR, 1,8 bis 2,9; Spannweite, 0,7 bis 5,3; $P < 0,001$).

Die ROC-Auswertung (Abbildung 6b, obere Reihe; Tabelle 8) für SUVmax und SUVmean zeigte optimale Cutoff-Werte von $>4,8$ / $>3,6$ sowie AUCs von 0,94 (95%-CI, 0,9 bis 0,98) und 0,94 (95%-CI, 0,9 bis 0,99). Alle Ergebnisse, einschließlich der SUV-Ratios, sind in Tabelle 7 aufgelistet.

5.1.2 Späte PET

5.1.2.1 Visuelle Auswertung

Bei der späten PET wurde ein Score von 1 für 11 LKs vergeben (7,1%), ein Score von 2 bei 55 LKs (35,5%), ein Score von 3 bei 45 (29,0%) und ein Score von 4 bei 44 LKs (28,4%). Die ROC-Auswertung (Abbildung 6, mittlere Reihe; Tabelle 8) zeigte einen optimalen Cutoff-Score > 3 (AUC, 0,94; 95%-CI, 0,9 bis 0,98). Entsprechende Sensitivität, Spezifität, PPV, NPV und Genauigkeit waren 91,7%, 90,8%, 75,0%, 97,3% und 91,0% (Tabelle 9).

Späte PET	RP	RN	FP	FN	Sensitivität		Spezifität		PPV		NPV		Genauigkeit		P
					%	95%-CI	%	95%-CI	%	95%-CI	%	95%-CI	%	95%-CI	
Visuell															
p4	33	108	11	3	91,7	77,5 bis 98,3	90,8	84,1 bis 95,3	75,0	59,7 bis 86,8	97,3	92,3 bis 99,4	91,0	85,3 bis 95,0	
p3+4	36	66	53	0	100	90,3 bis 100	55,5	46,1 bis 64,6	40,4	30,2 bis 51,4	100	94,6 bis 100	65,8	57,8 bis 73,2	<0,01
SUV (LK)															
SUVmax	32	106	13	4	88,9	73,9 bis 96,9	89,1	82,0 bis 94,1	71,1	55,7 bis 83,6	96,4	91,0 bis 99,0	89,0	83,0 bis 93,5	0,45
SUVmean	32	103	16	4	88,9	73,9 bis 96,9	86,6	79,1 bis 92,1	66,7	51,6 bis 79,6	96,3	90,7 bis 99,0	87,1	80,8 bis 91,9	0,18
SUV Ratios															
LK / Primarius	30	114	5	6	83,3	67,2 bis 93,6	95,8	90,5 bis 98,6	85,7	69,7 bis 95,2	95,0	89,4 bis 98,1	92,9	87,7 bis 96,4	0,61
LK / Leber max	31	101	18	5	86,1	70,5 bis 95,3	84,9	77,2 bis 90,8	63,3	48,3 bis 76,6	95,3	89,3 bis 98,5	85,2	78,6 bis 90,4	0,049
LK / Leber mean	30	108	11	6	83,3	67,2 bis 93,6	90,8	84,1 bis 95,3	73,2	57,1 bis 85,8	94,7	88,9 bis 98,0	89,0	83,0 bis 93,5	0,58
LK / MBPS	29	114	5	7	80,6	64,0 bis 91,8	95,8	90,5 bis 98,6	85,3	68,9 bis 95,1	94,2	88,4 bis 97,6	92,3	86,9 bis 95,9	0,75
RP, richtig positiv; RN, richtig negativ; FP, falsch positiv; FN, falsch negativ															

Tabelle 9: Diagnostische Performance der visuellen und semiquantitativen Auswertung für die späte PET

Die diagnostische Performance der visuellen und semiquantitativen Auswertung für die späte PET wird angezeigt. Die Genauigkeit von p4 wird mit den verbleibenden Parametern verglichen (McNemar-Test). Signifikante Ergebnisse fett gedruckt.

5.1.2.2 Semiquantitative Auswertung

Bei der späten PET war der SUVmax in malignen LKs (Median, 14,6; IQR, 7,9 bis 22; Spannweite, 4,1 bis 37,3) signifikant höher verglichen mit benignen LKs (Median, 3,3; IQR, 2,3 bis 4,5; Spannweite, 1,0 bis 23,9; $P < 0,001$). Selbiges wurde für den SUVmean in malignen LKs (Median, 6,8; IQR, 4,4 bis 10,6; Spannweite, 2,9 bis 17,1) und non-maligne LKs erhoben (Median, 2,4; IQR, 1,7 bis 3,1; Spannweite, 0,7 bis 8,6; $P < 0,001$).

Die ROC-Analysen (Abbildung 6, mittlere Reihe; Tabelle 8) für SUVmax und SUVmean zeigten einen optimalen Cutoff SUV von $>6,0$ / $>3,5$ und AUCs von 0,96 (95%-CI, 0,93 bis 0,99) oder 0,95 (95%-CI, 0,91 bis 0,98). Alle Ergebnisse, inklusive der SUV-Ratios, sind in Tabelle 9 dargestellt.

5.1.3 DTPI

5.1.3.1 Visuelle Auswertung

Änderungen des visuellen Scores von früher zu später PET wurden in 25 LKs beobachtet (16,1% aller LKs, Score 1 zu 2, $n = 2$; Score 2 zu 3, $n = 12$; Score 3 zu 4, $n = 9$; Score 3 zu 2, $n = 2$). Unter der Annahme von PET-Positivität bei einem Score >3 wurde eine Änderung des N-Stadiums bei 6 Patienten beobachtet (inkorrekt, $n = 4$, korrekt, $n = 2$). Entsprechend ergab eine Änderung des visuellen Scores zwischen früher und später PET eine AUC von 0,47 (95%-CI, 0,36 bis 0,57; Abbildung 6, untere Reihe; Tabelle 8 und 10), die in einer signifikant niedrigeren Genauigkeit des visuellen Scores mit DTPI im Vergleich zur frühen PET resultierte ($P < 0,001$).

5.1.3.2 Semiquantitative Auswertung

Der RI SUVmax in malignen LKs (Median, +31,6%; IQR, +20,6 bis +47,7%; Spannweite, 1,7 bis +116,5%) war signifikant höher als in nicht malignen LKs (Median, +13,3%; IQR, -2,4 bis +24,6%; Spannweite, -40,6 bis 161,5%; $P < 0,001$). Selbiges wurde beobachtet für RI SUVmean in malignen LKs (Median, +20,5%; IQR, +3,9 bis +30,5%; Spannweite, -9,8 bis +75,0%) und benignen LKs (Median, +4,6%; IQR, -6,4 bis 16,9%; Spannweite, -42,4 bis +113,8%; $P < 0,001$).

Die ROC-Analysen (Abbildung 6, untere Reihe; Tabelle 8) für RI SUVmax und RI SUVmean zeigten optimale Cutoff Werte von $>+19,4\%$ / $>+10,6\%$ sowie AUCs von 0,79 (95%-CI, 0,71 bis 0,86) und 0,71 (95%-CI, 0,63 bis 0,8). Alle Ergebnisse, inklusive der SUV-Ratios, sind in Tabelle 10 zu finden. Eine signifikant niedrigere Genauigkeit der semiquantitativen Messungen wurde für die DTPI verglichen mit der frühen PET beobachtet (jeweils $P < 0,001$).

Die Definition von PET-positiv als U zwischen 0,5 und 2 erreichte eine Sensitivität, Spezifität und Genauigkeit von 66,7%, 57,1% und 59,4% (Abbildung 9).

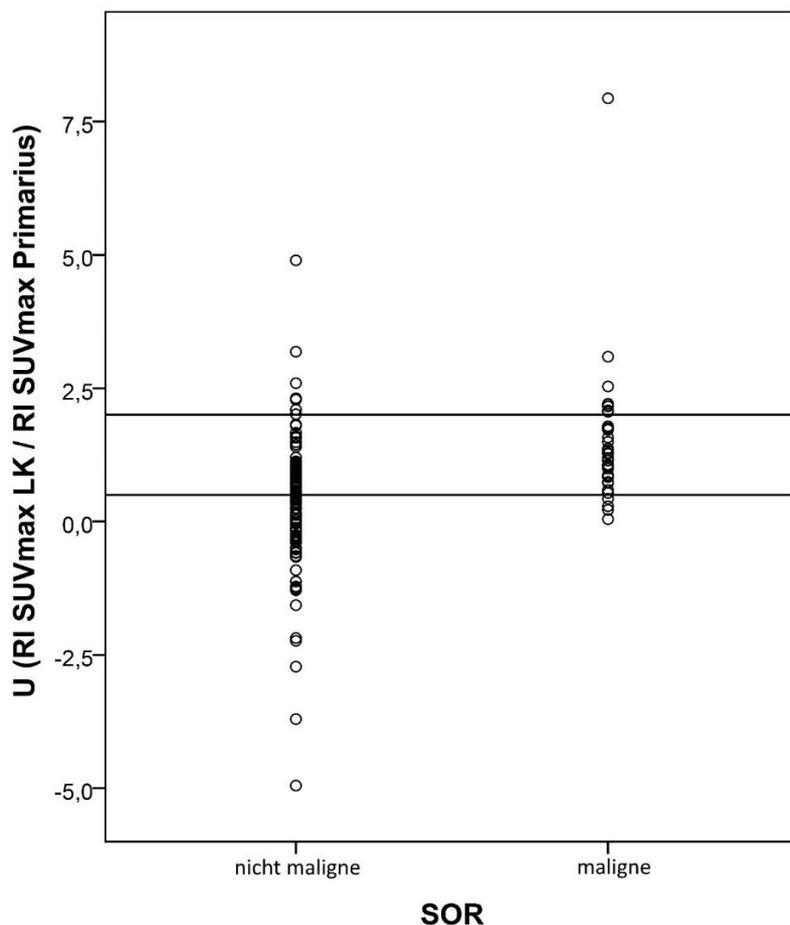


Abbildung 9: Streudiagramm von U in malignen und nicht malignen LKs

Die Verteilung von U ist durch den SOR in nicht maligne und maligne LKs aufgeteilt. Die horizontalen Linien repräsentieren ein U von 0,5 oder 2. Wie von Uesaka et al. vorgeschlagen (Uesaka et al. 2008), sind alle LKs dazwischen als PET-positiv definiert, allerdings zeigt das Diagramm eine zu hohe Überlappung zwischen beiden LK-Gruppen für eine verlässliche Unterscheidung.

DTPI	RP	RN	FP	FN	Sensitivität		Spezifität		PPV		NPV		Genauigkeit		P*	P**
					%	95%-CI	%	95%-CI	%	95%-CI	%	95%-CI	%	95%-CI		
Visuell																
Scoreänderung	36	2	117	0	100	90,3 bis 100	1,7	0,2 bis 5,9	23,5	17,1 bis 31,1	100	15,8 bis 100	24,5	18,0 bis 32,1		<0,01
RI (LK)																
SUVmax	31	75	44	5	86,1	70,5 bis 95,3	63,0	53,7 bis 71,7	41,3	30,1 bis 53,3	93,8	86,0 bis 97,9	68,4	60,4 bis 75,6	<0,01	
SUVmean	25	77	42	11	69,4	51,9 bis 83,7	64,7	55,4 bis 73,2	37,3	25,8 bis 50,0	87,5	78,7 bis 93,6	65,8	57,8 bis 73,2	<0,01	0,61
Δ Ratios																
LK / Primarius	24	94	25	12	66,7	49,0 bis 81,4	79,0	70,6 bis 85,9	49,0	34,4 bis 63,7	88,7	81,1 bis 94,0	76,1	68,6 bis 82,6	<0,01	0,05
LK / Leber max	26	75	44	10	72,2	54,8 bis 85,8	63,0	53,7 bis 71,7	37,1	25,9 bis 49,5	88,2	79,4 bis 94,2	65,2	57,1 bis 72,6	<0,01	0,41
LK / Leber mean	25	82	37	11	69,4	51,9 bis 83,7	68,9	59,8 bis 77,1	40,3	28,1 bis 53,6	88,2	79,8 bis 94,0	69,0	61,1 bis 76,2	<0,01	1,0
LK / MBPS	26	73	46	10	72,2	54,8 bis 85,8	61,3	52,0 bis 70,1	36,1	25,1 bis 48,3	88,0	79,0 bis 94,1	63,9	55,8 bis 71,4	<0,01	0,27
RP, richtig positiv; RN, richtig negativ; FP, falsch positiv; FN, falsch negativ																

Tabelle 10: Diagnostische Performance der visuellen und semiquantitativen Auswertung für die DTPI

Die diagnostische Performance der DTPI basierend auf der visuellen und semiquantitativen Auswertung wird angezeigt. Die Genauigkeit beruhend auf einer Änderung des visuellen Scores (P^*) oder RI SUVmax (P^{**}) wird mit den verbleibenden Parametern verglichen (McNemar-Test). Signifikante Ergebnisse fett gedruckt.

5.2 Multireader-Analyse

Insgesamt waren 278 LKs auswertbar (mediane LKs pro Patient, $n = 3$; Spannweite, 1 bis 6). Die Verteilung der LKs nach dem Stagingssystem von Mountain und Dresler (Mountain und Dresler 1997) war die folgende: EBUS 1 ($n = 2$), 2 ($n = 10$), 4 ($n = 60$), 5 ($n = 6$), 7 ($n = 54$), 10 ($n = 77$), 11 ($n = 40$) und 12 ($n = 29$). Vierundfünfzig von allen auswertbaren LKs (19,4 %) waren maligne.

5.2.1 Diagnostische Performance

Die höchste Gesamtgenauigkeit wurde durch Reader 3 erreicht, die folgenden Ergebnisse zeigte: Reader 3 ordnete 73 LKs (26,3 %) Score 1 zu, 91 LKs (32,7 %) Score 2, 48 (17,3 %) Score 3 und 66 LKs (23,7 %) Score 4 zu. Die ROC-Analyse (Abbildung 10) zeigte eine optimale diagnostische Genauigkeit bei einem visuellen Cutoff Score >3 , um maligne von non-malignen LKs zu unterscheiden (AUC, 0,94; 95%-CI, 0,91 bis 0,98). Sensitivität, Spezifität, PPV, NPV und Genauigkeit betragen bei diesem Cutoff Score 88,9%, 94,6%, 80,0%, 97,2% und 93,5% (Tabelle 11). Reader 1 und 2 erreichten ähnliche AUCs von jeweils 0,94 (jedes $P > 0,05$) sowie eine vergleichbare Genauigkeit von 92,1% bzw. 91,4% (jeweils $P > 0,05$). Wenn als Cutoff ein Score >2 gewählt wurde, waren die Sensitivität, Spezifität, PPV, NPV und Genauigkeit von Reader 3 96%, 58,3%, 45,6%, 98,8% und 77,0%.

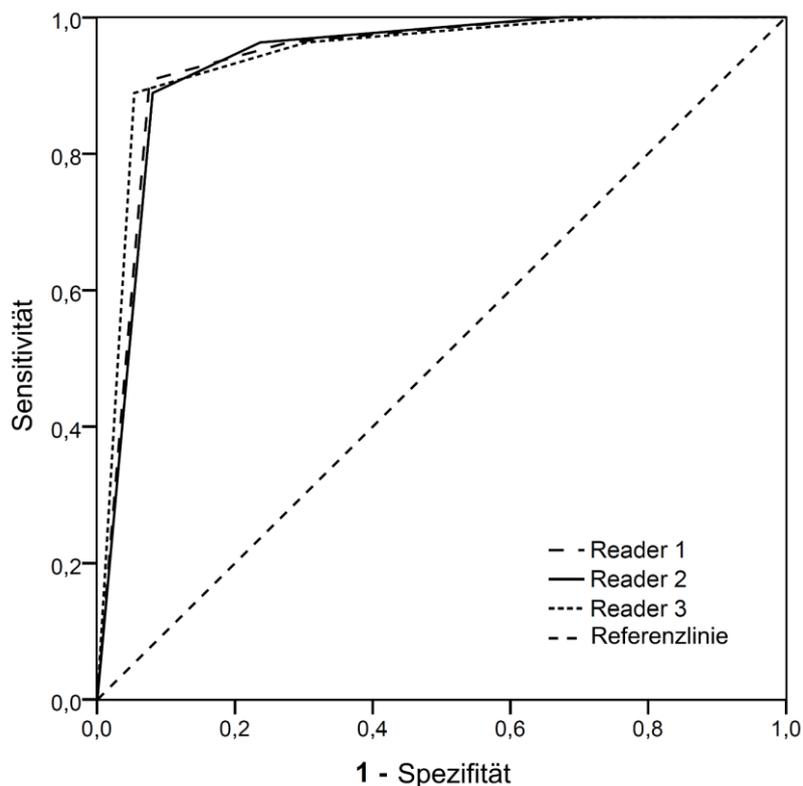


Abbildung 10: ROC-Kurven aller drei Reader für den visuellen Score

Reader	RP	RN	FP	FN	Sensitivität		Spezifität		PPV		NPV		Genauigkeit	
					%	95%-CI	%	95%-CI	%	95%-CI	%	95%-CI	%	95%-CI
1	49	207	17	5	90,7	79,7 bis 96,9	92,4	88,1 bis 95,5	74,2	62,0 bis 84,2	97,6	94,6 bis 99,2	92,1	88,3 bis 95,0
2	48	206	18	6	88,9	77,4 bis 95,8	92,0	87,6 bis 95,2	72,7	60,4 bis 83,0	97,2	93,9 bis 99,0	91,4	87,4 bis 94,4
3	48	212	12	6	88,9	77,4 bis 95,8	94,6	90,8 bis 97,2	80,0	67,7 bis 89,2	97,2	94,1 bis 99,0	93,5	90,0 bis 96,1

RP, richtig positiv; RN, richtig negativ; FP, falsch positiv; FN, falsch negativ

Tabelle 11: ROC-Analyse und diagnostische Performance für den visuellen Score

Korrekte und inkorrekte Bewertungen mit den korrespondierenden diagnostischen Werten werden für jeden Reader angezeigt.

5.2.2 Trennung nach anatomischen oder histologischen Charakteristika

Die ROC-Kurven für alle Reader sind getrennt nach anatomischen/histologischen Kriterien in Abbildung 11 dargestellt und mit ihren entsprechenden AUCs in Tabelle 12 zusammengefasst. Die AUCs der Reader 1, 2 und 3 waren nicht signifikant unterschiedlich; weder in den hilären LKs (jeweils 0,93; jeweils $P > 0,05$) noch in den mediastinalen LKs (0,95, 0,94 und 0,96; jeweils $P > 0,05$). Bei Patienten mit AC betrug die AUCs von Reader 1, 2 und 3 0,94, 0,94 und 0,93 (jedes $P > 0,05$), während LKs von SCC AUCs von 0,95, 0,94 und 0,95 zeigten (jeweils $P > 0,05$). Die AUCs von Reader 1, 2 und 3 bei LKs von Primärtumoren mit Grading G1/G2 waren 0,93, 0,92 und 0,91 (jeweils $P > 0,05$). Bei Grad G3/G4-Tumoren waren die AUCs 0,94, 0,95 und 0,96 für Reader 1, 2 bzw. 3 (jeweils $P > 0,05$).

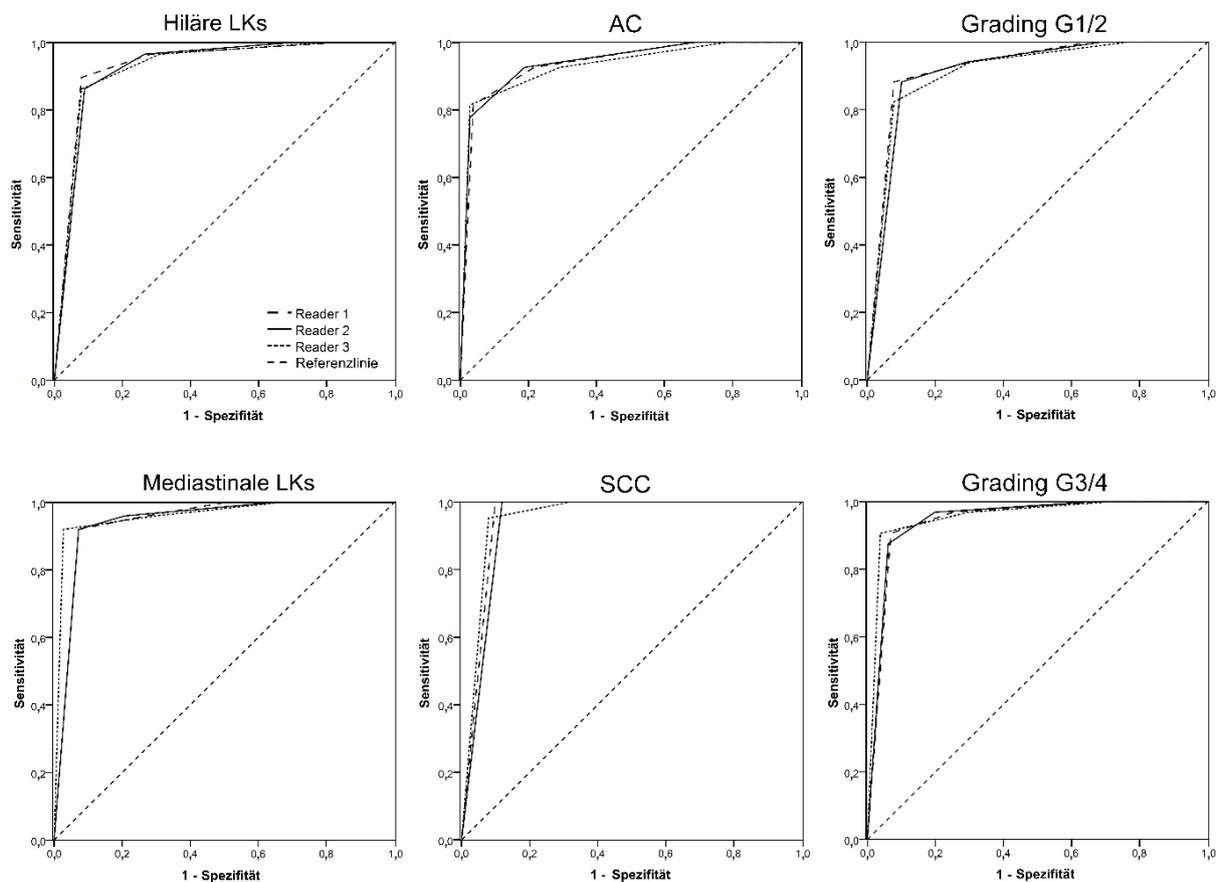


Abbildung 11: ROC-Kurven in Abhängigkeit der anatomischen oder histologischen Eigenschaften

Die ROC-Kurven der drei Reader werden nach der Aufteilung basierend auf der anatomischen Lage (linke Spalte), der Histologie des Primarius (mittlere Spalte) und dem histologischen Grading (rechte Spalte) angezeigt.

Subgruppen	AUC (95%-CI)		
	Reader 1	Reader 2	Reader 3
LK-Region			
hilär	0,93 (0,88 bis 0,98)	0,93 (0,88 bis 0,97)	0,93 (0,88 bis 0,98)
mediastinal	0,95 (0,9 bis 0,99)	0,94 (0,9 bis 0,99)	0,96 (0,92 bis 1,0)
Histologie			
AC	0,94 (0,89 bis 0,99)	0,94 (0,89 bis 0,99)	0,93 (0,87 bis 0,99)
SCC	0,95 (0,91 bis 0,99)	0,94 (0,9 bis 0,98)	0,95 (0,92 bis 0,99)
Grading			
G1/G2	0,93 (0,86 bis 0,99)	0,92 (0,85 bis 0,98)	0,91 (0,84 bis 0,99)
G3/G4	0,94 (0,9 bis 0,98)	0,95 (0,91 bis 0,99)	0,96 (0,92 bis 1,0)

Tabelle 12: AUCs aller Reader für die anatomischen/histologischen Subgruppen.

AUCs mit ihren 95%-CI für alle drei Reader werden, aufgeteilt nach ihrer LK-Region, histologischen Subtypen oder Grading angezeigt.

5.2.3 Interrater-Agreement

Übereinstimmende Bewertungen (PET-positiv vs. PET-negativ) von Reader 1 und 2 wurden in 274 von 278 LKs (98,6%) beobachtet, die zu einem Cohen's κ von 0,96 (95%-CI, 0,92 bis 1,0 [LoA, *almost perfect*]) führten. Die Vergleiche von Reader 1 und 3 sowie Reader 2 und 3 zeigten jeweils 270 (97,1%) übereinstimmend gewertete LKs mit einem vergleichbaren κ von 0,92 (95%-CI, 0,86 bis 0,97 [LoA, *almost perfect*]).

Insgesamt wurden 268 LKs (96,4%) übereinstimmend durch alle drei Reader gewertet, wodurch sich ein Fleiss' κ von 0,93 (95%-CI, 0,87 bis 1,0 [LoA, *almost perfect*]) ergab.

6. Diskussion

Nach der Diagnose des NSCLC muss eine prätherapeutische Beurteilung der thorakalen LKs erfolgen, um eine optimale Behandlung des Patienten zu gewährleisten (De Leyn et al. 2014, Harders et al. 2014). Die PET/CT hat dabei in den letzten Jahrzehnten eine immer größere Rolle beim Staging eingenommen (Gould et al. 2003, Volpi et al. 2018). Dennoch kann die PET/CT beim Stagingprozess häufig ein invasives Vorgehen nicht ersetzen und sogar ein häufigeres und weitreichenderes invasives Staging verursachen (Gunluoglu et al. 2011).

Ursächlich hierfür ist die geringe bzw. variable Genauigkeit der PET/CT (Gunluoglu et al. 2011, Schmidt-Hansen 2007). Angetrieben durch die Entwicklung neuer PET-Scanner mit höherer Auflösung, verbunden mit neuster Bildrekonstruktion, zielen Forschungsbemühungen daher darauf ab, die diagnostische Performance der PET/CT mittels DTPI zu verbessern (Bryant et al. 2006, Uesaka et al. 2008, Shinya et al. 2009). Ziel dieser Arbeit war es daher im ersten Auswertungsabschnitt, die Auswirkung der DTPI mit ihren zwei Zeitpunkten von früher und später PET auf die diagnostische Genauigkeit für das prätherapeutische LK-Staging bei NSCLC zu evaluieren. Angewendet wurden zu diesem Zwecke semiquantitative und visuelle Auswertungsmethoden.

6.1 DTPI-Studie

6.1.1 Semiquantitative Auswertung

Die höchste Genauigkeit von 76,1% für die DTPI-Auswertung erreichte die Ratio SUVmax (LK) / SUVmax (Primarius). Diese Genauigkeit war tendenziell höher als die der üblicheren einfachen RI SUVmax-Messungen (68,4%, $P = 0,05$). Semiquantitative Messungen zeigten signifikant höhere Genauigkeiten als die visuelle Befundung bei der DTPI-Auswertung. Insgesamt aber waren die erreichten Genauigkeiten der DTPI den Einzelzeitpunktmessungen von früher und später PET unterlegen. Insbesondere zeigte sich eine signifikant überlegene Genauigkeit der in der klinischen Routine eingesetzten frühen PET gegenüber der DTPI ($P < 0,001$). Diese Beobachtung wird von Shinya et al. gestützt, die 98 thorakale LKs von Patienten mit benignen und malignen Lungenrundherden untersuchten. Für die frühe PET erreichten sie unter Verwendung des SUVmax bei einem Cutoff von 3,6 eine Sensitivität und Spezifität von 87% und 88% sowie 92% und 93% für die späte PET bei einem Cutoff von 4,0. Im Gegensatz dazu zeigte der RI SUVmax bei einem Cutoff von +20,9% nur eine Sensitivität von 74% und eine Spezifität von 76% (Shinya et al. 2009).

Diese Ergebnisse stehen im Kontrast zu Uesaka et al., die 580 LKs in einer prospektiven Studie auswerteten, um maligne von nicht-malignen LKs zu unterscheiden. Die DTPI erreichte dort eine fast perfekte Genauigkeit. Als PET-positiv definierten die Autoren, in Analogie zum

Parameter U in der vorliegenden Studie, wenn das Verhältnis RI SUVmax (LK) und RI SUVmax (Primarius) zwischen 0,5 und 2 lag. Die entsprechende Sensitivität, Spezifität und Genauigkeit lag bei 98%, 93% und 97% (Uesaka et al. 2008). Bei Übertragung dieses Ansatzes auf die aktuelle Patientengruppe ergab sich eine Sensitivität, Spezifität und Genauigkeit von nur 66,7%, 57,1% und 59,4% aufgrund der großen Überschneidung von U bei malignen und nicht malignen LKs (Abbildung 9).

Dieses Ergebnis verdeutlicht, dass obwohl semiquantitative Cutoff Werte grundsätzlich in der Lage sind hohe Genauigkeiten beim thorakalen LK-Staging zu erreichen, sie jedoch nur unter gut kontrollierten Bedingungen anwendbar sind (Langen et al. 1993, Zasadny und Wahl 1993, Boellard et al. 2004, Jaskowiak et al. 2005, Akamatsu et al. 2014).

Der Einfluss von unterschiedlichen Rekonstruktionseinstellungen auf die semiquantitative Inter-time-point-Analyse bleibt, ebenso wie für die Einzelzeitpunktmessung, unklar. In der Literatur fanden sich für das LK-Staging mittels DTPI PET bei NSCLC für die frühe PET optimale SUV-Cutoffs zwischen 3,0 und 3,6 bzw. zwischen 3,3 und 4,0 für die späte PET. Die Cutoffs für beide Zeitpunkte waren damit niedriger als in der aktuellen Studie (frühe PET, 4,8; späte PET, 6,0). Der optimale Cutoff RI für die vorliegende Kohorte hingegen lag mit 19,4% innerhalb der in der Literatur gefundenen Spannbreite von 4,3 bis 20,9% (Shinya et al. 2009, Suga et al. 2009, Kim et al. 2012).

Die Ergebnisse deuten darauf hin, dass PSF- und TOF-basierte Rekonstruktionsverfahren (welche auch aktuell zum Einsatz kamen) keinen deutlichen Einfluss auf die DTPI-Analyse haben, da sie den SUV von malignen LKs zu beiden Zeitpunkten gleichermaßen erhöhen.

Beim Vergleich der semiquantitativen Einzelzeitpunktmessungen in der vorliegenden Studie zeigte sich, ähnlich der DTPI-Analyse, dass die späte PET-Befundung keine verbesserte Genauigkeit gegenüber der frühen PET erbrachte. Diese Beobachtung wird durch eine Metaanalyse von Shen et. al mit LKs von 654 Patienten mit NSCLC gestützt. Dort zeigten sich ähnliche Sensitivitäten (84% vs. 84%) und Spezifitäten (83% vs. 89%; $P > 0,05$) wie in der aktuellen Studie für die Einzelzeitpunktmessung und die DTPI (Shen et al. 2015). Die SUV-Ratios zwischen LKs und unterschiedlichen Referenzregionen zeigten weder für die frühe noch die späte PET überlegene Genauigkeit gegenüber LK SUVmax oder SUVmean (jeweils $P > 0,05$). Die diagnostische Performance war vergleichbar mit den Ergebnissen von Nguyen et al. (Nguyen et al. 2015). Ausnahme bildet hier LK / Primarius, der im aktuellen Patientenkollektiv eine Genauigkeit von 92% verglichen mit 74% aufwies.

6.1.2 Visuelle Auswertung

Die erhoffte bessere Genauigkeit durch die Anwendung der DTPI für das LK-Staging konnte nicht bestätigt werden, allerdings erreichte die visuelle Befundung über einen vierstufigen

Score in unserer Patientenkohorte eine hohe Genauigkeit für die frühe und späte PET. Hingegen erbrachte eine Score-Änderung von früher zu später PET eine geringe Genauigkeit, so dass durch die Anwendung der DTPI auch für die visuelle Auswertung kein zusätzlicher Nutzen für das LK-Staging entsteht.

Durch die Festlegung des visuellen Cutoff-Scores > 3 als PET-positiv (p4) für die frühe PET wurde eine hohe Sensitivität von 83,3% und eine sehr hohe Spezifität von 95,8% erreicht.

Bei der Anwendung eines niedrigeren Cutoffs (p3+4; Score > 2 ist PET-positiv) konnte sogar eine Sensitivität von 100% erreicht werden, die jedoch zu Lasten der Spezifität mit 63,9% ging, woraus sich eine Genauigkeit von 72,3% ergab.

Die Konstellation dieser Parameter ist klinisch relevant, da nach den ESTS-Leitlinien bei PET-positiven LKs oder hoher Wahrscheinlichkeit für Metastasen ein zusätzliches invasives Staging erfolgen soll (De Leyn et al. 2014). Auf die beiden eben genannten Cutoffs übertragen, wird die Anzahl der unnötigen Operationen maßgeblich vom NPV der FDG-PET/CT beeinflusst, während die Rate der vermeidbaren invasiven Stagingeingriffe vom PPV abhängt. Hierbei erreicht der Cutoff p3+4 zwar einen NPV von 100% und kann damit unnötige chirurgische Eingriffe verhindern, jedoch steigt möglicherweise durch den niedrigen PPV von 45,6% die Anzahl der vermeidbaren, invasiven Stagingprozeduren. Auf der anderen Seite könnte der p4 Cutoff durch einen NPV von 95% nicht nur chirurgische Eingriffe reduzieren, sondern mit einem PPV von 85,7% auch weniger unnötige invasive Stagingeingriffe bewirken.

6.1.3 Limitationen in Abhängigkeit von der Histologie

Bei der frühen PET wurden in der DTPI-Studie unter Anwendung des p4 Cutoffs 5 falsch positive LKs (Längsachsendurchmesser 12 bis 30 mm) festgestellt. Davon waren 4 LKs von Sinus-Histiozytose betroffen (Abbildung 8), und ein LK wies in erheblichem Maße Phagozytose von anthrakotischem Pigment auf. In der Literatur sind ähnliche Fälle erhöhter FDG-Aufnahme in benignen LKs aufgrund der eben genannten Pathologien beschrieben, die die falsch positiven Befunde erklären könnten (Yeh et al. 2009, Shim et al. 2005).

Auf der anderen Seite waren von den 6 falsch negativen LKs (Durchmesser 12 bis 22 mm) 5 LKs von Patienten mit ACs (muzinöse, mikropapilläre und partiell sarkomatoide entdifferenzierte papilläre AC) und 1 LK von einem Patienten mit einem schlecht differenzierten SCC. Auch hier finden sich Beispiele in der Literatur, speziell für den muzinösen Subtyp des AC, die eine niedrigere Sensitivität bei der LK-Beurteilung von ACs beschreiben (Schmidt-Hansen et al. 2014, Wang et al. 2012, Shim und Han 2010).

In solchen Fällen ist offenbar auch der visuelle Score in seiner diagnostischen Verlässlichkeit limitiert. Die Anwendbarkeit des Scores ist zudem auf Fälle mit einem mindestens ebenso hohen Uptake des Primärtumors wie der Leberhintergrund beschränkt.

6.2 Standardisierte visuelle Befundung

Es war unser Anliegen im zweiten Studienabschnitt, auf Grundlage der sehr guten Ergebnisse der visuellen Auswertung in der DTPI-Studie die Genauigkeit des 4-stufigen visuellen Scores mit einer geeigneten Patientenpopulation näher zu validieren und die Reproduzierbarkeit anhand des Interrater-Agreements von Readern mit unterschiedlicher Befundungserfahrung zu überprüfen. Außerdem wurde der Einfluss der anatomischen Lokalisation der LKs und der histologischen Eigenschaften des Primarius auf die diagnostische Genauigkeit untersucht.

6.2.1 Validierung des visuellen Scores

Reader 1-3 erreichten bei einem Cutoff Score von >3 (p_4) sehr gute Genauigkeiten von 92,1%, 91,4% und 93,5%. Diese Ergebnisse stehen im Einklang mit der Genauigkeit von 92,9% aus der visuellen Auswertung der DTPI-Studie. Alle drei Reader waren damit trotz unterschiedlicher Erfahrung in der klinischen PET-Befundung (gering, mittel, hoch) vergleichbar gut.

Ein Cochrane-Review von Schmidt-Hansen et al. (Schmidt-Hansen et al. 2014) gibt für die untersuchten Studien Sensitivitäten und Spezifitäten von 17 bis 100% bzw. 33 bis 100% für die FDG-PET im LK-Staging des NSCLC an. Die Kriterien für die PET-Positivität waren dabei unterschiedlich: „LK-Uptake $>$ Hintergrund“, „LK SUVmax $\geq 2,5$ “, „gemischte Kriterien“. Im Besonderen wiesen Studien zur visuellen LK-Beurteilung mit 82,3% bis 86,5% geringere Genauigkeiten aus als die aktuellen Auswertungen (Hellwig et al. 2007, Yen et al. 2008, Yang et al. 2008). Es muss dabei jedoch berücksichtigt werden, dass bspw. unterschiedliche technische Voraussetzung einen Vergleich erschweren.

Entscheidend für eine gute Beurteilungsmethode ist nicht nur ihre Genauigkeit, sondern auch ihre Reproduzierbarkeit in der klinischen Routine. Die notwendigen Kriterien dafür sind zweierlei: Gut definierte, allgemein akzeptierte, einfache Bewertungskriterien und höchste Objektivität während des Befundungsprozesses. Mit einem Fleiss' κ von 0,93 zeigte der 4-stufige Score bei den 3 Readern mit ihren unterschiedlichen Erfahrungswerten einen hohen Grad an Reproduzierbarkeit.

Analog zu der visuellen Auswertung in der DTPI-Studie wurde der Score zweigeteilt, um PET-positive Befunde (Score >3) von PET-negativen LK zu trennen (Score ≤ 3). Durch dieses Vorgehen wird die PET-Positivität alleine darüber bestimmt, ob die Uptakeintensität des LKs genauso intensiv erscheint wie die des Primarius. Typischerweise erscheint der Primarius „schwarz“, wenn die allgemein übliche „inverse grey“-Farbskala verwendet wird (Abbildung 5). Interessanterweise erwiesen sich LKs, die einen Score 3 erhielten, 3,3-mal so häufig histologisch/zytologisch als maligne wie LKs, die nur einen Score 2 erhielten, und kein LK mit

dem Score 1 wies Malignität auf. Es ist daher empfehlenswert für einen aussagekräftigeren Befund, auch LKs mit einem Score ≤ 3 weiter zu differenzieren.

6.2.2 Interrater-Agreement

Hofman et al. (Hofman et al. 2009) untersuchten das Interrater-Agreement zwischen drei Readern für die visuelle LK-Auswertung von 100 Patienten mit NSCLC. Die Fensterung erfolgte über einen festgelegten oberen Schwellwert von SUV = 10 unter Verwendung einer Grauskala. PET-Positivität wurde generell angenommen, wenn der Uptake in den LKs größer war als der in den MBPS, benigne dagegen, wenn der Uptake gleich bzw. niedriger war als der der MBPS. Es durften allerdings bei unklaren Fällen beispielsweise CT-Befunde für morphologische Fragestellungen wie Größe und Lage von Gewebe zu Hilfe genommen werden. Das Interrater-Agreement für das N-Stadium erreichte ein Cohen's κ von 0,75 bis 0,81. Schlechter schnitt das Agreement beim Vergleich von aortopulmonalen oder hilären LKs mit κ von 0,48 bis 0,71 versus mediastinalen LKs mit κ von 0,71 bis 0,88 ab. Derartige Unterschiede zwischen den LK-Lokalisationen bestätigten sich in der aktuellen Auswertung nicht. Die AUC der drei Befunder lag innerhalb der Subgruppe hilärer LK jeweils bei 0,93 und damit unwesentlich niedriger als bei den mediastinalen LKs mit AUCs zwischen 0,94 und 0,96. Das Interrater-Agreement von unerfahrenen und erfahrenen PET-Readern verglichen Smulders et al. (Smulders et al. 2007) mit der Befundung durch Experten, welche als SOR fungierten. Dabei wurden 30 Patienten mit NSCLC hinsichtlich des N-Stagings untersucht, spezifische Bewertungskriterien wurden nicht genannt. Es zeigte sich dabei eine merklich höhere Übereinstimmung zwischen erfahrenen Readern und den Experten mit $\kappa = 0,72$ als zwischen unerfahrenen Readern und Experten mit $\kappa = 0,58$. Insgesamt war die Übereinstimmung mit $\kappa = 0,65$ jedoch niedrig. In der aktuellen Multireader-Analyse war die Übereinstimmung zwischen dem unerfahrenen Reader 3 und dem erfahrensten Reader 1 mit $\kappa = 0,92$ geringer als zwischen Reader 3 und dem erfahreneren Reader 2 mit $\kappa = 0,96$. Insgesamt war aber die Übereinstimmung mit Fleiss' $\kappa = 0,93$ hoch, und das etwas höhere Agreement zwischen Reader 1 und 2 ist maßgeblich durch eine höhere Rate richtiger Befunde des Readers 3 bedingt (Tabelle 11). Ausgehend von diesen überzeugenden Ergebnissen für den unerfahrenen Reader hat der untersuchte visuelle Score unter Anwendung der standardisierten Kriterien erhebliches Potential Lernprozesse zu beschleunigen, während eine hohe diagnostische Genauigkeit gewahrt werden kann und insbesondere zweideutige Befunde vermieden werden können

6.2.3 Standardisierung

Eine geeignete und standardisierte Fensterung ist von Nöten, um einen hohen Grad an Objektivität zu ermöglichen und damit das Interrater-Agreement zu verbessern. Die von uns gewählte Fensterung „das Doppelte des SUVmean der Leber“ ist zu nahezu 100% reproduzierbar. Außerdem berücksichtigt diese Einstellung variierende Störfaktoren wie unterschiedliche Uptakeintensitäten des Primarius oder die Anwesenheit anderer hochintensiver Anreicherungen wie bspw. im Nierenbecken, der Harnblase oder im Hirngewebe. Dies steht im Gegensatz zu rigideren Fensterungseinstellungen wie „50% der maximalen Aktivität“. Ein Schwellwert relativ zur Leber und zum Primarius wahrt im Gegensatz zu starren Einstellungen wie „SUV = 10“ (Hofman et al. 2009) eine Unabhängigkeit von Parametern wie dem BMI, dem Plasmaglukosespiegel und dem Zeitpunkt nach Injektion (Zasadny und Wahl 1993, Hamberg et al. 1994).

Eine höhere Objektivität in der Dignitätsbeurteilung von LKs könnte prinzipiell, wie oben bereits erläutert, durch eine semiquantitative Auswertung erreicht werden. Diese ist jedoch von SUV-Cutoffs abhängig, für welche in der Literatur eine erhebliche Streubreite des SUVmax von 2,5 bis 5,2 berichtet wird und die damit einer ausreichenden Standardisierung im Wege steht (Bryant et al. 2006, Mountain und Dresler 1997, Hwangbo et al. 2009).

Weitere Vorteile bieten standardisierte Befundungen, neben der hohen diagnostischen Genauigkeit, durch die Möglichkeit, multizentrische Studien ohne eine zeitaufwendige zentrale Referenzbefundung realisieren zu können.

Angesichts der unterschiedlichen Fachdisziplinen, die bei onkologischen Patienten in der klinischen Routine in Diagnostik und Therapie involviert sind, vereinfachen standardisierte Befundungskriterien zudem die interdisziplinäre Kommunikation.

Wang et al. (Wang et al. 2015) zeigten bei Patienten mit NSCLC, die einer definitiven Strahlenbehandlung zugeführt wurden, dass auf Grundlage standardisierter visueller Kriterien für die Beurteilung des Ansprechens auf diese Therapie Patienten nach ihrem progressionsfreien Überleben und Gesamtüberleben unterschieden werden konnten (Mac Manus et al. 2013). Die SUV-basierte Befundung erreichte allerdings nur ein geringes Interrater-Agreement mit $\kappa = 0,43$. Dieses Ergebnis verdeutlicht, wie wichtig einheitliche Befundungskriterien sind.

Die Zuverlässigkeit der standardisierten visuellen PET-Befundung könnte durch unterschiedliche PET-Scanner (Schmidt-Hansen et al. 2014), Scannergenerationen und Szintillatorkristalle sowie Bildrekonstruktionseinstellungen beeinträchtigt sein (Lasnon et al. 2012, Hashimoto et al. 2018). In der Literatur werden Steigerungen des SUVmax in LKs von 43 bis 48% durch die Verwendung von PSF+TOF verglichen mit dem konventionellen ordered subset expectation maximization (OSEM) berichtet (Lasnon et al. 2012, Akamatsu et al. 2014). Weitergehende Untersuchungen sind daher notwendig, um die Anwendbarkeit des

Befundungsscores auch unter anderen Scannervoraussetzungen und Rekonstruktionseinstellungen zu evaluieren.

6.3 Limitationen

Limitationen zeigten beide Studien hinsichtlich ihres retrospektiven Charakters. Ein möglicher Selektionsbias ergibt sich aufgrund der oben genannten Einschlusskriterien (Punkt 4.1.1) und des hohen Anteils chirurgisch gewonnener Proben für den SOR (84% bzw. 85%). Dies ermöglichte zwar einen zuverlässigen SOR, impliziert aber, dass diese Patienten durch invasive bzw. non-invasive Methoden als operabel eingestuft wurden. Erwähnenswert ist außerdem, dass die diagnostische Genauigkeit aufgrund des geringen Anteils maligner LKs (23 bzw. 19%) hauptsächlich durch die Spezifität bestimmt wird.

6.4 Schlussfolgerung

Weder späte PET noch DTPI erbrachten einen Zugewinn an diagnostischer Genauigkeit für das thorakale LK-Staging beim NSCLC. Hingegen erbrachte die Anwendung des visuellen Scores bei standardisierter Fensterung sehr gute Ergebnisse und hat das Potential, die Wertigkeit der FDG-PET/CT im Stagingprozess zu verbessern, die Durchführung und Vergleichbarkeit von Studien zu erleichtern und Lernprozesse für unerfahrene Befunder zu beschleunigen.

7. Zusammenfassung

Das Hauptziel dieser Dissertation war die Optimierung der prätherapeutischen FDG-PET-basierten Dignitätsbeurteilung thorakaler LKs beim NSCLC. Das erste Teilprojekt evaluierte dafür die DTPI, verschiedene semiquantitative Parameter sowie einen visuellen Score. Dabei wurden 47 Patienten mit 155 LKs, die eine FDG-PET mit DTPI durchlaufen hatten (PET-Akquisitionszeitpunkte nach ca. 60 und 120 min), retrospektiv analysiert. Der Uptake der LKs wurde sowohl über einen definierten 4-stufigen visuellen Score bei einheitlicher Fensterung als auch semiquantitativ mit SUVmax, SUVmean und weiteren SUV-Ratios beurteilt. Die DTPI-Analyse beinhaltete RIs, Δ -Ratios und Änderungen im visuellen Score. Histologie oder Zytologie dienten als SOR. Die Genauigkeit wurde über eine ROC-Analyse bestimmt. Die Genauigkeit des DTPI war für alle Messungen niedrig (visuelle Auswertung, 24,5%; RI SUVmax, 68,4%; RI SUVmean, 65,8%; Δ -Ratios, 63,9 bis 76,1%) und signifikant niedriger gegenüber der frühen PET. Die Genauigkeit der frühen (Spannweite, 86,5 bis 92,9%) und späten PET (Spannweite, 85,2 bis 92,9%) waren vergleichbar. Bei der frühen PET war die Genauigkeit der visuellen Analyse (92,9%) vergleichbar der semiquantitativen Auswertung oder ihr überlegen (Spannweite, 86,5 bis 92,3%). Die Anwendung der späten PET oder der DTPI ergaben keinen Vorteil, jedoch erbrachte der angewendete visuelle Score gute Ergebnisse.

Auf diesen Erkenntnissen beruhte der zweite Auswertungsabschnitt, der zur retrospektiven Validierung des visuellen Scores an 86 Patienten mit 278 LKs durch drei unterschiedlich erfahrene, verblindete Reader diente, sowie zur Überprüfung des Interrater-Agreements bei Anwendung dieses Scores. Übereinstimmung wurde mittels Cohen's κ und Fleiss' κ überprüft. Die Reader erreichten ähnliche Genauigkeiten von 93,5%, 91,4% und 92,1%. Cohen's κ betrug zwischen 0,92 und 0,96 (Fleiss' κ , 0,93). ROC-Analysen zeigten keine signifikanten Unterschiede zwischen den 3 Readern.

Weder späte PET noch DTPI erbrachten einen Zugewinn an diagnostischer Genauigkeit für das LK-Staging beim NSCLC. Die Anwendung des visuellen Scores bei standardisierter Fensterung hingegen erlaubte eine hohe Genauigkeit und hat das Potential den Stellenwert der FDG-PET/CT im Stagingprozess zu verbessern, die Durchführung und Vergleichbarkeit von Studien zu erleichtern und Lernprozesse für unerfahrene Befunder zu beschleunigen.

8. Literaturverzeichnis

a) Zeitschriften/Lehrbücher/Links

1. Adams MC, Turkington TG, Wilson JM, Wong TZ: A systematic review of the factors affecting accuracy of SUV measurements. *AJR Am J Roentgenol.* 195(2): 310-20 (2010)
2. Alberg AJ, Brock MV, Ford JG, Samet JM, Spivack SD: Epidemiology of Lung Cancer Diagnosis and Management of Lung Cancer, 3rd ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest.* 143(5 Suppl): e1S–e29S (2013)
3. Alkadhi H, Leschka S, Stolzmann P, Scheffel H: *Wie funktioniert die CT?* Springer. Berlin Heidelberg 2011.
4. 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. *Eur J Radiol.* 83(1): 226-30 (2014)
5. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF): S3-Leitlinie Prävention, Diagnostik, Therapie und Nachsorge des Lungenkarzinoms. Langversion 1.0, 2018, AWMF-Registernummer: 020/007OL. URL: https://www.leitlinienprogramm-onkologie.de/fileadmin/user_upload/Downloads/Leitlinien/Lungenkarzinom/LL_Lunge_nkarzinom_Langversion_1.0.pdf (letzter Zugriff am: 23.05.2018)
6. Basu S, Alavi A: Partial volume correction of standardized uptake values and the dual time point in FDG-PET imaging: should these be routinely employed in assessing patients with cancer? *Eur J Nucl Med Mol Imaging.* 34(10): 1527-9 (2007)
7. Boellaard: Optimisation and harmonisation: two sides of the same coin? *Eur J Nucl Med Mol Imaging.* 40(7): 982-4 (2013)
8. Boellaard R, Krak NC, Hoekstra OS, Lammertsma AA: Effects of noise, image resolution, and ROI definition on the accuracy of standard uptake values: a simulation study. *J Nucl Med.* 45(9): 1519-27 (2004)
9. Boellaard R, O'Doherty MJ, Weber WA, Mottaghy FM, Lonsdale MN, Stroobants SG, Oyen WJ, Kotzerke J, Hoekstra OS, Pruim J, Marsden PK, Tatsch K, Hoekstra CJ, Visser EP, Arends B, Verzijlbergen FJ, Zijlstra JM, Comans EF, Lammertsma AA, Paans AM, Willemsen AT, Beyer T, Bockisch A, Schaefer-Prokop C, Delbeke D, Baum RP, Chiti A, Krause BJ: FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0. *Eur J Nucl Med Mol Imaging.* 37(1): 181-200 (2010)
10. 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. *Ann Thorac Surg.* 82(2):417-22 (2006)

11. Cerfolio RJ, Bryant AS, Ohja B: Restaging patients with N2 (stage IIIa) nonsmall cell lung cancer after neoadjuvant chemoradiotherapy: a prospective study. *J ThoracCardiovasc Surg.* 131(6): 1229–35 (2006)
12. De Leyn P, Doooms C, Kuzdzal J, Lardinois D, Passlick B, Rami-Porta R, Turna A, Van Schil P, Venuta F, Waller D, Weder W, Zielinski M: Revised ESTS guidelines for preoperative mediastinal lymphnode staging for non-small-cell lung cancer *Eur J Cardiothorac Surg.* 45(5): 787-98 (2014)
13. De Leyn P, Doooms C, Kuzdzal J, Lardinois D, Passlick B, Rami-Porta R, Turna A, Van Schil P, Venuta F, Waller D, Weder W, Zielinski M: Preoperative mediastinal lymph node staging for non-small cell lung cancer: 2014 update of the 2007 ESTS guidelines. *Transl Lung Cancer Res.* 3(4): 225-33 (2014)
14. DeLong ER, DeLong DM, Clarke-Pearson DL: Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44(3): 837–845 (1988)
15. Deutsche Gesellschaft für Hämatologie und Medizinische Onkologie e.V. (DGHO). Leitlinie Lungenkarzinom, nicht-kleinzellig. URL: <https://www.onkopedia.com/de/onkopedia/guidelines/lungenkarzinom-nicht-kleinzellig-nscl/@view/html/index.html> (letzter Zugriff am: 23.05.2018)
16. Dössel O: Bildgebende Verfahren in der Medizin - Von der Technik zur medizinischen Anwendung. 2. Auflage. Springer. Berlin Heidelberg 2016.
17. Fahey H: Data acquisition in PET imaging. *J Nucl Med Technol.* 30(2): 39-49. (2002)
18. Gallagher BM, Fowler JS, Gutterson NI, MacGregor RR, Wan CN, Wolf AP: Metabolic trapping as a principle of radiopharmaceutical design: some factors responsible for the biodistribution of [18F] 2-deoxy-2-fluoro-D-glucose. *J Nucl Med.* 19(10): 1154-61 (1978)
19. Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WE, Nicholson AG, Groome P, Mitchell A, Bolejack V: 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. *J Thorac Oncol.* 11(1): 39-51 (2016)
20. Gould MK, Kuschner WG, Rydzak CE, Maclean CC, Demas AN, Shigemitsu H, Chan JK, Owens DK: Test performance of positron emission tomography and computed tomography for mediastinal staging in patients with non-small-cell lung cancer: a meta-analysis. *Ann Intern Med.* 2;139(11): 879-92 (2003)
21. Groheux D, Quere G, Blanc E, Lemaignier C, Vercellino L, de Margerie-Mellon C, Merlet P, Querellou S: FDG PET-CT for solitary pulmonary nodule and lung cancer: Literature review. *Diagn Interv Imaging.* 97(10): 1003-1017 (2016)
22. 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. *Eur J Cardiothorac Surg.* 40(2): 287-90 (2011)
23. Halter G, Buck AK, Schirrmeister H, Aksoy E, Liewald F, Glatting G, Neumaier B, Mühlhling B, Nüssle-Kügele K, Hetzel M, Sunder-Plassmann L, Reske SN: Lymph node

- staging in lung cancer using [18F]FDG-PET. *Thorac Cardiovasc Surg.* 52(2): 96-101 (2004)
24. Hamberg LM, Hunter GJ, Alpert NM, Choi NC, Babich JW, Fischman AJ: The dose uptake ratio as an index of glucose metabolism: useful parameter or oversimplification? *J Nucl Med.* 35(8): 1308-12 (1994)
 25. Hammerschmidt S, Wirtz H: Lung cancer: current diagnosis and treatment. *Dtsch Arztebl Int.* 106(49): 809-18 (2009)
 26. Harders SW, Madsen HH, Hjorthaug K, Arveschoug AK, Rasmussen TR, Meldgaard P, Hoejbjerg JA, Pilegaard HK, Hager H, Rehling M, Rasmussen F: Mediastinal staging in Non-Small-Cell Lung Carcinoma: computed tomography versus F-18-fluorodeoxyglucose positron-emission tomography and computed tomography. *Cancer Imaging.* 14: 23 (2014)
 27. Hashimoto N, Morita K, Tsutsui Y, Himuro K, Baba S, Sasaki M: Time-of-flight information improved the detectability of sub-centimeter sphere using clinical positron emission tomography/computed tomography scanner. *J Nucl Med Technol.* pii: jnmt.117.204735 (2018)
 28. Hellwig D, Graeter TP, Ukena D, Groeschel A, Sybrecht GW, Schaefers HJ, Kirsch C: 18F-FDG PET for mediastinal staging of lung cancer: which SUV threshold makes sense? *J Nucl Med.* 48(11): 1761-6 (2007)
 29. 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. *Eur J Nucl Med Mol Imaging.* 36(2): 194-9 (2009)
 30. Houshmand S, Salavati A, Hess S, Werner TJ, Alavi A, Zaidi H: An update on novel quantitative techniques in the context of evolving whole-body PET imaging. *PET Clin.* 10(1): 45-58 (2015)
 31. Houshmand S, Salavati A, Segtnan EA, Grupe P, Høilund-Carlsen PF, Alavi A: Dual-time-point Imaging and Delayed-time-point Fluorodeoxyglucose-PET/Computed Tomography Imaging in Various Clinical Settings. *PET Clin.* 11(1): 65-84 (2016)
 32. Hwangbo B, Kim SK, Lee HS, Lee HS, Kim MS, Lee JM, Kim HY, Lee GK, Nam BH, Zo JI: Application of endobronchial ultrasound-guided transbronchial needle aspiration following integrated PET/CT in mediastinal staging of potentially operable non-small cell lung cancer. *Chest.* 135(5): 1280-1287 (2009)
 33. Jaskowiak CJ, Bianco JA, Perlman SB, Fine JP: Influence of reconstruction iterations on 18F-FDG PET/CT standardized uptake values. *J Nucl Med.* 46(3): 424-8 (2005)
 34. Jones CM, Brunelli A, Callister ME, Franks KN: Multimodality Treatment of Advanced Non-small Cell Lung Cancer: Where are we with the Evidence? *Curr Surg Rep.* 6(2): 5 (2018)
 35. Kelloff GJ, Hoffman JM, Johnson B, Scher HI, Siegel BA, Cheng EY, Cheson BD, O'shaughnessy J, Guyton KZ, Mankoff DA, Shankar L, Larson SM, Sigman CC, Schilsky RL, Sullivan DC: Progress and promise of FDG-PET imaging for cancer patient management and oncologic drug development. *Clin Cancer Res* 11(8): 2785-808 (2005)

36. Kim DW, Kim WH, Kim CG: Dual-time-point FDG PET/CT: Is It Useful for Lymph Node Staging in Patients with Non-Small-Cell Lung Cancer? *Nucl Med Mol Imaging.* 46(3):196-200 (2012)
37. Kinahan PE, Fletcher JW: Positron emission tomography-computed tomography standardized uptake values in clinical practice and assessing response to therapy. *Semin Ultrasound CT MR.* 31(6): 496-505 (2010)
38. Landis JR, Koch GG: The measurement of observer agreement for categorical data. *Biometrics* 33(1): 159-174 (1977)
39. Langen KJ, Braun U, Rota Kops E, Herzog H, Kuwert T, Nebeling B, Feinendegen LE: The influence of plasma glucose levels on fluorine-18-fluorodeoxyglucose uptake in bronchial carcinomas. *J Nucl Med.* 34(3): 355-9 (1993)
40. Lasnon C, Hicks RJ, Beaugregard JM, Milner A, Paciencia M, Guizard AV, Bardet S, Gervais R, Lemoel G, Zalcman G, Aide N: Impact of point spread function reconstruction on thoracic lymph node staging with 18F-FDG PET/CT in non-small cell lung cancer. *Clin Nucl Med.* 37(10): 971-6 (2012)
41. Li M, Wu N, Liu Y, Zheng R, Liang Y, Zhang W, Zhao P: Regional nodal staging with 18F-FDG PET-CT in non-small cell lung cancer: Additional diagnostic value of CT attenuation and dual-time-point imaging. *Eur J Radiol.* 81(8): 1886-90 (2012)
42. Li S, Zheng Q, Ma Y, Wang Y, Feng Y, Zhao B, Yang Y: Implications of false negative and false positive diagnosis in lymph node staging of NSCLC by means of ¹⁸F-FDG PET/CT. *PLoS One.* 8(10): e78552. (2013)
43. Mac Manus MP, Hicks RJ, Matthews JP, McKenzie A, Rischin D, Salminen EK, Ball DL: Positron emission tomography is superior to computed tomography scanning for response-assessment after radical radiotherapy or chemoradiotherapy in patients with non-small-cell lung cancer. *J Clin Oncol.* 21(7): 1285-92 (2003)
44. Mattes MD, Moshchinsky AB, Ahsanuddin S, Rizk NP, Foster A, Wu AJ, Ashamalla H, Weber WA, Rimner A: Ratio of Lymph Node to Primary Tumor SUV on PET/CT Accurately Predicts Nodal Malignancy in Non-Small-Cell Lung Cancer. *Clin Lung Cancer.* (6): e253-8 (2015)
45. Matthies A, Hickeson M, Cuchiara A, Alavi A: Dual time point 18F-FDG PET for the evaluation of pulmonary nodules. *J Nucl Med.* 43(7): 871-5 (2002)
46. Mohnike W, Hör G, Selbert H: PET-CT Atlas, 2. Auflage. Springer-Verlag. Berlin Heidelberg 2011
47. Mochizuki T, Tsukamoto E, Kuge Y, Kanegae K, Zhao S, Hikosaka K, Hosokawa M, Kohanawa M, Tamaki N: FDG uptake and glucose transporter subtype expressions in experimental tumor and inflammation models. *J Nucl Med.* 42(10): 1551-5 (2001)
48. Moloney F, Ryan D, McCarthy L, McCarthy J, Burke L, Henry MT, Kennedy MP, Hinchion J, McSweeney S, Maher MM, O'Regan K: Increasing the accuracy of 18F-FDG PET/CT interpretation of "mildly positive" mediastinal nodes in the staging of non-small cell lung cancer. *Eur J Radiol.* 83(5): 843-7 (2014)
49. Mountain CF, Dresler CM: Regional lymph node classification for lung cancer staging. *Chest.* 111(6): 1718-1723 (1997)

50. Nelson CA, Wang JQ, Leav I, Crane PD: The interaction among glucose transport, hexokinase, and glucose-6-phosphatase with respect to 3H-2-deoxyglucose retention in murine tumor models. *Nucl Med Biol.* 23(4): 533-41. (1996)
51. Nguyen P, Bhatt M, Bashirzadeh F, Hundloe J, Ware R, Fielding D, Ravi Kumar AS. Comparison of objective criteria and expert visual interpretation to classify benign and malignant hilar and mediastinal nodes on 18-F FDG PET/CT. *Respirology.* (1): 129-37 (2015)
52. Park YS, Lee J, Pang JC, Chung DH, Lee SM, Yim JJ, Yang SC, Yoo CG, Kim YW, Han SK: Clinical implication of microscopic anthracotic pigment in mediastinal staging of non-small cell lung cancer by endobronchial ultrasound-guided transbronchial needle aspiration. *J Korean Med Sci.* (4):550-4 (2013)
53. Popper H: Progression and metastasis of lung cancer. *Cancer Metastasis Rev.* 35(1): 75-91 (2016)
54. Postmus PE, Kerr KM, Oudkerk M, Senan S, Waller DA, Vansteenkiste J, Escriu C, Peters S; ESMO Guidelines Committee: Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol.* 28(suppl_4): iv1-iv21 (2017)
55. Jonathan M. Samet, Erika Avila-Tang, Paolo Boffetta, Lindsay M. Hannan, Susan Olivo-Marston, Michael J. Thun, Charles M. Rudin: Lung Cancer in never smokers: Clinical epidemiology and environmental risk factors. *Clin Cancer Res.* 15(18): 5626–5645 (2009)
56. Sanz-Viedma S, Torigian DA, Parsons M, Basu S, Alavi A: Potential clinical utility of dual time point FDG-PET for distinguishing benign from malignant lesions: implications for oncological imaging. *Rev Esp Med Nucl.* 28(3): 159-66 (2009)
57. Schmidt-Hansen M, Baldwin DR, Hasler E, Zamora J, Abaira V, Roqué I Figuls M: PET-CT for assessing mediastinal lymph node involvement in patients with suspected resectable non-small cell lung cancer. *Cochrane Database Syst Rev.* 13(11): CD009519 (2014)
58. Serke M, Stanzel F, Westhoff M: Geschlechtsspezifische Unterschiede beim Lungenkarzinom. *Pneumologie* 67(5): 270-9 (2013)
59. Shen G, Hu S, Deng H, Jia Z: Diagnostic value of dual time-point 18 F-FDG PET/CT versus single time-point imaging for detection of mediastinal nodal metastasis in non-small cell lung cancer patients: a meta-analysis. *Acta Radiol.* 56(6): 681-7. (2015)
60. Shim SS, Lee KS, Kim BT, Chung MJ, Lee EJ, Han J, Choi JY, Kwon OJ, Shim YM, Kim S: Non-small cell lung cancer: prospective comparison of integrated FDG PET/CT and CT alone for preoperative staging. *Radiology.* 236(3): 1011-9 (2005)
61. Shim SS, Han J: FDG-PET/CT imaging in assessing mucin-producing non-small cell lung cancer with pathologic correlation. *Ann Nucl Med.* 24(5): 357-62 (2010)
62. Shinya T, Rai K, Okumura Y, Fujiwara K, Matsuo K, Yonei T, Sato T, Watanabe K, Kawai H, Sato S, Kanazawa S. Dual-time-point F-18 FDG PET/CT for evaluation of intrathoracic lymph nodes in patients with non-small cell lung cancer. *Clin Nucl Med.* 34(4): 216-21 (2009)

63. Smulders SA, Gundy CM, van Lingen A, Comans EF, Smeenk FW, Hoekstra OS; Study Group of Clinical PET: 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. *Mol Imaging Biol.* (5): 318-22 (2007)
64. Spiro SG1, Gould MK, Colice GL: Initial evaluation of the patient with lung cancer: symptoms, signs, laboratory tests, and paraneoplastic syndromes: ACCP evidenced-based clinical practice guidelines (2nd edition). *Chest.* 132(3 Suppl): 149S-160S (2007)
65. Suga K, Kawakami Y, Hiyama A, Sugi K, Okabe K, Matsumoto T, Ueda K, Tanaka N, Matsunaga N: Differential diagnosis between (18)F-FDG-avid metastatic lymph nodes in non-small cell lung cancer and benign nodes on dual-time point PET/CT scan. *Ann Nucl Med.* ;23(6): 523-31. (2009)
66. Takeda A, Sanuki N, Fujii H, Yokosuka N, Nishimura S, Aoki Y, Oku Y, Ozawa Y, Kunieda E: Maximum standardized uptake value on FDG-PET is a strong predictor of overall and disease-free survival for non-small-cell lung cancer patients after stereotactic body radiotherapy. *J Thorac Oncol.* 9(1): 65-73 (2014)
67. Tournoy KG, Maddens S, Gosselin R, Van Maele G, van Meerbeeck JP, Kelles A: Integrated FDG-PET/CT does not make invasive staging of the intrathoracic lymph nodes in non-small cell lung cancer redundant: a prospective study. *Thorax.* 62(8): 696-701 (2007)
68. Townsend DW: Combined positron emission tomography-computed tomography: the historical perspective. 29(4): 232–235. (2008)
69. Uesaka D, Demura Y, Ishizaki T, Ameshima S, Miyamori I, Sasaki M, Fujibayashi Y, Okazawa H: Evaluation of dual-time-point 18F-FDG PET for staging in patients with lung cancer. *J Nucl Med.* 49(10): 1606-12 (2008)
70. Verschakelen JA, De Wever W, Bogaert J: Role of computed tomography in lung cancer staging. *Curr Opin Pulm Med.* 10(4): 248-55 (2004)
71. Volpi S, Ali JM, Tasker A, Peryt A, Aresu G, Coonar AS: The role of positron emission tomography in the diagnosis, staging and response assessment of non-small cell lung cancer. *Ann Transl Med.* 6(5): 95 (2018)
72. Von Schulthess: Cost considerations regarding an integrated CT-PET system. *Eur. Radiol.* 10(3): S377-S380 (2000)
73. Wang J, Welch K, Wang L, Kong FM. Negative predictive value of positron emission tomography and computed tomography for stage T1-2N0 non-small-cell lung cancer: a meta-analysis. *Clin Lung Cancer.* 13(2): 81-9 (2012)
74. Wang J, Wong KK, Piert M, Stanton P, Frey KA, Kong FS: Metabolic response assessment with 18F-FDG PET/CT: inter-method comparison and prognostic significance for patients with non-small cell lung cancer. *J Radiat Oncol.* 4(3): 249-256 (2015)
75. Warburg O, Posener K, Negelein E: Über den Stoffwechsel der Carcinomzelle. *Biochemische Zeitschrift.* 152:309-35 (1924)

76. Wood IS, Trayhurn P: Glucose transporters (GLUT and SGLT): expanded families of sugar transport proteins. *Br J Nutr.* 89(1): 3-9 (2003)
77. Yang W, Fu Z, Yu J, Yuan S, Zhang B, Li D, Xing L, Zhao D, Mu D, Sun X, Fang Y, Huang Y, Li W: Value of PET/CT versus enhanced CT for locoregional lymph nodes in non-small cell lung cancer. *Lung Cancer.* 61(1): 35-43 (2008)
78. Yeh DW, Lee KS, Han J, Yi CA, Lee HY, Chung MJ, Kim TS. Mediastinal nodes in patients with non-small cell lung cancer: MRI findings with PET/CT and pathologic correlation. *AJR Am J Roentgenol.* 193(3): 813-21 (2009)
79. Yen RF, Chen KC, Lee JM, Chang YC, Wang J, Cheng MF, Wu YW, Lee YC: 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? *Eur J Nucl Med Mol Imaging.* 35(7): 1305-15 (2008)
80. Zasadny KR, Wahl RL: Standardized uptake values of normal tissues at PET with 2-[fluorine-18]-fluoro-2-deoxy-D-glucose: variations with body weight and a method for correction. *Radiology.* 189(3): 847-50 (1993)

9. Publikationsverzeichnis

Teilpublikationen der Dissertation sind folgende:

Rogasch JM, Apostolova I, Steffen IG, **Steinkrüger FL**, Genseke P, Riedel S, Wertzel H, Achenbach HJ, Kalinski T, Schultz M, Schreiber J, Amthauer H, Furth C.

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. doi: 10.1016/j.ejrad.2016.05.004. Epub 2016 May 11.

Rogasch JM, Steffen IG, Riedel S, Apostolova I, Wertzel H, Achenbach HJ, **Steinkrüger FL**, Kalinski T, Schultz M, Schreiber J, Amthauer H, Furth C.

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. doi: 10.1007/s00330-015-4093-5. Epub 2015 Nov

10. Danksagung

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

Meinem Doktorvater, Herrn Prof. Dr. Holger Amthauer, danke ich besonders für die Überlassung des Themas und die Möglichkeit, auch nach seiner Berufung nach Berlin an die Charité diese Dissertation beenden zu dürfen. Er war als Ansprechpartner immer verfügbar und hat mich mit seiner freundlichen Art immer ermutigt diese Arbeit fertigzustellen.

Dank gilt natürlich auch meiner Familie, meiner Mutter Cecilia Prieto Lindholm, meinen drei Geschwistern und allen voran meinem im Februar verstorbenen Vater Jürgen Steinkrüger, der mich über all die Jahre während des Studiums und darüber hinaus unterstützt hat.

Speziell möchte ich Herrn Dr. Julian Rogasch danken, der mit seinem fachlichen Rat, seiner hilfreichen Kritik und letztlich seinem unermüdlichen Einsatz mir bei der Realisierung dieser Arbeit zu Seite stand.

11. Ehrenerklärung

Ich erkläre, dass ich die der Medizinischen Fakultät der Otto-von-Guericke-Universität zur Promotion eingereichte Dissertation mit dem Titel

Optimierte diagnostische Genauigkeit der F18-FDG-PET zur prätherapeutischen Dignitätsbeurteilung thorakaler Lymphknoten beim nicht-kleinzelligen Lungenkarzinom (NSCLC): Einfluss einer Dual Time Point-Diagnostik und eines visuellen Befundungsscores.

in der Klinik für Radiologie und Nuklearmedizin, Bereich Nuklearmedizin,

mit Unterstützung durch

Herrn Prof. Dr. med. Holger Amthauer

Herrn PD Dr. med. Christian Furth

Herrn Dr. Julian Rogasch

ohne sonstige Hilfe durchgeführt und bei der Abfassung der Dissertation keine anderen als die dort aufgeführten Hilfsmittel benutzt habe.

Bei der Abfassung der Dissertation sind Rechte Dritter nicht verletzt worden. Die Verwendung von Bildmaterial Dritter geschieht in jedem Fall mit Erlaubnis des jeweiligen Eigentümers, welche mir schriftlich vorliegen, und unter Einhaltung eventueller Auflagen zur Nutzung. Insbesondere sind entsprechende Abbildungen mit der zutreffenden Originalquelle gekennzeichnet.

Ich habe diese Dissertation bisher an keiner in- oder ausländischen Hochschule zur Promotion eingereicht. Ich übertrage der Medizinischen Fakultät das Recht, weitere Kopien meiner Dissertation herzustellen und zu vertreiben.

Magdeburg, den 07.10.2019

12. Lebenslauf

Persönliche Daten

Name: Ferdinand Ludwig Gunnar Alvaro
 Geboren am: 26.05.1989 in Bonn

Berufliche Tätigkeiten

Sept. 2018 – Juli 2019 Arzt im Kommunikationsteam bei Amboss GmbH

Akademische Ausbildung

Januar – August 2018 Promotionssemester im Bereich Medizin an der Universität Magdeburg
 2012 – 2017 Medizin an der Universität Magdeburg mit Abschluss Ärztliche Prüfung
 2015 Wissenschaftliche Hilfskraft in der Nuklearmedizin, Universitätsklinikum Magdeburg
 Sept. 2010 – März 2012 Medizin, Universität Luxemburg
 Okt. 2009 – Sept. 2010 VWL, Universität Göttingen

Praktisches Jahr

Juli – Okt. 2017 Bundeswehrkrankenhaus Berlin
 März – Juli 2017 Centre Hospitalier Universitaire (CHU) de Martinique
 Nov. 2016 – März 2017 Hospital de Clínicas Dr. Manuel Quintela, Montevideo

Wehrdienst

Juli 2008 – März 2009 Gebirgsjägerbataillon 232, Berchtesgaden

Schule

2006 – 2008 Kolleg St. Blasien, (Internat)
 2003 – 2006 Alexander-von-Humboldt-Schule, Caracas/Venezuela
 2002 – 2003 Sekundarschule Moos, Muri bei Bern/Schweiz
 2000 – 2002 Primarschule Seidenberg, Muri bei Bern/Schweiz
 1999 – 2000 Volksschule Herrenberg, Rapperswil-Jona/Schweiz
 1995 – 1999 Alexander-von-Humboldt-Schule, Lima/Peru

Sonstiges

Fremdsprachen
 Spanisch zweite Muttersprache
 Englisch verhandlungssicher
 Französisch sehr gute Kenntnisse

Interessen

Geschichte, Joggen, Wandern

Magdeburg, den 07.10.2019