

Aus der Klinik und Poliklinik für Nuklearmedizin
der Universitätsmedizin der Johannes Gutenberg-Universität Mainz

Diagnostische Sicherheit von Auswertelgorithmen
der FDG-PET/CT zur Rezidivbeurteilung beim lokal fortgeschrittenen
Lungenkarzinom nach Radiochemotherapie

Inauguraldissertation
zur Erlangung des Doktorgrades der Medizin
der Universitätsmedizin
der Johannes Gutenberg-Universität Mainz

Vorgelegt von

Alexander Dominik Brose
aus Heilbronn

Mainz, 2020

Wissenschaftlicher Vorstand:

1. Gutachter:

2. Gutachter:

Tag der Promotion:

08. Dezember 2020

Für Sam, Hal, Benedict und Henry

Inhaltsverzeichnis

I.	Abbildungs- und Tabellenverzeichnis	VII
II.	Abkürzungsverzeichnis	IX
1.	Einleitung	1
1.1.	Die Rolle der PET/CT in der Diagnostik des NSCLC	2
1.2.	Ziel und Motivation der Arbeit	2
2.	Literaturdiskussion	3
2.1.	Aktuelle Leitlinienempfehlungen beim NSCLC	3
2.1.1.	TNM-Klassifikation und UICC-Stadieneinteilung	3
2.1.2.	Therapeutische Optionen und Prognose	5
2.2.	Einsatz der PET/CT beim NSCLC	6
2.2.1.	Ausbreitungsdiagnostik (Staging)	6
2.2.2.	Beurteilung der Therapieansprache	8
2.2.3.	Herausforderung in der Rezidivdiagnostik	11
2.2.4.	Lokal- versus Fernrezidiv	14
2.3.	Herausforderungen in der modernen PET/CT-Diagnostik	15
2.3.1.	Die qualitative und quantitative Analyse von PET-Informationen	15
2.3.2.	Die Interobservervariabilität in der PET/CT-Diagnostik	18
2.4.	Die PET-Plan Studie	19
3.	Material und Methoden	24
3.1.	Die Auswahl der Daten aus der PET-Plan Studie	24
3.2.	Gerätetechnik und Bildakquisition	26
3.2.1.	Grundlagen der PET/CT	26
3.3.	Die qualitative Auswertung durch die <i>Observer</i>	28
3.3.1.	Erstellung des Fragebogens	29
3.3.2.	Rekrutierung der <i>Observer</i>	32
3.3.3.	Auswertung durch die <i>Observer</i>	32
3.4.	Die quantitative Auswertung	34
3.4.1.	Die Auswertung nach den PERCIST- und PREDIST-Kriterien	34
3.4.2.	Die Berechnung von MTV und TLG	35
3.5.	Die Festlegung des Goldstandards	36
3.5.1.	Qualitätssicherung	37
3.6.	Die statistische Auswertung	37
3.6.1.	Grundlagen der Interobservervariabilität und Cohens Kappa	37
3.6.2.	Odds Ratio und logistische Regression	40

3.6.3.	Ereigniszeitanalysen.....	41
3.6.4.	Sensitivität und Spezifität.....	42
3.6.5.	Weitere Berechnungen	42
3.6.6.	Datenverwaltung und Literaturrecherche	43
4.	Ergebnisse.....	44
4.1.	Charakterisierung des Patientenkollektivs.....	44
4.1.1.	Ergebnis des Goldstandards aus der klinischen Dokumentation	46
4.2.	Die qualitative Auswertung durch die vier <i>Observer</i>	47
4.2.1.	Überblick.....	47
4.2.2.	Die visuelle Auswertung der Lunge	48
4.2.3.	Die Auswertung des Mediastinum	52
4.2.4.	Urteil über Fernmetastasen	52
4.2.5.	Die Gesamtauswertung der <i>Observer</i>	52
4.2.6.	Die subjektive Beurteilungssicherheit der <i>Observer</i>	54
4.3.	Die quantitative Auswertung.....	54
4.3.1.	Überblick.....	54
4.3.2.	Ergebnisse der PERCIST- und PREDIST-Auswertungen	56
4.4.	Vergleich der verschiedenen Auswertungen	58
4.4.1.	Qualitative vs. quantitative Auswertung	58
4.4.2.	Vergleich der qualitativen Auswertung mit dem Goldstandard	59
4.4.3.	Vergleich der quantitativen Auswertung mit dem Goldstandard	61
4.4.4.	Die zeitliche Abhängigkeit der Sensitivität und Spezifität.....	61
4.5.	Evaluation der PET-Parameter	62
4.5.1.	Einfluss der Parameter aus der Baseline PET/CT auf das Outcome.....	62
4.5.2.	PET-Parameter in der Rezidivdiagnostik	63
5.	Diskussion	65
5.1.	Das Patientenkollektiv	65
5.2.	Die qualitative Auswertung.....	66
5.2.1.	Die visuelle Auswertung der Lunge	66
5.2.2.	Bewertung des Mediastinum	70
5.2.3.	Urteil über Fernmetastasen	71
5.2.4.	Gesamturteil	71
5.2.5.	Einfluss der Beurteilungssicherheit und der Kenntnisse	71
5.2.6.	Sensitivität und Spezifität anhand des Goldstandards.....	74
5.2.7.	Das Interobserver-Agreement im Vergleich zur Literatur.....	76
5.3.	Die quantitative Auswertung.....	78

5.3.1. Die Ergebnisse der PERCIST- und der PREDIST-Kriterien	78
5.3.2. Sensitivität und Spezifität im Vergleich zum Goldstandard	81
5.3.3. Die Bedeutung der Schwellenwert-Definition.....	81
5.4. Vergleich zwischen quantitativer und qualitativer Auswertung	83
5.5. Der Zeitpunkt der Rezidivdiagnostik.....	84
5.6. Abschätzung der Prognose anhand der PET-Parameter	85
5.7. Limitationen der Studie.....	87
5.7.1. Datenerhebung	87
5.7.2. Qualitative Auswertung	87
5.7.3. Quantitative Auswertung.....	87
5.7.4. Goldstandard	87
6. Zusammenfassung	88
7. Literaturverzeichnis.....	90
Danksagung	95

I. Abbildungs- und Tabellenverzeichnis

Tab. 2.1	TNM-Klassifikation des NSCLC	4
Tab. 2.2	UICC8-Stadien des NSCLC	4
Tab. 2.3	5-Jahresüberlebensraten des NSCLC	6
Abb. 2.4	Empfehlungen zum Zeitpunkt der Kontrolle der Therapieansprache	14
Tab. 3.1	Einschluss- und Ausschlusskriterien der PET-Plan Studie	25
Abb. 3.2	Beispiel für die Koregistrierung von CT und PET	27
Abb. 3.3	Fragebogen zur reproduzierbaren qualitativen Auswertung, Seite 1	30
Abb. 3.4	Fragebogen zur reproduzierbaren qualitativen Auswertung, Seite 2	31
Abb. 3.5	Vier-Felder-Tafel des Urteils beider Ärzte	38
Abb. 3.6	Sensitivität und Spezifität	42
Tab. 4.1	Charakterisierung des Patientenkollektivs	44
Tab. 4.2	Zeitintervall und Anzahl der erfolgten PET/CTs im Follow Up	45
Abb. 4.3	Zeitintervall zwischen Therapiebeginn und Rezidivverdacht	46
Tab. 4.4	Häufigkeit und Lokalisation bei vorliegendem Rezidiv	46
Abb. 4.5	Interpretation des IOA nach Landis und Koch	47
Tab. 4.6	Urteilshäufigkeit für das Vorliegen eines Rezidiv laut <i>Observer</i>	48
Tab. 4.7	IOA für die Beurteilung der Lunge	49
Tab. 4.8	Odds Ratio und relatives Risiko für die pulmonalen Risikokriterien	50
Tab. 4.9	Logistische Regressionsanalyse der pulmonalen Risikokriterien	51
Tab. 4.10	IOA für die Beurteilung der mediastinalen Lymphknotenstationen	52
Tab. 4.11	Übersicht über das IOA der Gesamtauswertung der <i>Observer</i>	53
Tab. 4.12	Beurteilungssicherheit der <i>Observer</i>	54

Tab. 4.13	Charakteristika der Baseline FDG-PET/CT	55
Abb. 4.14	PET-Parameter zum Zeitpunkt der Bestrahlungsplanung	55
Tab. 4.15	Ergebnisse der PERCIST-Auswertung abhängig vom Zeitpunkt	56
Tab. 4.16	Ergebnisse der PERCIST-Auswertung abhängig von der Wertung	56
Tab. 4.17	Ergebnisse der Auswertung nach den PREDIST-Kriterien	57
Tab. 4.18	IOA zwischen <i>Observer</i> -Konsens und PERCIST-Auswertung	58
Tab. 4.19	IOA zwischen <i>Observer</i> -Konsens und PREDIST-Auswertung	59
Tab. 4.20	Sensitivität und Spezifität der qualitativen Auswertung	60
Tab. 4.21	Sensitivität und Spezifität der quantitativen Auswertung	60
Tab. 4.22	Sensitivität und Spezifität in Abhängigkeit der Zeit	61
Tab. 4.23	Cox-Regression der PET-Parameter aus der Baseline-PET/CT	62
Abb. 4.24	Prognose der PET-Parameter zum Outcome der Patienten	63
Abb. 4.25	Die PET-Parameter im Follow Up	64
Tab. 4.26	Sensitivität und Spezifität der PET-Parameter im Follow Up	64
Abb. 5.1	Beispiel für die unsichere Zuordnung einer Läsion	73

II. Abkürzungsverzeichnis

¹⁸ F-FDG	¹⁸ F-Fluordesoxyglucose
3D-CRT	Dreidimensionale konformale Radiotherapie
ALK	Anaplastic lymphoma kinase (Tyrosinkinase, Protoonkogen)
AWMF	Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften
CT	Computertomographie
CECT	Contrast enhanced computed tomography
CMR	Complete Metabolic Remission
CR	Complete Response
CTV	Clinical Target Volume
EGFR	Epidermal growth factor receptor (Wachstumsfaktorrezeptor)
EORTC	European Organisation for Research and Treatment of Cancer
FDG	Fluordesoxyglucose
IASLC	International Association for the Study of Lung Cancer
IMRT	Intensitätsmodulierte Radiotherapie
ITV	Internal Target Volume
IOA	Interobserver- <i>Agreement</i> (Übereinstimmungswert der entsprechenden Interobservervariabilität)
IOV	Interobservervariabilität
IQR	Inter quartile range (Interquartilenabstand)
LBM	Lean Body Mass (fettbereinigte Körpermasse)
MTV	Metabolic Tumor Volume
n	Anzahl

NSCLC	Non small cell lung cancer (nicht-kleinzelliges Lungenkarzinom)
SCLC	Small cell lung cancer (kleinzelliges Lungenkarzinom)
OS	Overall Survival
PD	Progressive Disease
PET	Positronenemissionstomographie
PET/CT	Kombiniertes Verfahren von PET und CT
PERCIST	PET Response Criteria in Solid Tumors
PREDIST	PET Residual Disease in Solid Tumor
PMD	Progressive Metabolic Disease
PMR	Partial Metabolic Response
PR	Partial Response
PTV	Planning Target Volume
RECIST	Response Evaluation Criteria in Solid Tumors
ROS1	c-ros oncogene 1 (Tyrosinkinase, Protoonkogen)
SD	Stable disease
SMD	Stable metabolic disease
std. dev.	Standard deviation (Standardabweichung SD)
SUL	Standardized Uptake Value corrected for Lean Body Mass
SUV	Standardized Uptake Value
TLG	Total Lesion Glycolysis
UICC	Union for International Cancer Control
VOI	Volume of interest

1. Einleitung

In der Krebsstatistik für Deutschland stellen bösartige Neubildungen der Lunge und des Bronchialsystems bei beiden Geschlechtern die jeweils zweithäufigste maligne Tumorentität nach dem Prostatakarzinom und dem Mammakarzinom dar. Damit gehört das Bronchialkarzinom nicht nur zu den häufigsten Krebsneuerkrankungen, sondern ist in den deutschsprachigen Ländern mit rund 20 Prozent auch die häufigste Krebstodesursache innerhalb eines Jahres (1, 2).

Histopathologisch unterscheidet man das kleinzellige (SCLC) vom nicht-kleinzelligen Bronchialkarzinom (NSCLC). Letzteres stellt mit etwa 85 Prozent die häufigste Form von Lungenkrebs dar und wird wiederum in weitere histologische Subtypen (Plattenepithelkarzinom, Adenokarzinom und großzelliges Bronchialkarzinom) und anhand molekulargenetischer Merkmale (z.B. EGFR-Mutation, ALK-Translokation, ROS1-Translokation) unterschieden. Zu den Hauptrisikofaktoren an einem Bronchialkarzinom zu erkranken zählt insbesondere das Zigarettenrauchen, wobei Dauer und Menge des Konsums eine nicht unwesentliche Rolle spielen. Symptome die zur Detektion der Erkrankung führen, sind meist abhängig von Lage, Größe und Metastasierung des Tumors und treten häufig erst in einem fortgeschrittenen Stadium der Erkrankung auf. Ein kurativer Therapieansatz durch operative Resektion des Tumors ist häufig nur in einer frühen Phase der Erkrankung möglich.

Etwa zwei Drittel der Erkrankten gelten jedoch zum Zeitpunkt der Diagnosestellung aufgrund der Tumorgröße, einer Metastasierung oder einer vorbestehenden Lungenerkrankung als inoperabel. Für viele Patienten verbleibt dann zwar die Möglichkeit eines palliativen Therapieansatzes, welcher das Überleben der Erkrankten um einige Monate oder wenige Jahre verlängern kann. Jedoch besteht bei einem immer größer werdenden Teil der lokal fortgeschrittenen Stadien des nicht-kleinzelligen Lungenkarzinoms auch die Möglichkeit eines kurativen Therapieansatzes mittels multimodalem Konzept aus Bestrahlung und systemischer Therapie (3, 4).

Insgesamt hat sich die 5-Jahres-Überlebensrate dank neuer zielgerichteter Therapien in den letzten Jahren geringfügig verbessert, liegt aber zum Zeitpunkt der Diagnosestellung noch immer bei nur etwa 10 - 20%. Liegt bereits ein fortgeschrittenes Tumorstadium vor (UICC-Stadien IIIB und IV), so sinkt die 5-Jahres-Überlebensrate auf unter 5% (3).

1.1. Die Rolle der PET/CT in der Diagnostik des NSCLC

Die bildgebende Diagnostik mittels Computertomographie (CT) und Positronenemissionstomographie (PET) bildet einen der wesentlichen Grundpfeiler in der häufig interdisziplinären Planung des bestmöglichen Therapiekonzeptes. Dabei dient die Bildgebung nicht nur der primären Ausbreitungsdiagnostik als essentieller Aspekt vor der Behandlungsplanung, sondern wird auch während und nach der Therapie eingesetzt um das Ansprechen auf die Behandlung zu beurteilen und ein mögliches Wiederauftreten einer Krebserkrankung frühzeitig zu erkennen. In den letzten beiden Jahrzehnten konnte gezeigt werden, dass ein kombiniertes Verfahren (PET/CT) gegenüber der alleinigen CT überlegen ist, was die primäre Ausbreitungsdiagnostik, die Verlaufsbeurteilung unter Therapie, sowie die prognostische Aussagekraft einer Therapieansprache anbelangt (5-10). Hierfür finden sich in der Literatur neben zahlreichen Vorschlägen auch mittlerweile etablierte Einteilungen für das *Staging* und Kriterien zur Beurteilung des Therapieverlaufs, wie zum Beispiel die PERCIST-Kriterien. Betrachtet man jedoch die Wertigkeit der PET/CT hinsichtlich der Frage nach einem möglichen Rezidiv der Erkrankung, so mangelt es auf der einen Seite vor allem an gut validierten, objektivierbaren und reproduzierbaren Kriterien die ein Rezidiv definieren und dieses von posttherapeutischen Veränderungen, vor allem nach Bestrahlung, differenzieren. Auf der anderen Seite gibt es zu dieser Fragestellung bislang nur wenige Studien, die die häufig visuelle und somit subjektive Interpretation der FDG-PET/CT einer quantitativen und reproduzierbaren Auswertung anhand standardisierter Kriterien gegenüberstellen.

1.2. Ziel und Motivation der Arbeit

Diese Arbeit soll mit der Evaluation eigener Rezidivkriterien und deren Abgrenzung zu entzündlichen Veränderungen einen Beitrag dabei leisten, die Wertigkeit der FDG-PET/CT in der Rezidivdiagnostik beim lokal fortgeschrittenen Lungenkarzinom nach Radiochemotherapie zu überprüfen. In diesem Rahmen werden qualitative und quantitative Merkmale der Auswertung von Bilddaten aus der PET-Plan Studie miteinander verglichen. Die Sensitivität und Spezifität der Kriterien wird anhand des dokumentierten, klinischen Outcomes der Patienten ermittelt. Außerdem soll überprüft werden, inwiefern das Urteil abhängig vom befundenden Arzt ist und ob ein vordefinierter Algorithmus in der Auswertung helfen kann, die Sensitivität und Spezifität der PET/CT hinsichtlich der korrekten Diagnose zu erhöhen.

2. Literaturdiskussion

2.1. Aktuelle Leitlinienempfehlungen beim NSCLC

Neben der aktuellen S3-Leitlinie der AWMF zur „Prävention, Diagnostik, Therapie und Nachsorge des Lungenkarzinoms“ aus dem Jahr 2018 existiert im deutschsprachigen Raum noch eine weitere, durch die Deutsche Gesellschaft für Hämatologie und medizinische Onkologie veröffentlichte Leitlinie für das nicht-kleinzellige Lungenkarzinom aus dem Jahr 2019 (3, 4).

2.1.1. TNM-Klassifikation und UICC-Stadieneinteilung

Das TNM-Klassifikationssystem ist eine weltweit anerkannte und für die meisten malignen Tumorerkrankungen verwendete Beschreibung zur anatomischen Ausbreitung im Körper des erkrankten Patienten. TNM ist ein Akronym der Wörter *Tumor*, *Nodus*, *Metastase* und klassifiziert die lokale Ausdehnung und Größe des Primärtumors (T), betroffene lokale Lymphknotenstationen im Lymphabflussgebiet des Tumors (N) und gegebenenfalls vorhandene Metastasen, in anderen Organsystemen (M) als der primär betroffenen Körperregion. Sie wird von der Internationalen Vereinigung gegen Krebs (UICC) herausgegeben und dient der Einteilung in Erkrankungsstadien. An diesen wiederum orientieren sich nicht nur die in den Leitlinien ausgesprochenen Behandlungsempfehlungen, sondern auch Einteilungen von Patienten in Studien oder prognostische Einschätzungen für den Ausgang einer Therapie (11). In der durch die IASLC aktualisierte Auflage für das Bronchialkarzinom hat sich die Zuordnung des Lokalbefundes zum jeweiligen T-Stadium gegenüber der vorherigen Version geringfügig verändert. Die exakte lokale Tumorgöße wird noch stärker berücksichtigt und so wird beispielsweise ein zwischen 5 - 7 cm messender Primärtumor nun nicht mehr als T2b, sondern als T3 klassifiziert (12). Eine Infiltration des Zwerchfells, welche vormals als T3 galt, wird nun, aufgrund einer schlechteren Prognose als bisher angenommen, als T4-Stadium gewertet (13). Einen Überblick über die aktuelle TNM-Klassifikation des nicht-kleinzelligen Lungenkarzinoms gibt die Tabelle 2.1. Die Zuordnung der TNM-Klassifikation in ein UICC-Stadium ist der Tabelle 2.2. zu entnehmen.

2. Literaturdiskussion. Aktuelle Leitlinienempfehlungen beim NSCLC

Kategorie	Status	Kurzbeschreibung
T (Tumor)	Tis	Carcinoma in situ
	T1	≤ 3 cm, umgeben von Lungenparenchym, kein Kontakt zum Hauptbronchus
	T1a	≤ 1 cm
	T1b	> 1 und ≤ 2 cm
	T1c	> 2 und ≤ 3 cm
	T2	> 3 und ≤ 5 cm <u>oder</u> Kontakt zum Hauptbronchus <u>oder</u> Verschlussatelektase
	T2a	> 3 und ≤ 4 cm
	T2b	> 4 und ≤ 5 cm
	T3	> 5 und ≤ 7 cm <u>oder</u> Infiltration von Brustwand, Perikard, Nervus phrenicus <u>oder</u> Satellitenmetastase im ipsilateralen Lungenlappen
	T4	> 7 cm <u>oder</u> Infiltration von Mediastinum, Zwerchfell, Herzmuskel, Trachea, Ösophagus, große Gefäße <u>oder</u> ipsilaterale Satellitenmetase in anderem Lungenlappen
N (Lymphknoten)	N1	Ipsilateral peribronchial <u>und/oder</u> hilär <u>und/oder</u> intrapulmonal
	N2	Ipsilateral mediastinal <u>und/oder</u> infrakarinal
	N3	Kontralateral mediastinal <u>oder</u> kontralateral hilär <u>oder</u> supraklavikulär
M (Metastasen)	M1	Femmetastasen
	M1a	Kontralaterale Lunge, Pleura-/Perikardnodulus, maligner Erguss
	M1b	Singuläre extrathorakale Metastase
	M1c	Multiple extrathorakale Metastasen

Tabelle 2.1 TNM-Klassifikation des nicht-kleinzelligen Lungenkarzinoms, 8. Version (11).

	N0	N1	N2	N3
T1	IA	IIB	IIIA	IIIB
T2a	IB	IIB	IIIA	IIIB
T2b	IIA	IIB	IIIA	IIIB
T3	IIB	IIIA	IIIB	IIIC
T4	IIIA	IIIA	IIIB	IIIC
M1a	IVA	IVA	IVA	IVA
M1b	IVA	IVA	IVA	IVA
M1c	IVB	IVB	IVB	IVB

Tabelle 2.2 UICC8-Stadien des nicht-kleinzelligen Lungenkarzinoms (11).

2.1.2. Therapeutische Optionen und Prognose

Die Behandlung des NSCLC erfolgt in Abhängigkeit des UICC-Stadiums. Dabei fließen neben dem Allgemeinzustand und den Komorbiditäten des Patienten auch molekulargenetische Merkmale des Tumors in die Therapieentscheidung ein (14). Die kritische Abwägung zwischen den Behandlungsoptionen der chirurgischen Resektion, der Bestrahlung und der (neo-)adjuvanten systemischen Therapie bzw. deren Kombination ist die Aufgabe von interdisziplinären Tumorkonferenzen. Einen modifizierten Überblick über die 5-Jahres-Überlebensrate bietet Tabelle 2.3. Prinzipiell gilt beim NSCLC, wie bei anderen malignen Tumorentitäten auch, je geringer das Tumorstadium zu Beginn, desto größer die Aussicht auf Heilung; so können operationsfähige Patienten bis zum Stadium IIIA prinzipiell in kurativer Intention therapiert und der Chirurgie zugeführt werden (15). Während in Stadium I die alleinige operative Resektion die Methode der Wahl darstellt, wird für das Stadium II ein Konzept bestehend aus der Kombination von chirurgischer Resektion des betroffenen Lungenlappens, sowie einer (neo-)adjuvanten systemischen Chemotherapie vor bzw. nach der Operation angestrebt (16, 17). Im Stadium IIIA folgt der adjuvanten Chemotherapie in der Regel zusätzlich eine Bestrahlung der betroffenen mediastinalen Lymphknotenstationen zur Senkung des lokalen Rezidivrisikos. Es konnte bislang jedoch kein Einfluss der Nachbestrahlung auf die Gesamtüberlebenszeit gezeigt werden (3). Die Nachbestimmung des Nodalstatus erfolgt hier zumeist aus der histopathologischen Aufarbeitung der intraoperativen systematischen Lymphknotendissektion. In niedrigen Stadien (Stadium IA und IB) kann die alleinige Strahlentherapie ohne zusätzliche systemische Therapie bei nicht operationsfähigen Patienten ein kuratives Therapiekonzept darstellen (18, 19).

Bei den nicht operablen Stadien IIA/B und IIIA und den lokal fortgeschrittenen Stadien IIIB/C, wie im Kollektiv dieser Arbeit, stellt die kombinierte Radiochemotherapie die bestmögliche, noch kurativ intendierte Therapieform dar (20, 21). Da in diesen Fällen, aufgrund der fehlenden mediastinalen Lymphknotendissektion, keine histopathologische Informationen über gesicherte mediastinale Lymphknotenmetastasen vorliegt, kommt der Bildgebung eine besonders tragende Rolle in der Planung der Bestrahlung zu (s. Kapitel 2.3.1. Ausbreitungsdiagnostik beim NSCLC). Der klinisch etablierte Standard einer elektiven mediastinalen Lymphknotenbestrahlung wird mit den Ergebnissen klinischer Studien aus den letzten Jahren zunehmend in Frage gestellt (22, 23).

2. Literaturdiskussion. Einsatz der PET/CT beim NSCLC

Im Stadium der metastasierten Erkrankung (Stadium IVA/B) ist die Therapieintention hingegen meist palliativ und es stehen die lokale Kontrolle, sowie die gleichzeitige Vermeidung unnötiger Nebenwirkungen oder Komplikationen im Fokus einer leitliniengerechten Behandlung. Erwähnenswert ist in diesem Zusammenhang, dass bei etwa 35 – 40 % aller Patienten mit NSCLC bei Erstdiagnose bereits ein Stadium IV vorliegt (4). Die FDG-PET/CT besitzt in diesem Zusammenhang eine hohe diagnostische Wertigkeit zur Bestätigung oder zum Ausschluss von Fernmetastasen.

UICC-Stadium	5-Jahres-Überlebensrate
I	
IA	75 - 80 %
IB	55 - 60 %
II	
IIA	45 - 55 %
IIB	35 - 45 %
III	
IIIA	15 - 40 %
IIIB/IIIC	5 - 10 %
IV	
IVA	< 5 %

Tabelle 2.3 5-Jahresüberlebensraten des NSCLC modifiziert nach aktueller Leitlinie (3).

2.2. Einsatz der PET/CT beim NSCLC

Der PET-Bildgebung kommt in der Diagnostik und Therapiekontrolle des NSCLC aufgrund ihrer einzigartigen Möglichkeit nicht-invasiv in den lebendigen menschlichen Organismus zu blicken eine fundamentale und unentbehrliche Rolle zu verschiedenen Zeitpunkten des Erkrankungsverlaufes zu.

2.2.1. Ausbreitungsdiagnostik (Staging)

Die bildgebende Ausbreitungsdiagnostik ist Grundlage für die anschließende Einordnung in das vorliegende Stadium der Tumorerkrankung, bestimmt maßgeblich die Intention und Form der Therapie und dient als Prognosefaktor und Ausgangsbefund für Verlaufskontrollen (3). Die Bildgebung ist für die Ausbreitungsdiagnostik des Bronchialkarzinoms in der modernen Medizin somit

schlichtweg unverzichtbar. Dies spiegelt sich auch in den aktuellen Leitlinien wider, welche zwischen einer initialen Diagnostik bei neu aufgetretenen Symptomen und einer Ausbreitungsdiagnostik unterscheiden. Symptome, die zu einer weiteren Diagnostik führen, können zum Beispiel ein über 2 Wochen anhaltender Husten ohne Besserung sein. Vor allem bei positiver Raucheranamnese sollte dies zu einer weiteren Abklärung führen (4). Empfohlen wird hier zunächst die Diagnostik mittels kontrastmittelverstärkter CT zur Bestätigung, die bei positivem Befund eine histologische Sicherung nach sich ziehen sollte. Fällt der Befund der Histologie positiv für ein NSCLC aus, empfiehlt die Leitlinie im nächsten Schritt die primäre Ausbreitungsdiagnostik mittels Ganzkörper FDG-PET/CT durchzuführen (3). Dieses Vorgehen beruht vor allem auf der Überlegenheit der PET/CT gegenüber der CT bei der Genauigkeit des mediastinalen Lymphknoten-Stagings und der Detektion unerwarteter Fernmetastasen (5, 6, 10, 24). Auch die aktuelle AWMF-Leitlinie hebt eine höhere Sensitivität der PET gegenüber der CT bei der Dignitäts- und Größenbeurteilung von Lungenrundherden, insbesondere bei zusätzlichen Vorliegen von Atelektasen, hervor. Eingeschränkt wird die Aussagekraft der PET lediglich durch Tumoren mit vermindertem Glukosemetabolismus oder assoziierten entzündlichen Prozessen, die einen erhöhten Glukoseumsatz aufweisen (4). Durch eine PET/CT basierte Ausbreitungsdiagnostik kann außerdem die Rate an futilen Operationen um etwa die Hälfte gesenkt werden (25).

Dies wird besonders deutlich, wenn man die hohe Sensitivität der PET/CT beim N-Staging, einer rein morphologischen Begutachtung mittels CT gegenüberstellt. Bei der Anwendung eines Durchmesser von 10 mm in der kurzen Achse als Schwellenwert, wird hierfür eine Sensitivität von etwa 62 % berichtet. Sind vergrößerte Lymphknoten beschrieben, so liegt lediglich in 54 % der Fälle auch ein positiver Nodalbefund vor. Ist die CT negativ liegt in rund 17 % der Fälle dennoch ein positiver mediastinaler Lymphknotenbefund vor, wohingegen ein negativer PET/CT-Befund eine deutlich höhere Sicherheit bietet. Hier verbleibt eine Restwahrscheinlichkeit von 6 % für nicht-detektierte mediastinale Lymphknotenmetastasen (4, 24). Diese vermeintlich falsch negativen PET-Befunde können mit der Abschätzung der Befallswahrscheinlichkeit einzelner mediastinaler Lymphknotenstationen abhängig von der Lokalisation des Primärtumors umgangen werden (26). Dies ist bislang vor allem in Erwägung einer elektiven mediastinalen Bestrahlung relevant, wobei einige Studien aus den vergangenen Jahren vermehrt Zweifel an diesem Konzept wecken (22).

Ähnlich verhält es sich auch in Bezug auf unerwartete Fernmetastasen. Insbesondere für das in dieser Studie vorliegende Patientenkollektiv im UICC-Stadium II und III konnte gezeigt werden, dass sich durch eine ergänzende FDG-PET/CT Diagnostik die Rate von nicht erwarteten Fernmetastasen um 18 % bzw. 24 % in den jeweiligen Stadien erhöhte (27). Schließlich sind auch noch die neben den qualitativen Informationen, die in der PET/CT enthaltenen quantitativen Parameter, wie MTV und TLG, zu erwähnen, die sich in einigen Studien als vielversprechende Verlaufs- und Prognoseparameter erwiesen haben (28-30). Die breite Verfügbarkeit der CT im Vergleich zur PET/CT, sowie die geringeren Kosten bilden letztendlich die einzig verbleibende Rationale, die einen Einsatz der CT als primäres Diagnostikum bei hartem Verdachtsmoment für ein NSCLC rechtfertigen könnte.

2.2.2. Beurteilung der Therapieansprache

Wann immer möglich, sollte ein Patient mit Bronchialkarzinom im fortgeschrittenen Stadium in eine Studie eingeschlossen werden und die Therapie anhand eines Studienprotokolls erfolgen (3). Diese Empfehlung der Leitlinien beinhaltet konsequent auch die Beurteilung hinsichtlich einer Therapieansprache, die zumeist mit Untersuchungsgegenstand der Studie ist oder sogar einen der (meist sekundären) Endpunkte bildet. Entsprechend wird das bildgebende Diagnostikum in der Studie selbst definiert und häufig mit untersucht.

Den aktuell gängigen klinischen Standard zur Kontrolle des Ansprechens auf eine systemische Therapie bildet bei den soliden Tumoren, so auch beim NSCLC, die CT-Bildgebung. Die Evaluation der Therapieansprache, in Abgrenzung zu einem erneutem Progress, erfolgt dann zur besseren Objektivierbarkeit und Reproduzierbarkeit optimalerweise unter Anwendung der RECIST-Kriterien (31). Sie sehen eine morphologisch orientierte Einteilung der Tumorkläsionen anhand ihrer Größe vor und teilen die Therapieansprache in eine von vier möglichen Kategorien ein. Sie sehen eine Beurteilung der Erkrankung in regelmäßigen Abständen unter laufender systemischer Therapie, aber auch nach Therapieabschluss vor.

Da die Einschätzung des Verfahrens jedoch auf rein morphologischen Aspekten beruht, besitzt sie einen entscheidenden Nachteil: Sie wird dem unterschiedlichen Verhalten verschiedener Tumorentitäten unter laufender Therapie nicht gerecht, da in vielen Fällen das reine Volumen der Läsion nicht mit dem Grad der Therapieansprache korreliert (8). Gerade in Zeiten aufkommender neuer Therapiestrategien beim NSCLC,

wie der Bestimmung des Rezeptorstatus und zielgerichteter Therapie mittels Immunmodulatoren und Checkpoint-Inhibitoren, sind auch neue Strategien zur Bewertung einer metabolischen Ansprache gefragt. Eine Therapieansprache kommt hier nicht durch ein bloßes Schrumpfen des Tumors zum Ausdruck, sondern kann sich initial auch in Form eines sogenannten *Tumorflares* oder eines Pseudoprogress äußern, bei dem die Größe der Läsion initial sogar zunimmt (32). Beides beruht, einfach formuliert, auf einem Anschwellen des Tumors durch Einwanderung von Immunzellen und Zunahme des nekrotischen Gewebes. Die Evaluation solcher metabolischer Veränderungen unter Therapie, stellt die Domäne der FDG-PET/CT dar, vor allem in Abgrenzung zwischen therapeutischem Ansprechen, posttherapeutischer Veränderungen und erneutem Rezidiv (33). Im Bestreben auch hierfür objektivierbare und vom Untersucher unabhängige Ergebnisse erzielen zu können, wurden 1999 zunächst die EORTC-Kriterien (34), und 2009 dann die PERCIST-Kriterien definiert (8).

Bei den PERCIST-Kriterien wird in zwei Untersuchungen zu unterschiedlichen Zeitpunkten die jeweils höchste Intensität einer Läsion in einem Volumen von 1 ml bestimmt. Die Werte für diesen SUL_{peak} -Wert werden dann miteinander verglichen und anhand ihrer Änderung im Verlauf in vier Kategorien eingeteilt (35):

- „*Complete metabolic Response*“ (CMR) bedeutet in Analogie zu CR ein Verschwinden aller Läsionen oder kein vitales Tumorgewebe oberhalb der FDG-Intensität der Leber.
- „*Partial metabolic Response*“ (PMR) beschreibt einen prozentualen Abfall des SUL_{peak} von mehr als 30 % und einen absoluten Abfall um 0,8 Einheiten.
- „*Stable metabolic Disease*“ (SMD) erfüllt weder die Kriterien für PMR noch PMD; also eine Abnahme des SUL_{peak} im Follow-Up weniger als 30 % oder 0,8 Einheiten und gleichzeitig keine Zunahme um mehr als 30 % oder mehr als 0,8 Einheiten.
- „*Progressive metabolic Disease*“ (PMD) bezeichnet eine Zunahme des SUL_{peak} von mehr als 30 % oder 0,8 Einheiten gegenüber der Voruntersuchung oder eine neu aufgetretene Läsion.

Neben diesen Kriterien für die sogenannte Target-Läsion mit der höchsten Intensität sind noch weitere Kriterien für die Non-Target-Läsionen definiert, die beispielsweise bei einer individuellen Zunahme von mehr als 30 %, unabhängig von der Target-Läsion, ebenfalls eine PMD kennzeichnen.

Sowohl EORTC als auch PERCIST vergleichen PET-Scans vor und nach Therapie und ordnen die Ergebnisse den 4 gleichen Kategorien zu. Dabei unterscheiden sie sich jedoch unter anderem in der Auswahl der Zielläsion und den zugrunde liegenden Schwellenwerten zur Abstufung zwischen den Kategorien und den dafür verwendeten SUV-Parametern (9). Die PERCIST-Kriterien verwenden für die Intensitätsmessung im PET den SUL, was sie weniger fehleranfällig macht, beispielsweise gegenüber starken Gewichtsschwankungen der Patienten im Therapieverlauf (34). Darüber hinaus verhindert eine geforderte absolute Veränderung von mindestens 0,8 Einheiten ein Überschätzen von bereits im Baseline geringen SUL-Werten (9).

Für beide metabolischen Kriterien konnte in den vergangenen Jahren anhand zahlreicher Studien belegt werden, dass sie den klassischen RECIST-Kriterien überlegen sind (36-38). Dies beruht nicht zuletzt auf der besseren Differenzierung zwischen echtem Progress und Pseudoprogress (39). Obwohl in PET-basierten Studien meist beide Kriterien Anwendung finden, wird den PERCIST-Kriterien mittlerweile aufgrund ihrer einfacheren klinischen Anwendbarkeit, ihrer besseren Reproduzierbarkeit und klarer Definitionsweise der Vorzug gegeben (40). So greift auch diese Arbeit die PERCIST- und nicht die EORTC-Kriterien auf. Ein wesentlicher Unterschied zwischen beiden Kriterien bezüglich des Outcomes konnte bislang jedoch nicht gezeigt werden (41). Einigkeit der Studienergebnisse besteht jedoch darin, dass eine frühe Kontrolle mittels PET/CT nach Therapieabschluss ein Ansprechen besser einschätzen kann als konventionelle Bildgebung (7, 42).

Eine kürzlich publizierte Weiterentwicklung der PERCIST-Kriterien stellen die PREDIST-Kriterien dar, welche von einer italienischen Forschungsgruppe im Rahmen der Radiochemotherapie beim lokal fortgeschrittenen Rektumkarzinom definiert wurden (43). Sie basieren ebenfalls auf der Messung des SUL_{peak} -Wertes, weisen allerdings zwei essentielle Unterschiede zu den PERCIST-Kriterien in der Definition eines Progress auf. Einerseits definieren sie einen Schwellenwert in der Kontrolluntersuchung, dessen Überschreiten einen Progress definiert, andererseits benötigt man so in logischer Konsequenz keinen Vergleichswert aus der Voruntersuchung. Die PREDIST-Kriterien unterteilen lediglich in die zwei Kategorien *CMR* und *Residual Disease* (44). Die Definition des Schwellenwertes im Follow Up der PREDIST-Kriterien ist dabei dieselbe wie im Baseline der PERCIST-Kriterien und erfolgt anhand der basalen Stoffwechselsituation der Leber. Bereits Wahl *et al.* hatten diese Schwellenwertdefinition in ihrer ursprünglichen Publikation der PERCIST-Kriterien für das Follow Up diskutiert, sich dann aber für eine weichere Definition

entschieden, nach der eine Läsion im Follow Up lediglich über dem Niveau des Leberstoffwechsels liegen muss (8). Ein Vorteil des individuell am Stoffwechsel des Patienten definierten Schwellenwertes im Follow Up ist seine Unabhängigkeit von der Voruntersuchung und somit von Abweichungen in Gerätetechnik oder Zeitpunkt der Untersuchungen.

2.2.3. Herausforderung in der Rezidivdiagnostik

Für gewöhnlich erfolgt die Nachsorge der Patienten nach kurativer Radiochemotherapie in den ersten beiden Jahren in vierteljährlichen Abständen, danach halbjährlich. Neben der klinischen Untersuchung schlägt die Leitlinie zu diesen Terminen auch die Durchführung einer geeigneten Bildgebung vor, die in der klinischen Praxis meist durch eine CT erfolgt. Empfehlungen zur Anwendung der PET/CT gibt die Leitlinie zwar nicht explizit, hebt jedoch die hohe Sensitivität (99%) und Spezifität (89%) der PET/CT in der Diagnostik des Lokalrezidivs beim NSCLC hervor und unterstreicht den Vorteil der gleichzeitigen Erkennung neu aufgetretener Fernmetastasen (4).

Prinzipiell kann in der Rezidivdiagnostik des NSCLC zwischen lokoregionärem Progress, mediastinalem (*nodalem*) Progress und „*distant recurrence*“, also Fernmetastasierung, unterschieden werden (45). Der Begriff lokoregionär umfasst dabei in aller Regel den ipsilateralen Lungenflügel und Hilus, sowie die ipsilateralen mediastinalen Lymphknoten. Manche Autoren subsumieren das gesamte Mediastinum unter den Begriff *lokoregionär*, was in der Differenzierbarkeit hinsichtlich einer elektiven nodalen Bestrahlung nicht sinnvoll erscheint, jedoch die Unterscheidung zwischen *local* und *distant recurrence* erleichtert. Letzteres ist als Fernmetastase beim NSCLC definiert als Metastase im kontralateralen Hemithorax oder einer extrathorakalen Region (46, 47).

Die häufigsten Lokalisationen der Fernmetastasen sind in absteigender Reihenfolge meist die kontralaterale Lunge, das Skelett, das Gehirn, die Leber, die Nebennieren oder an unerwarteten Lokalisationen wie zum Beispiel einem abdominalen Lymphknoten. Sie alle können dem Betrachter bei geringer Größe in der CT entgehen oder nicht miterfasst sein. In der PET heben sich Fernmetastasen von FDG-aviden Tumoren wie dem NSCLC meist als FDG-avide Foci vom übrigen Bild ab und sind somit in der PET leichter zu diagnostizieren (42, 46). In der Regel ist die PET/CT also der reinen CT in der Diagnostik einer Fernmetastasierung auch beim Rezidiv

überlegen. Einzige Ausnahme stellt in manchen Quellen die Hirnmetastasierung dar, da sie sich in manchen Fällen nur wenig in der Intensität vom übrigen Hirnparenchym abhebt (46).

Auch bei der Beurteilung eines möglichen nodalen Mediastinalrezidivs sind rein morphologische Kriterien höchst unzuverlässig. So können bereits bestrahlte Lymphknoten längere Zeit noch vergrößert imponieren, während frühe nodale Metastasen keine Lymphknotenvergrößerung verursachen müssen (6). Anhand der FDG-PET/CT kann man hier differenzieren, welcher Lymphknoten einen hohen Glukosemetabolismus aufweist und somit ein Rezidiv vermuten lässt (7). Entsprechend wird die Sensitivität für die PET/CT im Vergleich zur CT deutlich höher beziffert (75 – 85 % *versus* 50 – 70 %). Entsprechendes gilt auch für die Spezifität der PET/CT gegenüber der CT (85 – 90 % *versus* 65 – 85 %) (46).

Im Vergleich zur CT bietet die FDG-PET/CT ebenfalls Vorteile bei der Bewertung hinsichtlich eines lokoregionären, pulmonalen Rezidivs, vor allem durch ihre Darstellung des Glukosemetabolismus im Lungenparenchym. So kann es unmittelbar nach einer Radiotherapie beispielsweise zu typischen pulmonalen Veränderungen eines frühen Bestrahlungs-assoziierten Lungenschadens kommen, welcher sich üblicherweise mit Milchglastrübungen, Konsolidierungen und Atelektasen äußert, die das Erkennen der morphologischen Tumorausdehnung in der reinen CT erschweren können (47, 48). Später gehen diese Veränderungen in einen fibrotischen Umbau des Lungenparenchyms über und Erschweren damit das Erkennen eines lokoregionären Tumorrezidivs innerhalb kompakter oder flächiger Narbenareale (48). In der Literatur gibt es einige Studien zur Risikostratifizierung dieser Veränderungen bezüglich des tatsächlichen Vorliegens eines Rezidiv mittels morphologischer CT Diagnostik (48, 49). Doch stößt die rein morphologische Darstellung hier an ihre Grenzen. Spricht ein Tumor zum Beispiel auf neue zielgerichtete Therapien zur zytostatischen Tumorkontrolle nicht mehr an, so bleibt ein mögliches „Wiederaufflammen“ des Tumormetabolismus bei reiner Größenbeurteilung möglicherweise zunächst unerkannt. Die PET kann diese Steigerung im Glukosemetabolismus aufzeigen (5, 50).

Dennoch gibt es in der Bewertung eines möglichen lokoregionären Rezidiv auch einige Limitationen in der FDG-basierten PET/CT-Diagnostik. Hierunter zählen postradiogene Veränderungen wie die PET Pneumonitis, Lungenfibrosierung, Pleurodese aber seltener auch granulomatöse Entzündungsreaktionen, wie zum Beispiel sarkomatoide Reaktionen bei der Behandlung mit Immunmodulatoren (46).

Die mit einer Radiotherapie assoziierte Pneumonitis tritt in der Regel innerhalb der ersten drei Monate nach Therapieabschluss auf (51), weswegen einige Autoren eine Therapieevaluation mittels FDG-PET/CT auch erst nach dieser Zeitspanne empfehlen (Abb. 2.4). Die Fibrosierung des bestrahlten Lungenparenchyms ist hingegen ein Phänomen, das in der Regel erst einige Monate später auftritt. Nach spätestens drei Jahren kann man aber bei bis zu 99 % der bestrahlten Patienten Dichteveränderungen des Lungenparenchyms messen (52). Aufgrund dieser Veränderungen werden in der Rezidivdiagnostik häufig nur fokale FDG-Akkumulationen im Parenchym gewertet, während diffuses Enhancement des Lungenparenchyms oder der Pleura nach Abschluss einer Radiotherapie häufig als unspezifisch entzündlich gewertet werden kann (42). Generell scheint eine Unterscheidung zwischen posttherapeutischen Veränderungen und einem Rezidiv nicht anhand fest definierter Kriterien auszumachen zu sein. Die Studie einer australischen Arbeitsgruppe kommt zum Schluss, dass eine Unterscheidung generell möglich sei, wenn der Observer auf eine strikte interne Konsistenz seiner Bewertung achte und ihm alle nötigen klinischen Informationen über die Therapie des Patienten vorlägen (51). Sie stellten fest, dass eine vorliegende Pneumonitis die Rezidivdiagnostik zwar erschwere, jedoch keinen Impact auf die tatsächliche Diagnose eines Rezidivs hat. Die Beobachtung, dass das Ausmaß an Pneumonitis und Bestrahlung miteinander korrelierten, ließ die Gruppe vermuten, dass die Radiosensitivität von Tumor und normalem Gewebe in Zusammenhang stehen müsse. Es erlaubt jedoch keinen direkten Rückschluss darauf, ob die Intensität der Pneumonitis in der FDG-PET/CT auch mit dem Ansprechen der Therapie in Verbindung steht (53). Dieselbe Arbeitsgruppe konnte ihre eigenen Ergebnisse später nochmals in einer weiteren Studie reproduzieren und Einteilungen des unterschiedlichen Erscheinens von einer Strahlenpneumonitis in der PET/CT im ersten halben Jahr nach Radiotherapie definieren. Die Veröffentlichung hebt noch einmal hervor, dass speziell in zweideutigen und fragwürdigen Fällen, eine qualitative Auswertung gegenüber der quantitativen Auswertung (z. B. mittels PERCIST) im Vorteil sein kann (51).

Einheitliche, quantitative Messkriterien zur Diagnostik eines Rezidiv in der PET/CT gibt es – im Unterschied zur Bewertung der Therapieansprache – jedoch bis heute nicht.

Therapieform	Zeitintervall
Chemotherapie	10 Tage
Operation	6 Wochen
Radiofrequenzablation	6 - 8 Wochen
Immuntherapie	8 - 9 Wochen
Radiotherapie	3 Monate

Abbildung 2.4 Empfehlungen zum Zeitpunkt der Kontrolle der Therapieansprache mittels FDG-PET/CT nach Kandathil *et al.* (46).

2.2.4. Lokal- versus Fernrezidiv

In der Regel ist mit den Begriffen des lokalen Progress und des lokoregionären Rezidiv dasselbe, in unterschiedlichen Zeiträumen während und nach der Therapie gemeint. Die Abgrenzung zur Fernmetastase fällt hingegen leichter und ist zeitlich unabhängig zur Behandlung. Um die zeitlich aufwändige, visuelle Auswertungen in dieser Arbeit zu erleichtern, wurden die intrathorakalen Metastasen im jeweiligen Bereich (pulmonal oder mediastinal) mit abgefragt. Erst später, zum Vergleich mit dem Goldstandard zum tatsächlichen Outcome der Patienten wurde dann lediglich zwischen lokoregionärem Rezidiv (pulmonal und/oder mediastinal) und Metastase (also *distant recurrence*) unterschieden.

In einer Analyse chirurgisch therapierter Patienten mit NSCLC konnte 2010 gezeigt werden, dass es sowohl für das lokoregionäre Rezidiv als auch für das Rezidiv in Form der Fernmetastase einen ersten Peak nach etwa einem Jahr nach Therapieabschluss gibt (54). In der heutigen Zeit wird bei immer besser werdender lokaler Kontrolle durch die Strahlentherapie, das Rezidiv in Form der neu auftretenden Fernmetastasierung und dessen Therapie immer relevanter. Insbesondere aufgrund der vielfältigen neuen, Therapiemöglichkeiten, mit Immunmodulatoren und zielgerichteter molekularer Therapie (42).

In der Literatur finden sich Zahlen, die besagen, dass etwa 34 – 50 % der Patienten mit NSCLC einen lokoregionären Progress erleiden und bei 50 – 66 % Fernmetastasen auftreten. Betrachtet man explizit nur das UICC-Stadium III, so werden in älteren Studien bei bis zu 85 % der Patienten nach Radiochemotherapie lokale Rezidive diagnostiziert (46). Diese Zahlen dürften mit zunehmend genauerer Applikation der Radiotherapie in den letzten Jahren etwas gesunken sein und wie in der Studie von

2. Literaturdiskussion. Herausforderungen in der modernen PET/CT-Diagnostik

Cauro *et al.* eher bei 50% liegen (45). Die 2015 von Consonni *et al.* publizierten Rezidivraten der Teilnehmer der EAGLE-Studie bestätigen die Annahme, dass ein Rezidiv, unabhängig vom UICC-Stadium, in Form einer Fernmetastasierung wahrscheinlicher sei, als in Form eines Lokalrezidivs (55). Gleichzeitig implizieren die veröffentlichten Zahlen zur Mortalität nach einem Rezidiv, die Wichtigkeit einer frühen Diagnose, um den Patienten möglichst schnell einer adäquaten Therapie zuführen zu können. Sodass gelten sollte, die FDG-PET/CT in der Rezidivdiagnostik spätestens dann in Betracht zu ziehen, wenn in der CT Diagnostik der Verdacht eines möglichen Rezidiv geäußert wird (42). Die Routinebildgebung in der Nachsorge nach Therapieansprache sollte aber vorerst weiterhin durch die CT erfolgen (56).

2.3. Herausforderungen in der modernen PET/CT-Diagnostik

2.3.1. Die qualitative und quantitative Analyse von PET-Informationen

Bei der Einordnung von Bilddaten aus der PET kann man im Wesentlichen zwischen qualitativen und (semi-) quantitativen Informationen unterscheiden (9).

Einen qualitativen Eindruck gewinnt der Untersucher daraus, dass die Aktivitätsbelegung des Gewebes im Körper aus den Koinzidenzmessungen entsprechend berechnet und durch eine Farbskala, nach Intensität kodiert, in ein Bild umgewandelt, übertragen wird. Für reine PET-Bilder ist hierfür eine Graustufenskala, für fusionierte PET/CT-Bilder meist eine Farbskala üblich (Abb. 3.2 A, C). Stärkere Anreicherungen des radioaktiv-markierten Tracers im Gewebe führen dann zu einer farblich unterscheidbaren Kodierung im Vergleich zu anderem Gewebe. Der Untersucher kann diese Anreicherung dann beschreiben in dem er diese optisch mit Nachbarstrukturen, dem Blutkreislauf oder anderen Organen, wie zum Beispiel der Leber, vergleicht. Er kann zusätzlich auch beschreibende Einteilungen vornehmen und dadurch zum Beispiel fokale von flächigen oder diffus verteilten Anreicherungen abgrenzen. Bei all diesen Merkmalen liegt die Stärke ihrer Ausprägung sozusagen jedoch „im Auge des Betrachters“ und sie sind in ihrer Wahrnehmung somit subjektiv.

Einen Ansatz, die Erhebung dieser qualitativen Merkmale der FDG-PET/CT objektivierbar und reproduzierbar zu gestalten, unternehmen beispielsweise die Deauville-Kriterien in der PET-Bildgebung von Lymphomen (57). Sie stellen ein einfaches Werkzeug zur visuellen Interpretation einer sichtbaren FDG-Akkumulation dar, in dem diese mit zwei relativ konstanten Referenzpunkten verglichen und einer

2. Literaturdiskussion. Herausforderungen in der modernen PET/CT-Diagnostik

Fünf-Punkte-Skala zugeordnet wird (58). Als Referenzpunkte dienen hierbei der mediastinale Blutpool und der basale Stoffwechsel der Leber.

Einem ähnlichen Prinzip folgen die Peter Mac-Kriterien in der Bewertung der Therapieansprache bei soliden Tumoren (7). Für diese Einteilung ist eine sequentielle Bildgebung notwendig. Sie sieht einen Vergleich zwischen dem Ausgangsbefund (Baseline-PET) und der erneuten Bildgebung nach Therapie (Follow-Up-PET) vor. Der Befunder muss eine graduelle Einteilung des Unterschiedes zwischen beiden Bildgebungen vornehmen und das Ergebnis des Follow-Up-PET einer von vier Kategorien (*complete response, partial response, no response, progressive disease*) zuordnen.

Um die qualitative Bildinformation der FDG-PET/CT einem quantitativen Wert zuordnen zu können, wurde der „Standard Uptake Value“ (SUV) definiert. Er berechnet sich wie folgt (59):

$$\text{SUV} = \frac{\text{Aktivitätskonzentration} \left[\frac{\text{Bq}}{\text{g}} \right] \times \text{Körpergewicht} [\text{g}]}{\text{applizierte Aktivität} [\text{Bq}]}$$

Der SUV ist also dimensionslos und bezieht sich klassischerweise auf das Körpergewicht. In der Literatur findet sich hierfür gelegentlich der Zusatz SUV_{bw} für *body weight*. Der Wert ist abhängig von der Aktivitätskonzentration und diese wiederum vom Körpergewebe und Zeitpunkt der Messung. Bei theoretischer, gleichmäßiger Verteilung des Tracers im gesamten Körper würde sein Wert überall genau 1 betragen. Da die Distribution des Tracers im Körper aber inhomogen, entsprechend des jeweiligen Gewebes verläuft, werden Werte zwischen 1 und 2 für gewöhnlich als normal betrachtet. Ein Wert deutlich über 2 ist meist pathologischer Genese. Mithilfe des SUV können, bezogen auf eine bestimmte dreidimensionale Region (VOI) in der PET, der Mittelwert der FDG-Akkumulation (SUV_{mean}) und der Wert des Voxels mit der höchsten Intensität (SUV_{max}) bestimmt werden. Obwohl dieser Wert auf der Basis eines einzelnen Voxels zunächst sehr fehleranfällig erscheint, konnte in zahlreichen Studien gezeigt werden, dass ein hoher SUV_{max}-Wert mit einer hohen Wahrscheinlichkeit für Malignität assoziiert ist (60-62). Er zeigt jedoch eine hohe Fehleranfälligkeit bei erhöhtem Bildrauschen und kann in seinem Wert zwischen verschiedenen PET-Scannern variieren (35). Außerdem überschätzt ein hoher SUV_{max} den FDG-Uptake partiell nekrotischer oder zerfallender maligner Läsionen,

2. Literaturdiskussion. Herausforderungen in der modernen PET/CT-Diagnostik

wie zum Beispiel nach beginnender Chemotherapie (9). Auch der Bezug des SUV auf das Körpergewicht kann zum Beispiel bei adipösen Patienten eine Fehlerquelle darstellen. Der sogenannte *Standard Uptake Value corrected for Lean Body Mass* (SUL, auch SUV_{lbm}) bezieht diese potentielle Fehlerquelle mit ein und stützt sich in seiner Berechnung im Gegensatz zum SUV auf die fettfreie Körpermasse. Diese wird üblicherweise anhand der Formel nach James berechnet (63). Aus ihm lässt sich der SUL_{peak} ableiten. Dazu nutzt man entweder entsprechende Softwarelösungen oder bewegt iterativ ein vordefiniertes, kugelförmiges VOI mit dem Durchmesser von 1,2 cm ($\cong 1$ ml) über die Läsion mit der höchsten Intensität in der PET bis man den höchsten Durchschnittswert (den *Peak*) erreicht. Der SUL_{peak} berechnet sich dann als die Summe der höchsten SUL-Werte der Voxel mit insgesamt 1 ml Volumen geteilt durch die Anzahl der Voxel und stellt somit den Durchschnitt der höchsten SUL-Werte im PET innerhalb eines Volumens dar (35). Im Gegensatz zum SUV_{max} bezieht sich der SUL_{peak} also nicht nur auf ein singuläres Voxel, sondern auf mehrere Voxel mit der höchsten Intensität und insgesamt 1 ml Volumen. Das macht ihn insgesamt weniger störanfällig für Bildrauschen. Er ist aufgrund der Inklusion mehrerer Voxel auch robuster gegenüber Ausreißern und weniger empfindlich gegenüber größerer Voxelgrößen (8, 35). Das Argument, der SUV_{max} sei im Gegensatz zum SUL_{peak} vom Observer unabhängig bestimmbar, relativiert sich in Zeiten Software-gestützter semiautomatisierter Auswertungsmöglichkeiten (9).

Anhand der mittleren SUV-, beziehungsweise SUL-Werte kann man nun für bestimmte Läsionen, wie zum Beispiel soliden Tumoren, weitere Werte zur genaueren Charakterisierung berechnen. Definiert man einen Schwellenwert und umkreist alle Voxel einer Tumorkläsion, die oberhalb dieser Schwelle liegen, so erhält man das metabolische Tumolvolumen (*metabolic tumor volume*, MTV). Für die Definition der Schwelle gibt es bislang keinen einheitlichen Standard, in den meisten Studien findet jedoch der SUL_{mean} der Leber plus dessen zweifache Standardabweichung Anwendung (30, 35). Multipliziert man das MTV anschließend mit dem mittleren SUL (SUL_{mean}) der Läsion, erhält man die *total lesion glycolysis* (TLG). Sie ist ein virtuelles Maß für die Tumorklast im Körper. Die TLG für die Gesamttumorklast im Körper bei mehreren einzelnen fokalen Läsionen, zum Beispiel Primärtumor und Metastasen, berechnet sich aus der Summe aller einzelnen TLG. Auf Basis dieser (semi-)quantitativen PET-Parameter existieren Klassifikationssysteme zur Therapieansprache, wie die bereits erwähnten EORTC-, PERCIST- oder PREDIST-Kriterien (8, 34, 43).

2.3.2. Die Interobservervariabilität in der PET/CT-Diagnostik

Die qualitative Beurteilung einer PET/CT-Diagnostik hängt immer vom auswertenden Arzt und somit in logischer Konsequenz von der Gewohnheit der Auswertung und der Erfahrung des Arztes ab (9). Die Art der Bildrekonstruktion und ihre technische Auslesung variiert ferner zwischen den jeweiligen Instituten, welche die Bilder erstellen. Quantitative Messwerte, wie der SUVmax, sind dabei zwar unabhängig vom gemessenen Volumen, können aber in ihrer Rekonstruktion durch den Computer und in ihrer Interpretation durch den Menschen Fehlern unterliegen (58). Auch erschweren posttherapeutische Veränderungen häufig die Abgrenzung zu einem Rezidiv und sorgen so für eine hohe Variabilität in der Interpretation gleicher Bilder durch verschiedene *Observer*.

Quantitative Auswertelgorithmen, wie die PERICST-Kriterien, sollen dazu beitragen diese Variabilität der subjektiven Beurteilung zu reduzieren und die Ergebnisse der PET/CT in klinischen Studien vergleichbarer zu gestalten. So konnten in Studien hohe Übereinstimmungswerte für die Auswertungen anhand der PERCIST- und EORTC-Kriterien durch zwei unterschiedliche *Observer* gezeigt werden, wie zum Beispiel bei Turgeon et al. mit 95 % und 98 % (64). Außerdem ist die Übereinstimmung zwischen beiden Kriterien in der Beurteilung einer Therapieansprache mit 95 % ebenfalls als fast perfekt einzuschätzen (41).

Die PERICST-Kriterien sind so konstruiert, dass sie in ihrer Interpretation bewusst wenig Spielraum für subjektive Interpretation lassen, um sie so möglichst reproduzierbar zu gestalten. Die Limitation der subjektiven Interpretation der Befunde ist durch die dadurch fehlende Diskriminierung zwischen Pneumonitis- und Rezidiv-typischen Veränderungen aber auch gleichzeitig ihr größter Nachteil (51). Von der anderen Seite betrachtet, kann ihre Implementierung in die subjektive Auswertung jedoch dazu beitragen diese zu verbessern, wie Fledelius *et al.* zeigen konnten (65). Fledelius *et al.* stellten in ihrer Untersuchung zwischen rein visueller Interpretation und der Auswertung nach PERCIST-Schema durch 8 *Observer* (Assistenzärzte) fest, dass die Übereinstimmung zwischen den *Observern* bei PERCIST höher lag. Sie berichten von Kappa-Werten von 0,76 für PERCIST und von 0,60 für eine visuelle, subjektive Interpretation. Hingegen kommen Turgeon *et al.* zu dem Schluss, dass vorher fest definierte visuelle Kriterien zu einer höheren Übereinstimmung zwischen den *Observern* führen, als EORTC- und PERCIST-Kriterien. Sie stellten fest, dass Abweichungen der PERCIST-Kriterien vor allem in der Differenzierung zwischen Progress und benignen postradiogenen Veränderungen zustande kommen und

schließen so daraus, dass visuelle Kriterien besser zwischen diesen differenzieren (64). Sie betonen dabei auch die notwendige Expertise und die Erfahrung des befundenden Arztes, was gleichzeitig auch den wichtigsten Erklärungsansatz für die Diskrepanz der Ergebnisse zwischen beiden Veröffentlichungen darstellt. Die fast perfekten Übereinstimmungswerte ($\kappa = 0,95$) von denen hier berichtet werden, dürften dabei nicht die klinische Routine widerspiegeln, die von Wahl *et al.* für das mediastinale Staging des NSCLC ($\kappa = 0,65$) eher im Bereich der moderaten Übereinstimmung angesiedelt werden (8).

Ein guter Kompromiss liegt dabei in einem Mittelweg zwischen beiden Kriterien. So konnten Nestle *et al.* zeigen, dass ein Harmonisierungsprozess mit mehrmaliger Fehleranalyse und Reevaluation definierter Auswertekriterien dazu beitragen kann, die Interobservervariabilität deutlich zu verbessern. Die hier berichteten, zu Anfangs lediglich moderaten Übereinstimmungswerte ($\kappa = 0,53$) führten zur Diskussion der Fehlerquellen und deren Identifizierung zu verbesserten Interpretationskriterien. Nach Implementieren der verbesserten Kriterien, verbesserte sich auch das Interobserveragreement in der Auswertung ($\kappa = 0,66$), wobei als unsicher gewertete Beurteilungen weiterhin mit niedrigen Übereinstimmungswerten assoziiert waren (66). Diese Arbeit soll an die Erkenntnisse dieses Harmonisierungsprozesses anknüpfen und nutzt diese in der *Observer*-Auswertung.

2.4. Die PET-Plan Studie

Die PET-Plan Studie ist eine prospektive, multizentrische Phase-III Studie, deren Ergebnisse Anfang des Jahres 2020 im *Lancet Oncology* veröffentlicht wurden (22). Sie untersucht die Möglichkeit einer Zielvolumenreduktion mithilfe einer PET/CT-basierten Bestrahlungsplanung in der Therapie von inoperablen Patienten mit lokal fortgeschrittenem nicht-kleinzelligem Lungenkarzinom in den Stadien II und III. Also jene Patienten, bei denen trotz fortgeschrittenem Stadium noch eine kurative Therapieintention mittels kombinierter Radiochemotherapie besteht. Eben diese Patienten profitieren gleichzeitig aber auch von einem möglichst niedrigen Nebenwirkungsprofil einer Bestrahlung der besonders strahlensensiblen Organe im Thorax, wie dem Herzen oder der Lunge. Sodass sich in Zeiten der immer genauer werdenden Bestrahlungapplikation zum Einen die Frage stellt, wie groß das lokale

Bestrahlungsvolumen ausfallen muss, um ein möglichst niedriges Risiko für ein lokales Rezidiv gewährleisten zu können. Zum Anderen - und fast noch essentieller - ist jedoch die Frage, wie die Bestrahlungsplanung die exakteste und somit beste Möglichkeit bietet das klinisch relevante Tumolvolumen zu definieren. So scheint es bei fortschreitender Technik in der Bestrahlung nur logisch auch in der Technik zur Bestrahlungsplanung weiter voranzuschreiten und die moderne PET/CT-Bildgebung der konventionellen CT-basierten Bestrahlungsplanung vergleichend gegenüber zu stellen.

In einem Studiendesign auf Nichtunterlegenheit wurden die Patienten hierfür in einem 1:1-Verhältnis randomisiert und zwei Therapiearmen zugeteilt:

Im konventionellen Arm A erfolgte die Therapieplanung der Zielvolumina sowohl mit den Informationen aus der CT-, als auch der FDG-PET/CT-Untersuchung. In das Bestrahlungsfeld miteinbezogen wurden neben dem Primärtumor auch nachgeschaltete Atelektasen. Als mediastinale Lymphknotenmetastasen galten jene Lymphknotenstationen, die entweder in der PET/CT FDG-avid (PET-positiv) waren, deren Kurzachsendurchmesser in der CT über ein Zentimeter betrug (CT-positiv), die durch eine Biopsie gesichert wurden oder die abhängig von der Lokalisation des Primärtumors als wahrscheinlich betroffen galten (26). Diese wurden den entsprechenden mediastinalen Lymphknotenlevel nach Mountain-Dressler zugeordnet. Die Zuordnung zur entsprechenden Lymphknotenstation erfolgte dabei anhand des Atlas nach *Chapet et al.* (67). Alle als betroffen gewerteten Lymphknotenstationen wurden in die Bestrahlungsplanung miteinbezogen und ebenfalls bestrahlt.

Im experimentellen Arm B erfolgte die Bestrahlungsplanung des Primärtumors und der mediastinalen Lymphknoten rein anhand der PET-positiven Befunde und, falls vorhanden, histologisch gesicherter Lymphknotenmetastasen. PET-negative, aber in der CT vergrößerte Lymphknoten wurden nicht miteinbezogen und somit nicht elektiv bestrahlt.

Für die Zielvolumenplanung mittels PET/CT stand in beiden Fällen ein zuvor von derselben Arbeitsgruppe publizierter semiautomatischer Konturierungsalgorithmus zur Verfügung (68). Die Radiatio erfolgte in beiden Armen mittels dosiseskaliertem IMRT oder 3D-CRT mit einer Dosis von 60 – 74 Gray als fraktionierte Bestrahlung (à 2 Gray) begleitet von einer platinhaltigen systemischen Chemotherapie. Die elektive Bestrahlung des Mediastinums in Arm A erfolgte mit einer Dosis von 50 Gy/2 Gy.

Die Ergebnisse, der zur Bestrahlungsplanung an verschiedenen Zentren angefertigten PET/CT-Untersuchungen, unterlagen nach einer initialen Auswertung durch die Zentren selbst, in einem zweiten Schritt dem Urteil eines vorher einheitlich geschulten Experten. Dieser entstammte einem Gremium, bestehend aus 11 erfahrenen Nuklearmedizinern, welche zuvor einen gemeinsamen Harmonisierungsprozess zur verbesserten Befundauswertung durchliefen (66). Dieses Prozedere konnte die Übereinstimmung in der Auswertung der mediastinalen Lymphknotenlevel nachweislich verbessern und betont die elementare Wichtigkeit gemeinsamer und einheitlich definierter Kriterien zur Auswertung qualitativ interpretierbarer Bildgebung. Außerdem konnte diese Form der Qualitätssicherung dazu beitragen, die Bestrahlungsplanung reproduzierbar und einheitlicher zu gestalten und trug somit wesentlich zur besseren Vergleichbarkeit der Ergebnisse bei (69). Die Auswertestrategien dieser Dissertation bauen auf diesen Erkenntnissen auf und nutzen die im Harmonisierungsprozess definierten Strategien zur Bewertung und Zuordnung von FDG-Akkumulationen in der PET-Bildgebung um möglichst nahtlos an die guten Übereinstimmungswerte in der visuellen Auswertung anzuknüpfen.

Der primäre Endpunkt der PET-Plan Studie wurde gebildet durch die Bestätigung eines lokoregionären Progress durch die FDG-PET/CT. Overall Survival (OS) und progressionsfreie Überlebenszeit (inklusive neu aufgetretener Metastasen) bildeten die sekundären Endpunkte. Hierfür erfolgte die Nachbeobachtung der Patienten über einen medianen Zeitraum von 29 Monaten. Im diesem Rahmen bekamen die Patienten regelmäßige CT-Untersuchungen. War hier nicht zweifelsfrei zu klären, dass kein Rezidiv vorlag, bekamen die Patienten zur Bestätigung oder Ablehnung der Diagnose zusätzlich eine weitere PET/CT-Untersuchung. Auch die Auswertung dieser Untersuchung erfolgte durch die Unterstützung des hierfür einheitlich geschulten Expertengremiums und floss maßgeblich in die Definition des Goldstandards der Studie mit ein.

Die Ergebnisse der PET-Plan Studie konnten zum einen zeigen, dass eine Zielvolumenreduktion mithilfe einer PET-basierten Bestrahlungsplanung keine Nachteile hinsichtlich des lokalen Rezidivrisikos beim NSCLC birgt. Zum Anderen stützen sie, in Bezug auf die sekundären Endpunkte, ebenfalls die Hypothese, dass eine elektive Bestrahlung des Mediastinums keinen zusätzlichen Benefit erbringen könnte. Vielmehr erkaufte man sich dadurch die Nebenwirkungen einer höheren Bestrahlungsdosis, welche sich dann negativ auf die Gesamtüberlebenszeit auswirken

könnten. Dies betrifft vor allem die Nebenwirkungen an besonders strahlensensiblen Thoraxorganen wie dem Herzen, dem mediastinalen Lymphabflusssystem oder dem umgebenden gesunden Lungenparenchym, in Form von postradiogenen entzündlichen Veränderungen. Wenngleich eine Verbesserung der Gesamtüberlebenszeit durch die Zielvolumenreduktion gegenüber der konventionellen Bestrahlung letztendlich nicht gezeigt werden konnte, so hat die Studie dennoch eine weitere wichtige Kernaussage: Nämlich, dass der Schlüssel zu einer exakteren Applikation der Bestrahlungsdosis nicht in der rein morphologischen Planung, sondern in der funktionell metabolischen Zielvolumenplanung liegen könnte.

Deutlich wird durch die Korrelation des initialen Tumorzvolumens mit dem Outcome der Patienten ebenfalls, dass dieses nicht nur einen wichtigen prognostischen Faktor hinsichtlich der Wahrscheinlichkeit eines lokoregionären Rezidiv spielt, sondern auch Einfluss auf die Überlebenszeit nimmt, wie bereits in Studien zuvor gezeigt werden konnte (70). Nicht diskutiert werden in diesem Zusammenhang hingegen die Rolle der aus der PET/CT ableitbaren funktionellen Tumorcharakteristika (s. Kapitel 2.4.1). Diese lassen sich hervorragend aus den, im Rahmen der Studie nach standardisierten Protokollen, einheitlich durchgeführten Untersuchungen zur Bestrahlungsplanung bestimmen und weiter analysieren. In dieser Arbeit wird dabei zum einen der Einfluss der PET-Parameter vor Therapie auf die Wahrscheinlichkeit eines Rezidiv überprüft. Zum anderen wird die Aussagekraft der PET-Parameter im *Follow Up* zum Vorliegen eines möglichen Rezidiv beleuchtet.

Neben der Schlussfolgerung einer aus onkologischer Sichtweise vertretbaren Zielvolumenreduktion bei gleichwertiger lokaler Kontrolle und verminderter Toxizität im Vergleich zur konventionellen Bestrahlung, lässt die PET-Plan Studie also auch die Feststellung zu, dass hierfür der metabolischen Tumorbildgebung mittels FDG-PET/CT eine sehr bedeutende Rolle zukommt. Der logische Schluss daraus ist, die Wertigkeit der PET/CT mit Hilfe der im Verlauf akquirierten PET/CT-Untersuchungen auch über die Bestrahlungsplanung hinausgehend zu untersuchen. Denn die PET-Plan Studie definiert in diesem Zusammenhang zwar eine einheitliche Zuordnung zur Lokalisation eines Rezidiv, lässt aber in der Interpretation über das tatsächliche Vorliegen eines Rezidiv und die Abgrenzung zu rein entzündlichen Prozessen (aus nuklearmedizinischer Sicht) sehr viel Spielraum. Somit ist das Ergebnis der Auswertung trotzdem subjektiv geprägt und hängt stark von der Erfahrung des auswertenden Arztes ab. Um stärker auf diese Thematik einzugehen, wurden die PET-Bildgebungen bei Verdacht auf Rezidiv in dieser Arbeit erneut ausgewertet. Für die

visuelle, qualitative Form der Auswertung durch die *Observer* wurden vorher Zuordnungskriterien für das Vorliegen eines pulmonalen Rezidiv definiert.

Diese sollten es ermöglichen genauer zwischen Rezidiv und entzündlichem Prozess zu differenzieren und somit zu einer möglichst einheitlichen Interpretation der FDG-PET/CT beitragen. Zweifelsfrei bietet die PET-Plan Studie zur Überprüfung dieser Differenzierung das optimale Patientenkollektiv. Die Form der klinisch gängigen, visuellen Auswertung wird in dieser Arbeit außerdem mit einer semiautomatisierten, quantitativen Auswertung verglichen um zu untersuchen wo die jeweiligen Vorteile und Limitationen einer subjektiven und objektivierbaren Auswertung der PET/CT in der Rezidivdiagnostik nach Radiochemotherapie liegen.

3. Material und Methoden

In dieser Arbeit wurde zur Untersuchung der in der Zielsetzung formulierten Fragestellungen ein Teil des Patientenkollektivs einer prospektiv dokumentierten, randomisierten Studie verwendet. Unter erneuter, nun retrospektiver Betrachtung der PET/CT-Bilddatensätze des ausgewählten Kollektivs, wurden diese von vier *Readern* hinsichtlich der Frage nach einem Rezidiv, beziehungsweise einer lokalen Progression, anhand qualitativer Merkmale ausgewertet. Die zusätzlich vorgenommene, quantitative Auswertung nach PERCIST-/PREDIST-Protokoll erfolgte mittels semiautomatischer Schwellenwertanalyse durch eine speziell hierfür konzipierte, kommerzielle Softwarelösung.

3.1. Die Auswahl der Daten aus der PET-Plan Studie

Die in dieser Arbeit verwendeten Patienten- und Bilddaten entstammen dem Kollektiv der oben beschriebenen PET-Plan Studie. Die Verwendung der Daten innerhalb dieser retrospektiven Datenauswertungen fand unter Zustimmung der Studienleitung und im Rahmen des positiven Ethikvotum für die Studie und deren Teilprojekte statt. Die ursprüngliche PET-Plan Studie ist eine prospektive, multizentrische, Phase III-Studie an Zentren in Deutschland, Österreich und der Schweiz (PET-Plan; ARO-2009-09). Die Rekrutierung der insgesamt 311 Studienteilnehmer erfolgte von Mai 2009 bis Dezember 2016 anhand der im Studienprotokoll vordefinierten Einschluss- und Ausschlusskriterien (Tabelle 3.1). Die Registrierung der Studie ist mit der Kennung NCT00697333 unter [ClinicalTrials.gov](https://clinicaltrials.gov) einsehbar. Die Veröffentlichung der Studienergebnisse erfolgte im April 2020 in der Fachzeitschrift *Lancet Oncology* (22).

Innerhalb der Studie erfolgte eine Nachsorge der Patienten mittels körperlicher Untersuchung, Laborwertkontrolle und CT-Bildgebung nach Leitlinie jeweils 3, 6, 9, 12, 18, 24, 36 und 48 Monate nach Beendigung der Radiochemotherapie. Entstand bei einer der Kontrolluntersuchungen der Verdacht auf ein Rezidiv der Erkrankung, so wurde eine weitere Bildgebung mittels PET/CT durchgeführt.

3. Material und Methoden. Die Auswahl der Daten aus der PET-Plan Studie

Einschlusskriterien	Ausschlusskriterien
NSCLC, histologisch bestätigt	Neuroendokrine Tumordifferenzierung
UICC-Stadium II-III	Pancoast Tumor
Nach interdisziplinärem Konsensus inoperabel	Supraclaviculäre Lymphknotenmetastasen
ECOG \leq 2	Maligner Pleura- oder Perikarderguss
Älter als 18 Jahre	Vorweg erfolgte Chemotherapie
Bestrahlungsplanung nach Studienprotokoll durchführbar	Überlagernde Entzündung in der Planungs PET/CT
Simultane Radiochemotherapie möglich	
Lungenfunktionsparameter FeV1 > 1l oder > 35%	

Tabelle 3.1 Einschluss- und Ausschlusskriterien der Patientenkohorte der PET-Plan Studie.

Für die Auswertungen zu dieser Arbeit wurden zu Beginn aus den insgesamt 311 Studienteilnehmern der PET-Plan Studie diejenigen ausgesucht, bei denen im Verlauf des Nachbeobachtungszeitraumes nach Therapieabschluss in der CT-Bildgebung der Verdacht auf ein Rezidiv geäußert wurde und für die laut klinischem Dokumentationssystem eine anschließende PET/CT-Bildgebung stattfand. Dies traf auf 98 der 311 Patienten aus fünf Zentren zu. Von den fünf angefragten Zentren betrug der Rücklauf aus vier Zentren insgesamt 122 PET/CT-Bilddatensätze von insgesamt 89 Patienten. Hiervon mussten weitere 8 Bilddatensätze von 7 Patienten ausgeschlossen werden, bei denen entweder die Daten zur Bestrahlungsplanung unvollständig waren oder die Koregistrierung von CT- und PET-Komponente nicht gelang. Insgesamt verblieben somit 114 PET/CT-Bilddatensätze (bestehend jeweils aus der PET/CT zur Bestrahlungsplanung und der PET/CT bei Verdacht auf Rezidiv) von 82 Patienten zur qualitativen und quantitativen Auswertung. Die Datensätze wurden pseudonymisiert und gemischt. Dabei wurde darauf geachtet, dass zwei Follow-Up-Untersuchungen desselben Patienten in der Abfolge der pseudonymisierten Daten nicht zu nah beieinander lagen. Jedem Follow-Up Datensatz bei Rezidivverdacht wurde das entsprechende Baseline PET/CT pseudonymisiert zugeordnet (beispielweise PAT001 für das Follow-Up- und BAS001 für das korrespondierende Baseline-PET/CT).

3.2. Gerätetechnik und Bildakquisition

Zur Verwendung kamen in dieser Studie koregistrierte Bilddatensätze von PET/CT-Kombinationsgeräten aus 4 verschiedenen Zentren (Siemens Biograph 16, Philips Gemini GXL, Philips Gemini TF16). Die Anfertigung der Bilder folgte dem im Studienprotokoll einheitlich definierten, gängigen klinischen Ablauf zur Akquise von Ganzkörper-PET/CT-Bildern beim nüchternen Patienten (Glukosekonzentration im venösen Blut <150 mg/dl). Der PET-Scan erfolgte ≥ 60 min nach Injektion der im jeweiligen Zentrum üblicherweise verwendet Aktivität der ^{18}F -FDG. Die Positionierung des Patienten entsprach der Lagerung während der Bestrahlung mit über den Kopf erhobenen Armen.

3.2.1. Grundlagen der PET/CT

Die Bilderzeugung eines Computertomographen beruht auf den Prinzipien der Röntgendiagnostik, bei der mithilfe einer Röntgenröhre elektromagnetische Strahlung erzeugt wird. Die Schwächung der Strahlung im Patienten erlaubt es Rückschlüsse auf die Gewebezusammensetzung des Körpers zu ziehen. Aus diesen Informationen berechnet ein Computer anschließend virtuelle Bilder des Patientenkörpers. Im Gegensatz zu einer planaren, zweidimensionalen Röntgenthoraxaufnahme entsteht als Endergebnis ein dreidimensionaler, aus Volumenelementen (Voxeln) aufgebauter Schichtdatensatz, welcher am Bildschirm eines Computers in jeder Raumrichtung betrachtet werden kann.

Die Positronenemissionstomographie (PET) bildet im Gegensatz zur CT nicht anatomische, sondern metabolische Informationen aus dem Körper des Patienten ab. Sie beruht auf der Koinzidenzmessung von Energie, die beim Zerfall radioaktiver Isotope entsteht. Geschieht dieser als β^+ -Zerfall, wird ein Positron emittiert, welches im Abstand einiger Millimeter (abhängig von seiner Energie, welche wiederum vom freisetzenden Isotop abhängt) mit einem Elektron rekombiniert. Aus dem Verbrauch beider Teilchen wird Energie frei, die in Form zweier 180° zueinander ausgerichteter γ -Quanten freigesetzt wird (Vernichtungsstrahlung). Diese Gammastrahlen werden schließlich von einem ringförmig umschließenden Detektorsystem erfasst, dem sogenannten PET-Scanner. Der Ort ihrer Aussendung, der auf einer geraden Linie zwischen beiden Detektorelementen liegt, kann berechnet werden. Im Unterschied zum Detektorsystem der CT sind die Detektoren des PET-Scanners stationär

angebracht und rotieren nicht. Auch ist die Bildakquisitionszeit mit einigen Minuten deutlich länger im Vergleich zur CT.

Bei der PET/CT sind beide Techniken hintereinander in der Gantry eines gemeinsamen Gerätes verbaut. Es ermöglicht eine sequentielle Akquisition von CT- und PET-Bildern ohne dazwischen den Patienten umlagern zu müssen. So entstehen miteinander verknüpfte (koregistrierte) Bilder, die anatomische und metabolische Bildinformationen in einem gemeinsamen Koordinatensystem abbilden. Aus ihnen lassen sich Fusionsbilder erstellen, die die anatomische Zuordnung der metabolischen PET-Informationen erleichtern (Abb. 3.2). Außerdem dient die Bildinformation aus der CT der Ableitung von Schwächungskarten zur Kompensation der PET-Rohdaten. In diesen liegen neben *echten* Koinzidenzen auch *falsche* Koinzidenzmessungen vor, welche beispielsweise durch zwei zufälligerweise gleichzeitig registrierte γ -Quanten aus unterschiedlichen Vernichtungsreaktionen entstehen können und dann Messfehler verursachen. Aus dieser Schwächungskorrektur (und weiterer Korrekturen wie der Streustrahlen- und der Zerfallskorrektur) resultieren dann exaktere PET-Bilder, was einen weiteren Vorteil einer Koregistrierung beider Modalitäten darstellt.

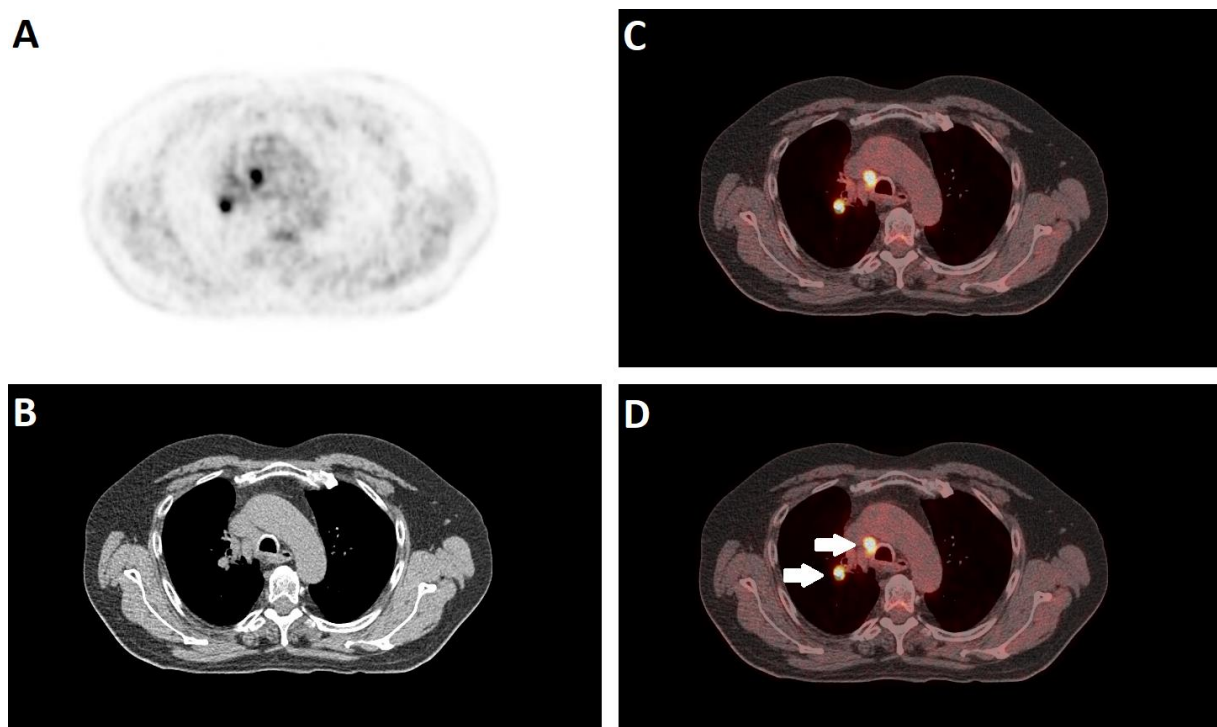


Abbildung 3.2 Beispiel für die Koregistrierung von CT und PET.

Das Bildbeispiel zeigt das schwächungskorrigierte PET (A), die axiale CT-Schicht (B) und das resultierende Fusionsbild (C) eines Patienten mit einem Bronchialkarzinom des rechten Oberlappens (nicht abgebildet). Auf Höhe des Aortenbogens zu sehen sind eine pulmonale Satellitenmetastase und eine mediastinale Lymphknotenmetastase (Pfeilmarkierung, D).

3. Material und Methoden. Die qualitative Auswertung durch die Observer

Die bei der PET/CT verwendeten Isotope werden chemisch an stoffwechselaktive Moleküle gebunden und dem Patienten als „Tracer“ verabreicht. Der am häufigsten verwendete Tracer für die PET Bildgebung ist ¹⁸F-FDG. Dabei handelt es sich um ein radioaktiv markiertes Glukoseanalogon, welches sich im Körper in besonders stoffwechselaktiven Zellen anreichert. Dies betrifft Organe mit physiologischer Weise hohem Energiebedarf und somit hohem Glukosestoffwechsel wie das Herz und das Gehirn oder zum Beispiel Muskulatur bei Anspannung. Die Zellen transportieren die radioaktiv markierte Glukose ins Zellinnere, wo sie nach dem ersten Schritt der Glykolyse akkumuliert, da sie vom nachfolgenden Enzym, aufgrund des Fluor-Isotopes, nicht weiter verstoffwechselt werden kann. In den meisten gesunden Zellen im Körper verhindert die Anwesenheit von Sauerstoff durch den Pasteur-Effekt jedoch die anaerobe Glykolyse, führt die Glucose in Form von Pyruvat dem Citratzyklus zu und senkt somit insgesamt den Glukosebedarf der Zelle (71). Zu Beginn des 20. Jahrhunderts konnte Otto Warburg in seinen Untersuchungen zum Stoffwechsel von Tumoren bereits zeigen, dass Krebszellen diesen Mechanismus unterdrücken und unter hohem Glukosebedarf Energie aus der aeroben Glykolyse gewinnen (72). Aber nicht nur Malignome, sondern auch entzündlich verändertes Gewebe weist einen hohen Glukoseumsatz auf und reichert auf diese Weise ¹⁸F-FDG in den Zellen an (51). Gerade in der Rezidivdiagnostik nach Radiatio stellt diese Tatsache den Auswerter häufig vor Probleme.

Im Unterschied zur CT, die eine rein morphologische Momentaufnahme eines unentwegt stoffwechselaktiven, lebendigen Organismus darstellt, gibt die FDG-PET/CT also Aufschluss über den Glukosestoffwechsel im Körper und erlaubt Rückschlüsse auf physiologische und pathologische Prozesse im Organismus. Zur richtigen Interpretation der mittels FDG-PET/CT akquirierten Bilddaten benötigt es neben Kenntnissen über den physiologischen Glukosemetabolismus im Körper auch die richtige Analysestrategie, um unphysiologische Anreicherungen einer benignen oder malignen Genese zuordnen zu können (34).

3.3. Die qualitative Auswertung durch die Observer

Zur Erleichterung der Durchführbarkeit und zur standardisierten Erfassung der Antworten wurde für qualitative Auswertung durch die vier *Reader* ein einheitlicher Fragebogen mit 2 Seiten pro Bilddatensatz erstellt. Jeder *Reader* bekam somit 228 Seiten zur Auswertung, aufgeteilt auf jeweils zwei Auswertebögen bereitgestellt. Die

Auswertung des ersten Antwortbogens fand im Zeitraum von April bis August 2019 statt, die des zweiten zwischen September 2019 bis März 2020. Die Antwortquote auf die Fragen und die Rücklaufquote der Fragebögen betrug 100%.

3.3.1. Erstellung des Fragebogens

Der Erstellung des Fragebogens ging eine systematische Literaturrecherche auf Pubmed mit den Begriffen „qualitative“, „quantitative“, „criteria“, „recurrence“, „fdg-pet“, „lung cancer“ und „nsccl“ voraus. Die Erstellung des vorläufigen Fragebogens erfolgte in enger Zusammenarbeit mit einem der PET/CT-Gutachter aus dem Expertengremium des Harmonisierungsprozesses (66). Nach Erstellung des Fragebogens erfolgte ein Testdurchgang an 25 Patientendatensätzen durch einen der *Observer* zur Praxistauglichkeit des Fragebogens und zur Evaluation der benötigten zeitlichen Dauer der Auswertungen. Nach dem Testdurchgang erfolgte eine erneute Überarbeitung des Fragebogens zur Optimierung der zeitlichen Durchführbarkeit. Die beiden Seiten des Fragebogens sind in der Abbildung 3.2 und der Abbildung 3.3 dargestellt. Sie enthalten zwei in binäre Antwortmöglichkeiten („ja/nein“, „trifft zu/trifft nicht zu“, „sicher/unsicher“) gegliederte Abschnitte zur Bewertung der Lunge und des Mediastinums, sowie ein Feld mit der Frage nach Fernmetastasen und ein Feld zur Freitextergänzung. Dabei orientieren sich die abgefragten Kriterien in ihrer Interpretation an den durch Hicks *et al.* definierten subjektiven Auswertekriterien (42). Um die zeitlich aufwändige, visuelle Auswertung anhand des Fragebogens zu erleichtern, wurde die intrapulmonale Metastase zunächst bei der Bewertung der Lunge mit abgefragt und fiel in der Bewertung durch die Observer auch in die Kategorie des pulmonalen Rezidiv. Erst später, im Vergleich mit dem Goldstandard zum tatsächlichen Outcome der Patienten wurde dann lediglich zwischen lokoregionärem Rezidiv (pulmonal und/oder mediastinal) und Metastase (also *distant recurrence*) in der Observer-Auswertung unterschieden. Hierzu wurden alle als zutreffend markierten Wertungen für eine intrapulmonale Metastase und eine extrathorakale Metastase als *distant recurrence* zusammengefasst. Pulmonale Rezidive die insgesamt nur aufgrund einer intrapulmonalen Metastase zustande kamen, flossen nicht mehr als solche in die Bewertung eines Lokalrezidivs mit ein. Die übrigen pulmonalen Rezidive und die mediastinalen Rezidive wurden als lokoregionäres Rezidiv zusammengefasst.

3. Material und Methoden. Die qualitative Auswertung durch die Observer

Seite 1 – Ohne Kenntnisse über Baseline PET-Scan und Radiotherapie

PET/CT – ID: PAT

PET/CT – Datum:

Bewertung der **Lunge**

	<u>trifft zu</u>	<u>trifft nicht zu</u>	<u>Evaluierung sicher</u>	<u>Evaluierung unsicher</u>
Im Vergleich zur Leber				
Fokale FDG-Akkumulation				
- in LK 10-11R				
- in LK 10-11L				
- im zentralen Lungenparenchym				
- im peripheren Lungenparenchym				
Fibrotisches Areal mit FDG-Uptake ... malignomverdächtig?				
Intrapulmonale Metastase?				

Insgesamt V.a. pulmonales Rezidiv? Ja Nein sicher unsicher

Bewertung der **Lymphknotenstationen** im Mediastinum

LK-Station	<u>Positiv</u>	<u>Negativ</u>	<u>Sicher</u>	<u>Unsicher</u>
1-2R				
1-2L				
3A				
3P				
4R				
4L				
5				
6				
7				
8				

(nicht bewertete Stationen werden als ‚sicher negativ‘ gewertet)

V.a. mediastinales Rezidiv? Ja Nein sicher unsicher

extrathorakale Metastasen? Ja → Wo?:

Sonstige Anmerkungen:

Abbildung 3.3 Fragebogen zur reproduzierbaren qualitativen Auswertung, Seite 1.

3. Material und Methoden. Die qualitative Auswertung durch die Observer

Seite 2 – Kenntnisse über Baseline PET-Scan und Radiotherapie

PET/CT – ID: PAT

Baseline-Scan: BAS

Bewertung der Lunge

	<u>trifft zu</u>	<u>trifft nicht zu</u>	<u>Evaluierung sicher</u>	<u>Evaluierung unsicher</u>
<i>(Im Vergleich zur Leber)</i>				
Fokale FDG-Akkumulation				
- in LK 10-11R	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
- in LK 10-11L	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
- im zentralen Lungenparenchym	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
- im peripheren Lungenparenchym	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Fibrotisches Areal mit FDG-Uptake	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... malignomverdächtig?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Intrapulmonale Metastase?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Insgesamt V.a. pulmonales Rezidiv?	<input type="checkbox"/> Ja	<input type="checkbox"/> Nein	<input type="checkbox"/> sicher	<input type="checkbox"/> unsicher
➔ Vergleich des SUV _{max} für Beurteilung	<input type="checkbox"/> hilfreich	<input type="checkbox"/> nicht hilfreich		

Bewertung der Lymphknotenstationen im Mediastinum

LK-Station	bestrahlt?	LK-Station	<u>Positiv</u>	<u>Negativ</u>	<u>Sicher</u>	<u>Unsicher</u>	
1-2R	<input type="checkbox"/>	1-2R	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
1-2L	<input type="checkbox"/>	1-2L	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
3A	<input type="checkbox"/>	3A	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
3P	<input type="checkbox"/>	3P	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
4R	<input type="checkbox"/>	4R	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
4L	<input type="checkbox"/>	4L	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
5	<input type="checkbox"/>	5	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
6	<input type="checkbox"/>	6	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
7	<input type="checkbox"/>	7	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
8	<input type="checkbox"/>	8	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
10-11R	<input type="checkbox"/>	(nicht bewertete Stationen werden als ‚sicher negativ‘ gewertet)				<input type="checkbox"/>	<input type="checkbox"/>
10-11L	<input type="checkbox"/>						

V.a. mediastinales Rezidiv?	<input type="checkbox"/> Ja	<input type="checkbox"/> Nein	<input type="checkbox"/> sicher	<input type="checkbox"/> unsicher
-----------------------------	-----------------------------	-------------------------------	---------------------------------	-----------------------------------

Anmerkungen:

Abbildung 3.4 Fragebogen zur reproduzierbaren qualitativen Auswertung, Seite 2.

3.3.2. Rekrutierung der *Observer*

Die Rekrutierung der Auswerter für diese Studie erfolgte bei zwei der vier *Readern* aus dem Expertengremium der PET-Plan Studie. Beide *Reader* sind Fachärzte für Nuklearmedizin mit langjähriger Berufserfahrung. Die zwei weiteren *Reader* (ebenfalls Nuklearmediziner) erhielten ein mündliches Teaching und eine schriftliche Zusammenfassung über die wesentlichen Schritte des Harmonisierungsprozesses zur Auswertung vor Beginn. Außerdem zur Verfügung gestellt wurden allen *Readern* die Veröffentlichung der Panel-Harmonisierung, der Atlas zur korrekten Zuordnung der Lymphknotenstationen im Mediastinum und die Veröffentlichung zur Befallswahrscheinlichkeit der mediastinalen Lymphknotenstationen in Abhängigkeit zur Lokalisation des Primärtumors in der Lunge (26, 66, 67).

3.3.3. Auswertung durch die *Observer*

Alle vier *Reader* erhielten zu Beginn des entsprechenden Auswertzeitraums einen handschriftlich auszufüllenden Auswertbogen und die zugehörigen PET/CT-Bilddatensätze. Die anonymisierten Bilddatensätze, im gängigen DICOM-Format, konnten in das eigene System eingelesen, archiviert und betrachtet werden. Das Setting wurde bewusst so gewählt, dass die *Reader* die Auswertungen in ihrem gewohnten Arbeitsumfeld und Software-Umgebung mittels eigenem PACS-Viewer vornehmen konnten. Bei der Auswertung wurden die *Observer* dazu angehalten, in mehreren Schritten vorzugehen:

In einem ersten Schritt wurde zunächst nur das Follow-Up-PET/CT bei Rezidivverdacht betrachtet und dieses anhand der ersten Seite des Fragebogens bewertet. Dabei waren die entsprechenden binären Antwortmöglichkeiten der geschlossen gestellten Fragen anzukreuzen, die nach Meinung des Auswerter zutrafen. Die Bewertung der Lunge sieht dabei eine Einschätzung bezüglich der Lage einer möglichen fokalen FDG-Akkumulation im Lungenparenchym, die Zuordnung der Dignität eines möglicherweise postradiogenen, fibrotischen Areals und eine Einschätzung hinsichtlich einer möglicherweise neu aufgetretenen Läsion vor. Anschließend sollte der *Reader* beurteilen, ob es sich in Betrachtung der Gesamtkonstellation um ein pulmonales Rezidiv handele oder nicht. Bei der Auswertung des Mediastinums wurde neben der Zuordnung eines FDG-aviden Herdbefundes zum entsprechenden Lymphknotenlevel ebenfalls ein Urteil auf die Frage eines möglichen Rezidiv verlangt. Zusätzlich wurde bei allen Bewertungen die

3. Material und Methoden. Die qualitative Auswertung durch die Observer

subjektive Sicherheit der abgegebenen Einschätzung („sicher/nicht sicher“) und die Frage nach einer Fernmetastase („ja/nein“) erfragt.

Im zweiten Schritt erhielten die *Reader*, nach Umblättern der Seite im Auswertebogen, Kenntnis über die Bestrahlungsplanung des Patienten und erhielten das Baseline-PET/CT zur vergleichenden Betrachtung mit dem unmittelbar zuvor bewerteten Follow-Up-PET/CT.

In einem dritten Schritt war der *Reader* dann mit den neu gewonnenen Erkenntnissen über die Ausgangssituation und das Bestrahlungsfeld dazu angehalten die entsprechende zweite Seite des Fragebogens auszufüllen. Hier erfolgte eine ähnlich strukturierte, erneute Bewertung der zuvor abgefragten *Items* in Lunge und Mediastinum. Zusätzlich wurde wieder nach der subjektiven Bewertungssicherheit der Kriterien gefragt. Außerdem fand sich auf der zweiten Seite auch ein Feld, bei dem die *Reader* gefragt wurden, ob sie nach Vorliegen des Baseline-Scans den SUVmax-Wert beider Bilddatensätze verglichen und diesen als hilfreich zur Urteilsbildung empfunden hatten.

Während des Auswertevorgangs galt, dass nach Weiterblättern auf die nächste Seite innerhalb des Auswertebogens nicht mehr auf die vorigen Seiten zurückgeblättert werden durfte. Auch Änderungen auf der ersten Seite durften nach Kenntnis über die Bestrahlungsplanung nicht mehr durchgeführt werden. Bei der Betrachtung der PET/CT-Datensätze war es den Auswertern jedoch erlaubt, die „Maximum Intensity Projektion“ der PET-Komponente zu verwenden, sowie alle Schichten der axialen Ebenen auch separat als CT- oder PET-Bilder zu betrachten. Dabei war ebenso die Verwendung von Tools wie Zoom und Größenmessung, sowie eine Änderung der Kontrasteinstellungen gestattet.

Anschließend an den jeweiligen Untersuchungszeitraum erfolgte die elektronische Erfassung und tabellarische Darstellung der binär kodierten Fragen im Excel- und im SPSS-Datenformat, wobei „0“ meist für eine negative Antwort und „1“ für eine positive Antwort verwendet wurde.

3.4. Die quantitative Auswertung

3.4.1. Die Auswertung nach den PERCIST- und PREDIST-Kriterien

Die Durchführung der Auswertung nach PERCIST erfolgte mithilfe der kommerziellen Software Hybrid3D Tumor Finder. Ihr liegt ein semiautomatischer, Schwellenwert-basierter Algorithmus zugrunde, der speziell für diese Anwendung von der Firma HERMES angeboten wird (Hybrid3D Tumor Finder Version 2.2, HERMES MEDICAL SOLUTIONS, Stockholm, Schweden). In einem ersten Schritt wurden zunächst die PET/CT-Bilddaten zum Zeitpunkt der Bestrahlungsplanung und die zugehörigen Follow-Up Untersuchungen archiviert. Damit eine Läsion im Baseline-PET als messbar gilt, fordern die PERCIST-Kriterien die Definition eines sogenannten Schwellenwertes, der die obere Grenze der physiologischen FDG-Belegung im Organismus definiert (8). Hierzu wurde ein VOI mit einem Durchmesser von 3 cm innerhalb des rechten Leberlappens und unter Vermeidung von großen Gefäßen oder fokalen Parenchymveränderungen definiert (35). Aus dem gemessenen Mittelwert des SUL in der Leber wurde anhand der Formel $[1,5 \times \text{SUL}_{\text{mean}} + 2 \times \text{SD}]$ der Schwellenwert definiert. Unter Ausschluss aller physiologischen FDG-Anreicherungen (Gehirn, Nieren, Harnblase, lymphatisches Gewebe der Schleimhäute, Muskulatur, ...) wurden dann automatisch alle Areale mit FDG-Akkumulation oberhalb des Schwellenwerts definiert. Gleichzeitig wurde innerhalb der definierten Volumina nach einem etwa 1 ml messenden Volumen gesucht, welches den höchsten SUL_{peak} -Wert aufweist. Dieser SUL_{peak} dient, nach den PERCIST-Kriterien, dann als Ausgangswert für alle späteren Vergleiche mit nachfolgenden Untersuchungen. Aus der entsprechenden Follow-Up Untersuchung wurde dann ebenfalls der höchste SUL_{peak} -Wert bestimmt. Dieser muss nicht zwangsläufig aus der gleichen Läsion, wie in der Untersuchung zuvor stammen; eine Schwellenwertdefinition anhand der Leber ist für seine Bestimmung (nach PERCIST) ebenfalls nicht notwendig. Die prozentuale Veränderung zwischen beiden wurde anhand der Formel $[100 \times (\text{Follow-Up-Wert} - \text{Baseline-Wert}) / \text{Baseline-Wert}]$ berechnet und anschließend der entsprechenden der vier möglichen Kategorien zugeordnet. Dabei lag eine komplette metabolische Remission (CMR) vor, wenn kein vitales Tumorgewebe oberhalb des FDG-Niveaus der Leber mehr messbar war. Bei mehr als 30% prozentualen und mehr als 0,8 Einheiten absolutem Abfall des SUL_{peak} handelte es sich um eine partielle metabolische Remission (PMR). Betrug die Abnahme des SUL_{peak} im Follow-Up weniger als 30% oder 0,8 Einheiten und gab es gleichzeitig auch keine Zunahme um mehr als 30% oder mehr als 0,8 Einheiten, so lag eine stabile metabolische Erkrankung vor (SMD). Eine Zunahme des SUL_{peak} von

mehr als 30% oder 0,8 Einheiten gegenüber der Voruntersuchung oder eine neu aufgetretene Läsion kennzeichneten einen metabolischen Progress (PMD). Neben diesen Kriterien für die Target-Läsion mit der höchsten Intensität sind noch weitere Kriterien für die Non-Target-Läsionen definiert, die in der Auswertung ebenfalls Berücksichtigung fanden (35).

In Abweichung zu den PERCIST-Kriterien erfolgte die Bestimmung des SUL_{peak} -Wertes im Follow-Up bei den PREDIST-Kriterien ebenfalls anhand einer Schwellenwertdefinition (43, 44). Zur Berechnung des Schwellenwertes wurde der SUL_{mean} -Wert des Leberstoffwechsels im Follow Up wie in der Baseline-Untersuchung bestimmt und der Schwellenwert wieder anhand der Formel $[1,5 \times SUL_{mean} + 2 \times SD]$ bestimmt. Um als Rezidiv zu gelten, musste der SUL_{peak} im Follow Up oberhalb dieser individuellen Schwelle liegen. Fand sich in der Untersuchung kein SUL_{peak} oberhalb des Schwellenwertes, wurde die Untersuchung als CMR gewertet.

3.4.2. Die Berechnung von MTV und TLG

In der gängigen Literatur herrscht für die Berechnung des metabolischen Tumervolumens (MTV) eine hohe Heterogenität an Definitionen für die korrekte Schwelle vor. Gängige Methoden zur Festlegung einer Schwelle stellen beispielsweise die Bestimmung des SUL-Mittelwerts eines 3 cm messenden VOI innerhalb des rechten Leberlappens. Der Wert zur Schwellenbestimmung liegt dann über der Summe aus SUL-Mittelwert und doppelter Standardabweichung ($SUL_{mean} + 2 \text{ SDs}$). Eine weitere Möglichkeit zur Festlegung eines Schwellenwertes stellt die Bestimmung des SUL_{max} und das Festlegen auf einen bestimmten Bruchteil hiervon dar, wie zum Beispiel 50% des SUL_{max} -Wertes beim sogenannten TLG_{50} (30). Aufgrund der hohen Heterogenität in der Literatur und aus Gründen der besseren Vergleichbarkeit mit den anderen erhobenen Parametern fiel die Entscheidung zur Definition eines Schwellenwertes auf dieselbe wie bei der PERCIST-Auswertung. Hierfür wurde ein VOI mit einem Durchmesser von 3 cm im rechten Leberlappen platziert und die Schwelle als $[1,5 \times SUL_{mean} + 2 \times SD]$ definiert. Die Berechnung des entsprechenden metabolischen Tumervolumens (MTV) in der PET/CT zur Bestrahlungsplanung erfolgte dann anhand einer semiautomatischen Schwellenwertkonturierung mittels HERMES (Hybrid3D Tumor Finder Version 2.2, HERMES MEDICAL SOLUTIONS, Stockholm, Schweden).

Im Falle einer singulären Läsion berechnet sich die „Total Lesion Glycolysis“ (TLG) aus dem MTV multipliziert mit dem Mittelwert des SUL. Soll ein Gesamtwert für die TLG mehrerer Tumorkläsionen ermittelt werden, so multipliziert man jeweils die einzelnen Volumina mit dem jeweiligen SUL-Mittelwert und addiert diese Werte im Anschluss.

3.5. Die Festlegung des Goldstandards

In der PET-Plan Studie stellte die Zeit bis zum lokoregionären Progress den primären Endpunkt dar. Overall Survival (OS) und progressionsfreie Überlebenszeit bildeten die sekundären Endpunkte. Der Goldstandard der Studie ist definiert als Diagnose einer lokalen Progression zu jedem Beobachtungszeitpunkt zudem zunächst eine CT- und nachfolgend eine PET/CT-Bildgebung entstand und das Rezidiv bestätigte. Neben dem primären Endpunkt definierten selbstverständlich auch neu diagnostizierte Fernmetastasen und neu aufgetretene thorakale Metastasen das Vorliegen eines Rezidiv zum Beobachtungszeitpunkt. In die Beurteilung gingen neben dem aufwändig standardisierten Reading der PET/CT-Bildgebung auch alle Informationen zum klinischen Verlauf der Patienten, nachfolgende Bildgebung, sowie histologische Sicherungen mit ein. Laut elektronischer Studiendokumentation (eCRF) lag bei Abschluss der Studie bei insgesamt 35 der 82 Patienten mit Rezidivverdacht ein lokoregionäres Rezidiv vor. Bei 42 der 82 Patienten wurde einer der sekundären Endpunkte (Tod, Metastasierung) erreicht. In der Dokumentation des Goldstandards wurde ein Rezidiv also entweder als lokal (primärer Endpunkt) oder als Metastase (sekundärer Endpunkt) dokumentiert. Neu aufgetretene, kontralaterale pulmonale Metastasen fielen dabei in die Kategorie Metastasen.

Im Gegensatz zur Dokumentation des Goldstandards wurden neu aufgetretene, pulmonale Metastasen im Fragebogen zur qualitativen Auswertung durch die *Observer* zunächst einem pulmonalen Rezidiv zugeordnet. Da für die Zuordnung „Metastase oder Lokalrezidiv“ die Kenntnis des Ausgangsbefundes zwingend erforderlich ist, wurde dies in der Abfrage der Observerauswertung zunächst als eigener Punkt der Lungenbeurteilung abgefragt. Zum späteren Vergleich mit dem Goldstandard erfolgte die Unterteilung der zu bewertenden Kategorien im Auswertebogen nicht mehr in pulmonal, mediastinal und extrathorakal, sondern in lokoregionäres (pulmonales oder mediastinales) Rezidiv und Metastase (pulmonal und extrathorakal), entsprechend der gängigen Unterteilung in der Literatur (siehe Kapitel 3.3.1).

3.5.1. Qualitätssicherung

Die interne Qualitätssicherung dieser Arbeit ist eng angelehnt an die Qualitätssicherung innerhalb der PET-Plan Studie und beruht, in den wesentlichen Punkten, auch bei dieser Arbeit auf den hierzu veröffentlichten Kriterien (66). Unter anderem beruhen die Outcome Daten der Patienten auf demselben diagnostischen Goldstandard zur Definition des lokalen Progress bzw. Rezidivs. Auch sind weite Teile der Interpretation und des Readings der PET/CT-Bilddaten dieselben und profitieren somit unmittelbar vom Panel-Harmonisierungsprozess. Lediglich die zu bewertenden Kriterien weichen von denen der PET-Plan Studie ab. Gleich bleibend ist hingegen die anatomische Zuordnung pulmonaler Läsionen, sowie die Einordnung in die mediastinalen Lymphknotenlevel und die zuvor bestimmte Sensitivität zur Beurteilung einzelner, nicht eindeutiger Läsionen, wie zum Beispiel eine niedrige *Reader-Sensitivität* bei einer schwachen FDG-Akkumulation im kontralateralen Hilus bei niedrigem prädiktivem Wert (26).

3.6. Die statistische Auswertung

3.6.1. Grundlagen der Interobservervariabilität und Cohens Kappa

Die Bestimmung der Interobservervariabilität (IOV) dient als Maß für die Übereinstimmung zwischen Bewertern. Die Bedeutung der IOV liegt darin zu quantifizieren, wie unabhängig, objektiv und zuverlässig die Bewertung eines Objektes durch den Bewertenden ist. Sie erlaubt es Rückschlüsse zu ziehen, inwiefern eine Abhängigkeit zwischen dem Messverfahren und der durchführenden Person besteht. Das Präfix „*Inter-*“ steht dabei für den Vergleich des Urteils von zwei verschiedenen *Observern*. Da die Übereinstimmungswerte der IOV streng genommen nicht die Abweichung, sondern die Übereinstimmung der Bewerter messen, werden sie in dieser Arbeit deshalb als Interobserver-Agreement bezeichnet und mit IOA abgekürzt.

Die Berechnungen der Interobservervariabilität beruht auf den folgenden mathematischen Überlegungen:

Dieselben Items sollen anhand kategorialer Merkmale („ja“ oder „nein“) von verschiedenen Betrachtern beurteilt und die Übereinstimmung der Ergebnisse ermittelt werden. In einem exemplarischen Szenario sind hierfür die Urteile zweier Ärzte (die *Observer*) für jeweils zehn Patientenfälle (Beurteilungsobjekte) in einer Vier-Felder-

3. Material und Methoden. Die statistische Auswertung

Tafel aufgetragen (Abb. 3.4). Beide Ärzte beurteilen, ob der jeweilige Patient krank ist entweder mit „ja“ für krank oder mit „nein“ für „nicht krank“.

		<u>Arzt 2</u>	
		Ja	Nein
<u>Arzt 1</u>	Ja	6	0
	Nein	3	1

Abbildung 3.5 Vier-Felder-Tafel der Urteile beider Ärzte.

In diesem Beispiel stimmen beide Ärzte in der Bewertung von 6 Patienten mit der Antwort „ja“ bei 1 der 10 Fälle mit der Antwort „nein“ überein. In 3 der 10 Fälle bewertete der erste Arzt drei der Patienten mit „nein“ (als nicht krank), während der zweite Arzt diese mit „ja“ (als krank) bewertete. Die globale Übereinstimmung (7 von insgesamt 10 Fälle) wird als p_0 bezogen auf die Gesamtheit der zu beurteilenden Objekte angegeben, wobei 0 keiner (null-prozentige) und 1 einer kompletten (hundertprozentige) Übereinstimmung entspricht. In diesem Beispiel liegt die Übereinstimmung bei $p_0 = 0,7$. Das Interobserver-Agreement läge somit also bei einer 70%-igen Übereinstimmung.

In diesem Beispiel ist der Sachverhalt relativ offensichtlich und einfach zu berechnen. In komplizierteren Fällen, in denen zwei *Observer* hinsichtlich der Übereinstimmung in mehreren Beurteilungsobjekten und -kategorien verglichen werden sollen, dient folgende mathematische Formel der Berechnung:

$$p_0 = \frac{\sum_{i=1}^z h_i}{N}$$

p_0 = berechneter Übereinstimmungswert

z = Beurteilungskategorie

i = Beurteilungsobjekt

N = Anzahl der Beurteilungsobjekte

h = Beurteilungshäufigkeit

h_i = Randhäufigkeit (entspricht der Summer einer Spalte der Vier-Felder-Tafel)

3. Material und Methoden. Die statistische Auswertung

Allerdings wird in dieser Formel das Szenario, dass die *Observer* mit einer gewissen Wahrscheinlichkeit auch zufällig zur Übereinstimmung kommen können nicht berücksichtigt. Hierfür gibt es eine weitere Formel, in der die zu erwartende Übereinstimmung unter Berücksichtigung einer zufälligen Übereinstimmung p_e angegeben wird:

$$p_e = \frac{1}{N^2} \sum_{i=1}^z h_i \cdot h_i$$

Für das oben genannte Beispiel der Bewertung durch die zwei Ärzte ergibt sich hieraus der Wert $p_e = 0,58$, also eine zufällige Übereinstimmung in 5,8 der 10 Fälle.

Mittels der Berechnung von Cohens Kappa (κ) ergibt sich nun die Möglichkeit unter Einbeziehung beider Größen p_0 und p_e einen Wert für die Übereinstimmungsgüte zu bekommen, nachdem die zunächst grobe Übereinstimmung durch die zufällig zustande gekommene Übereinstimmung korrigiert wurde (73):

$$\kappa = \frac{p_0 - p_e}{1 - p_e}$$

Für das Beispiel lässt sich hiermit ein κ -Wert von $\kappa = 0,29$ errechnen.

Während ein κ -Wert von 0 dafür steht, dass eine Übereinstimmung rein zufällig zustande kommt, weisen Werte zwischen 0 und 1 auf eine objektivierbare Übereinstimmung zwischen den Urteilen der beiden Bewerter hin. Eine genauere Einteilung der κ -Werte und damit eine Einschätzung der Güte der Urteile gelingt mithilfe der Interpretation nach Landis und Koch (74):

Dabei gilt $\kappa < 0$	schlechte Übereinstimmung („poor“)
$\kappa = 0,00 - 0,20$	leichte Übereinstimmung („slight“)
$\kappa = 0,21 - 0,40$	ausreichende Übereinstimmung („fair“)
$\kappa = 0,41 - 0,60$	mittlere Übereinstimmung („moderate“)
$\kappa = 0,61 - 0,80$	erhebliche Übereinstimmung („substantial“)
$\kappa = 0,81 - 1,00$	(fast) perfekte Übereinstimmung („(almost) perfect“)

Die Berechnung der Übereinstimmung zweier *Observer* mittels Cohens Kappa-Wert bietet also den entscheidenden Vorteil, dass sie neben der reinen Betrachtung der miteinander übereinstimmenden Werte auch mit einbezieht, dass Übereinstimmungen zufällig zustande kommen können. Das oben genannte Beispiel der zwei Ärzte zeigt, dass die rohe Übereinstimmung von 70% die wahre Übereinstimmung deutlich überschätzt hätte, die bei lediglich um die 30% lag. Eine weitere Möglichkeit eine solche wahre Übereinstimmung zweier *Observer* zu berechnen bietet Scotts Pi-Wert. Er unterscheidet sich nur unwesentlich in der Berechnung des p_e -Wertes von der Formel zu Cohens Kappa-Wert (75). Beide Formeln besitzen jedoch einen entscheidenden Nachteil: Sie bieten lediglich die Möglichkeit endlich viele kategoriale Urteile von genau zwei *Observern* einzuschätzen. Möchte man hingegen die Übereinstimmung von mehr als nur 2 *Observern* berechnen, so benötigt man hierfür die Kappa-Wertberechnung nach Fleiss, welche auch in dieser Arbeit für den Vergleich zwischen den 4 *Observern* verwendet wurde (76). War dies nicht möglich, so wurde der Mittelwert aus den Werten nach Cohens Kappa bestimmt.

Für die qualitativen Auswertungen der Fragebögen durch die *Observer* entsprachen die Beurteilungsobjekte den aufgeführten Items auf den Fragebögen. Die Beurteilungskategorien bildeten entsprechend binäre kategoriale Dimensionen aus, das heißt sie konnten nur zwei mögliche Werte annehmen (beispielsweise „zutreffend“ oder „nicht zutreffend“, „sicher“ oder „unsicher“). Die Beurteilungshäufigkeit entsprach bei der Auswertung entweder der Anzahl der Patienten ($n = 84$) oder der Gesamtzahl der verfügbaren PET/CT-Untersuchungen ($n = 114$).

3.6.2. Odds Ratio und logistische Regression

Die Odds Ratio ist ein Zusammenhangsmaß, welches angibt, um wieviel sich die Eintrittswahrscheinlichkeit eines Ereignisses erhöht, wenn ein bestimmter Risikofaktor als Bedingung vorliegt. Wertet man zum Beispiel die Merkmale „gesund“ oder „erkrankt“ als Ereignis und die Merkmale „exponiert“ oder „nicht exponiert“ als Risikofaktor, so ergibt sich für die Berechnung der Odds Ratio folgende Formel:

$$\text{Odds Ratio} = \frac{\text{Anzahl exponierter Erkrankter} \times \text{Anzahl nicht exponierter Gesunder}}{\text{Anzahl exponierter Gesunder} \times \text{Anzahl nicht exponierter Erkrankter}}$$

Ergibt das berechnete Ergebnis einen Zahlenwert größer 1, so erhöht sich das Risiko des Eintretens des Ereignisses bei Vorliegen des Risikofaktoren. Liegt der Zahlenwert zwischen 0 und 1 so verringert sich das Risiko bei Vorliegen des Faktors. Bei einem Zahlenwert von genau 1 ändert das Vorliegen des Risikofaktors die Eintrittswahrscheinlichkeit des Ereignisses nicht; das Risiko ist also genau so hoch, wie ohne das Vorliegen des Faktors. Will man diesen Umstand für mehrere, voneinander unabhängige Risikofaktoren gleichzeitig untersuchen, so benötigt man das Modell einer logistischen Regressionsanalyse. Sie dient der Berechnung der Odds für die unabhängigen Variablen unter gleichzeitiger Berücksichtigung aller Faktoren.

Bei der Berechnung der Odds Ratio in dieser Studie galt die Wertung bezüglich eines Rezidiv als eintretendes Ereignis und die bewerteten Items als unabhängige Risikofaktoren für das Eintreten des Ereignisses. Die Berechnung der Odds diente der Einschätzung eines Zusammenhangs vor der weiteren Berechnung des logistischen Regressionsmodells, in dem alle zugehörigen Items gleichzeitig entweder einzeln oder gruppiert einbezogen wurden.

3.6.3. Ereigniszeitanalysen

Cox-Regressionsanalyse

Im Gegensatz zur logistischen Regressionsanalyse, die den Einfluss unabhängiger Variablen auf eine kategoriale, abhängige Variable schätzt, benötigt man für die Schätzung des Einflusses unabhängiger Variablen auf die Dauer bis zum Eintreten eines bestimmten Ereignisses eine Methode der Ereigniszeitanalyse. Um die Hazardraten für die quantitativen Parameter der PET/CT zu berechnen wurden Cox-Regressionsanalysen durchgeführt. Als eintretendes Ereignis galt der dokumentierte Zeitpunkt des tatsächlichen Gesamtrezidivs.

Kaplan-Meier-Schätzer

Die Darstellung der Daten in einer Kaplan-Meier-Kurve dient der Abschätzung der Überlebensfunktion im Rahmen einer rechts-zensierten Ereigniszeitanalyse. Dabei wird für eine Grundgesamtheit an Probanden in einem Graphen erfasst, zu welchem Zeitpunkt ein bestimmtes Ereignis trat oder nicht trat. Die daraus resultierende Kurve fällt nach rechts ab und besitzt gegen null gehende Intervalllängen. Werden zwei Bedingungen für das Eintreten eines Ereignisses miteinander verglichen, lässt sich mithilfe eines Log-Rank-Tests prüfen, ob diese einen signifikant unterschiedlichen Einfluss auf das Überleben der Probanden besitzen. In dieser Arbeit dient dieses

Verfahren der Testung des Einfluss der PET-Parameter zum Zeitpunkt der Bestrahlungsplanung auf das Eintreten eines Rezidiv.

3.6.4. Sensitivität und Spezifität

Sensitivität und Spezifität spiegeln zwei weitere Gütekriterien zu Einschätzung eines Testverfahrens wider. Während die Sensitivität angibt wie gut ein Test einen tatsächlich Erkrankten auch als krank erkennt, bezeichnet die Spezifität eines Tests wie häufig ein tatsächlich Gesunder auch als gesund erkannt wird. So sollte ein gutes Screening-Verfahren beispielsweise aus zwei Testungen bestehen, bei denen der erste Test eine hohe Sensitivität aufweist um einen Erkrankten sicher zu erkennen. Der zweite Test sollte dann aus der vorselektierten Gruppe diejenigen zuverlässig erkennen die gesund sind.

$$\text{Sensitivität} = \frac{\text{Anzahl richtig Positiver}}{\text{Anzahl richtig Positiver} + \text{Anzahl falsch Negativer}}$$

$$\text{Spezifität} = \frac{\text{Anzahl richtig Negativer}}{\text{Anzahl richtig Negativer} + \text{Anzahl falsch Positiver}}$$

Abbildung 3.6 Sensitivität und Spezifität.

Beide Maßzahlen nehmen Werte zwischen 0 und 1 an und werden für gewöhnlich in Prozent angegeben. Für die Testung der Güte der qualitativen und quantitativen Urteile wurde bei der Berechnung der Goldstandard zugrunde gelegt.

3.6.5. Weitere Berechnungen

Die demographischen Patientendaten, sowie die Messdaten zum Baseline-Zeitpunkt wurden als deskriptive Statistik tabellarisch zusammengefasst. Die Berechnungen der Interobservervariabilitäten und der Odds, sowie das Erstellen der logistischen Regressionsanalysen und der Ereigniszeitanalyse mittels Cox Regressionsmodell erfolgten mithilfe von SPSS. Die Testung zur Mittelwertabweichung des IOA erfolgte als gepaarter *t*-Test. Die Berechnung von Sensitivität und Spezifität erfolgte durch Excel. Bei allen Ergebnissen wurde ein *p*-Wert < 0,05 als statistisch signifikant erachtet. Die *p*-Wert Testung erfolgte, wenn die Richtung des Effektes klar war, nur einseitig. Der Konsensus der qualitativen *Observer*-Meinung wurde aus der mehrheitlich als richtig bewerteten Aussagen gebildet. Wurde ein Urteil zu gleichen

Teilen als „zutreffend“ beziehungsweise „nicht zutreffend“ markiert, so wurde im Konsensus durch einen fünften *Observer* ermittelt welches Urteil als zutreffend gilt.

3.6.6. Datenverwaltung und Literaturrecherche

Die Ausgabe der Auswertebögen erfolgte gedruckt auf Papier, die Auswertung erfolgte an den jeweiligen PACS-Viewer der Institute. Für die Auswertung nach PERCIST-Kriterien wurde die Software HERMES Hybrid 3D Viewer (HERMES MEDICAL SOLUTIONS) genutzt. Die Erfassung der Ergebnisse der Auswertebögen und der PERCIST-Auswertung erfolgte elektronisch mithilfe von Microsoft Excel für Windows (Microsoft Office 365 Enterprise) und SPSS (Version 23, IBM), ebenso die Berechnungen und die grafische Darstellung. Zur Literaturrecherche wurde das Portal Pubmed (<https://www.ncbi.nlm.nih.gov>) genutzt; für die Verwaltung der verwendeten Literatur wurde Endnote X9 (Thomson Reuters) verwendet. Die Verschriftlichung der Ergebnisse erfolgte mit dem Programm Microsoft Word für Windows (Microsoft Office 365 Enterprise).

4. Ergebnisse

4.1. Charakterisierung des Patientenkollektivs

Die Charakterisierung des Patientenkollektivs ergab die in der Tabelle 4.1 dargestellten Ergebnisse. Zur Übersicht und besseren Vergleichbarkeit mit den Ergebnissen der PET-Plan Studie, erfolgt die Darstellung einmal als Gesamtkollektiv und einmal in Unterteilung in den jeweiligen konventionellen und experimentellen Therapiearm. Aus dem hier vorliegenden Gesamtkollektiv von 82 Patienten entstammen 42 Patienten dem Arm A mit der konventionellen Bestrahlungsplanung und 35 Patienten dem Arm B mit der rein PET-basierten Bestrahlungsplanung. 5 Patienten aus dem Gesamtkollektiv sind dem Beobachtungsarm C zugeordnet und werden hier nicht separat aufgelistet.

Patientencharakteristika	Gesamt n= 82	Arm A konventionelle Bestrahlungsplanung n=42	Arm B PET-basierte Bestrahlungsplanung n=35
Alter zu Therapiebeginn Median (IQR), [Jahre]	64,8 (57,2 - 71,6)	65,0 (57,1 - 72,6)	65,4 (59,8 - 71,4)
Geschlecht, n (%)			
Männlich	62 (76)	30 (70)	27 (75)
Weiblich	20 (24)	12 (30)	8 (25)
Fettbereinigte Körpermasse (LBW) zu Therapiebeginn Mittelwert (SD), [kg]	56 (8,4)	55 (8,4)	57 (8,6)
ECOG, n (%)			
0	41 (50)	21 (50)	14 (40)
1	33 (40)	19 (45)	20 (57)
2	0	0	0
keine Angabe	8 (10)	2 (5)	1 (3)
UICC-Stadium (V7) zu Therapiebeginn, n (%)			
II A	2 (2)	1	1
II B	2 (2)	2	0
III A	31 (38)	16	15
III B	39 (48)	21	18
keine Angabe	8 (10)	2	1
Histologische Differenzierung zu Therapiebeginn, n (%)			
Plattenepithelkarzinom	38 (46)	19	19
Adenokarzinom	27 (33)	16	11
Großzelliges Karzinom	0	0	0
Andere	7 (9)	3	4
keine Angabe	10 (12)	2	1

Tabelle 4.1 Charakterisierung des Patientenkollektivs.

Bei allen 82 Patienten entstand der Verdacht auf das Vorliegen eines Rezidiv bei der durchgeführten CT-Diagnostik innerhalb der klinischen Nachsorge. Die Patienten erhielten daraufhin eine Bildgebung mittels FDG-PET/CT um ein Rezidiv zu verifizieren oder auszuschließen. Konnte das Vorliegen eines lokoregionären Rezidiv in der

PET/CT bestätigt werden, so war der primäre Endpunkt der PET-Plan Studie erreicht und es erfolgte keine weitere PET-Bildgebung im Rahmen der Studie. Lag laut PET/CT kein Lokalrezidiv oder ein Rezidiv in Form einer Metastasierung (sekundärer Endpunkt) vor, so blieben die Patienten weiterhin in der Nachbeobachtung und konnten weitere PET-Bildgebung bei erneutem Verdacht auf ein Lokalrezidiv erhalten. Aus diesem Grund liegen für einige der Patienten in dieser Studie mehr als nur ein PET/CT in der Nachsorge vor. Die Anzahl der PET/CT-Untersuchungen pro Patient und das Zeitintervall zwischen Therapiebeginn und erneuter PET/CT-Untersuchung sind der Tabelle 4.2 zu entnehmen. Von insgesamt 82 Patienten erhielten 64 genau eine PET/CT-Untersuchung und 18 mehr als eine PET/CT-Untersuchung im Zeitraum der Nachsorge.

Follow Up Daten	Gesamt
Zeitraum Therapiebeginn - Rezidivverdacht	
Median [Monate]	12
Mittelwert [Monate]	18
Minimum - Maximum [Wochen]	8 - 318
Anzahl der Untersuchungen	114
< 6 Monate nach Therapiebeginn	21
6 - 12 Monate nach Therapiebeginn	36
> 12 Monate nach Therapiebeginn	57
Anzahl der Patienten	82
Patienten mit n =1x FU-PET/CT	64
Patienten mit n > 1x FU-PET/CT	18
davon n=2	10
davon n=3	3
davon n=4	4
davon n=5	1

Tabelle 4.2 Zeitintervall und Anzahl der erfolgten PET/CTs im Follow Up.

Die Abbildung 4.3 stellt das vergangene Zeitintervall von Therapiebeginn der Radiochemotherapie bis zur PET/CT beim Zeitpunkt des Rezidivverdachts dar. Dem *Box Plot* ist in Ergänzung zur Tabelle 4.2 zu entnehmen, dass etwa die Hälfte aller PET/CT-Untersuchungen innerhalb von 12 Monaten nach Beginn der Radiochemotherapie erfolgte. Etwa ein Fünftel der Untersuchungen erfolgte bereits im ersten halben Jahr nach Beginn der Therapie.

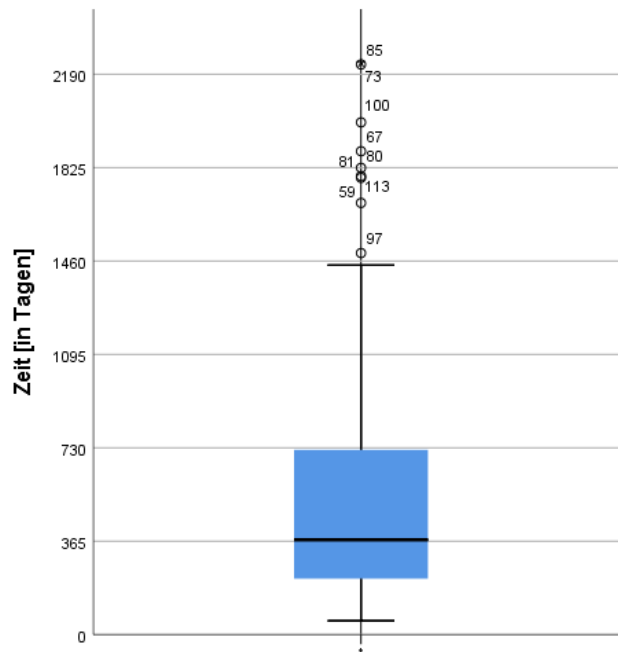


Abbildung 4.3 Zeitintervall zwischen Therapiebeginn und Rezidivverdacht.

4.1.1. Ergebnis des Goldstandards aus der klinischen Dokumentation

Die Zuordnung der primären und sekundären Endpunkte aus dem klinischen Dokumentationssystem der PET-Plan Studie ergab für das vorliegende Patientenkollektiv dieser Arbeit, die in der Tabelle 4.4 dargestellten Ergebnisse. Von 114 Fällen mit Rezidivverdacht, bestätigten sich 78 davon tatsächlich als Rezidiv. Davon lagen in 50 Fällen ein lokoregionärer Progress und in 56 Fällen eine Metastasierung vor. Hierbei lagen jeweils zu einem Drittel ein rein lokoregionärer Progress (22 von 78), eine Metastasierung (28 von 78) oder beides gleichzeitig (28 von 78) vor (nicht separat dargestellt). Die Lokalisationen pulmonal und mediastinal beziehen sich dabei rein auf den Ort des lokoregionären Rezidiv. Pulmonale Metastasen wurden zu der Kategorie Metastase gezählt.

	Anzahl	Pulmonal	Mediastinal	Lokoregionär	Metastase	Gesamt-Rezidiv
Gesamter Zeitraum						
Rezidiv	114 (100)	42 (36,8)	23 (20,2)	50 (43,9)	56 (49,1)	78 (68,4)
Zeitraum						
< 6 Monate	21 (18,4)	6	6	10	15	17
> 6 Monate	93 (81,6)	36	17	40	41	61

Tabelle 4.4 Häufigkeit und Lokalisation bei vorliegendem Rezidiv.

4.2. Die qualitative Auswertung durch die vier *Observer*

Die visuell geprägten, qualitativen Auswertungen der vier *Observer* wurden zunächst erfasst und für jeden *Observer* einzeln ausgewertet. Nach einer zusammenfassenden Darstellung der Ergebnisse, bezüglich der Rezidivbewertungen der einzelnen *Observer*, erfolgte die Berechnung der Übereinstimmungen zwischen den Urteilen. Die Farbkodierung der Übereinstimmungswerte in den Tabellen folgt der Interpretation nach Landis und Koch (Abb. 4.5). Für die Lunge erfolgte eine Risikoanalyse bezüglich der Faktoren, die zur Diagnose eines pulmonalen Rezidiv führten. Außerdem wurde der Einfluss der subjektiven Bewertungssicherheit überprüft. Abschließend wurden die Ergebnisse für die Lunge, das Mediastinum und Fernmetastasen einander gegenübergestellt.







Farbkodierung						
kappa-Wert	0	0 - 0,2	0,21 - 0,4	0,41 - 0,6	0,61 - 0,8	0,81 - 1,0
Übereinstimmung	schlecht	leicht	ausreichend	moderat	erheblich	(fast) perfekt

Abbildung 4.5 Interpretation des IOA nach Landis und Koch.

4.2.1. Überblick

Da auch der Einfluss der Kenntnisse über die Bestrahlungsplanung und das Staging untersucht werden sollte, fand die Auswertung einer jeden PET/CT durch die *Observer* immer doppelt statt: einmal ohne und einmal mit Kenntnissen über das *Staging* und die Bestrahlung. Tabelle 4.6 liefert einen Überblick, wie häufig sich die *Observer* für das Vorliegen eines Rezidivs entschieden haben, jeweils ohne und mit Kenntnis über die Ausgangsuntersuchung und die Bestrahlungsplanung.

Nach Kenntnis über die Bestrahlungsplanung und das Staging entschieden sich die vier *Observer* im Durchschnitt in 71 der 114 Fälle für das Vorliegen eines pulmonalen Rezidiv, was etwa 63 % der Fälle entspricht. Für das Vorliegen eines mediastinalen Rezidiv entschieden sich die *Observer* im Durchschnitt bei 38 %, entsprechend 41 von 114 Fällen. Fernmetastasen lagen laut Meinung der *Observer* bei 28 %, entsprechend 31 von 114 Fällen, vor. Der Observer-Konsens liegt mit 59 %, 28 % und 25 % für ein pulmonales, mediastinales Rezidiv und Fernmetastasen etwas unter dem gebildeten Mittelwert der *Observer*-Meinungen. Ebenso weichen Mittelwert der *Observer* und Konsens beim Gesamturteil voneinander ab, bei dem im Mittel bei 81 (72 %) und laut

4. Ergebnisse. Die qualitative Auswertung durch die vier Observer

Expertenkonsensus bei 75 der 114 Fälle (66 %) insgesamt ein Rezidiv vorlag. Der gepaarte t-Test zur Mittelwertabweichung zwischen den Entscheidungen der *Observer* ohne und mit Kenntnis über die Bestrahlungsplanung ergab mit $p = 0,218$ keinen signifikanten Unterschied zwischen beiden Szenarien.

Rezidivlokalisierung	Observer 1	Observer 2	Observer 3	Observer 4	Mittelwert	Observer-Konsens
Lunge, n (%)						
ohne Kenntnis	65 (57)	68 (60)	85 (75)	62 (54)	70 (61)	66 (58)
mit Kenntnis	71 (62)	71 (62)	85 (75)	60 (53)	71 (63)	67 (59)
Mediastinum, n (%)						
ohne Kenntnis	36 (32)	38 (33)	63 (55)	33 (29)	43 (37)	34 (30)
mit Kenntnis	37 (33)	38 (33)	65 (57)	32 (28)	41 (38)	32 (28)
Thorax, n (%)						
ohne Kenntnis	69 (61)	72 (63)	93 (82)	68 (60)	76 (67)	71 (62)
mit Kenntnis	75 (66)	73 (64)	93 (82)	66 (58)	77 (68)	70 (61)
Fernmetastasen, n (%)	30 (26)	31 (27)	41 (36)	27 (24)	31 (28)	28 (25)
Gesamturteil, n (%)						
ohne Kenntnis	74 (65)	78 (68)	100 (88)	72 (63)	79 (71)	76 (67)
mit Kenntnis	80 (70)	80 (70)	99 (87)	69 (61)	81 (72)	75 (66)

Tabelle 4.6 Urteilshäufigkeit für das Vorliegen eines Rezidiv in der Bewertung der *Observer*.

4.2.2. Die visuelle Auswertung der Lunge

Für die visuelle Auswertung der Lunge durch die *Observer* wurden die Lokalisation fokaler FDG-Akkumulationen, malignitätssuspekter FDG-Uptake innerhalb postradiogener Veränderungen und das Vorliegen intrapulmonaler Metastasen abgefragt. In Abwägung der vorliegenden Veränderungen entschied sich der *Observer* dann für oder gegen das Vorliegen eines pulmonalen Rezidiv. Die Werte für das IOA dieser Entscheidungen der vier *Observer* sind in der Tabelle 4.7 jeweils ohne und mit Kenntnissen über den Ausgangsbefund und die Bestrahlungsplanung dargestellt.

Aus dem Vergleich beider Tabellen geht hervor, dass sich die Übereinstimmung zum Urteil über fokale FDG-Akkumulationen im peripheren Lungenparenchym mit Kenntnissen über den Ausgangsbefund und die Bestrahlungsplanung verschlechtert. Sie sinkt von 0,440 (moderate Übereinstimmung) auf 0,388 (ausreichende Übereinstimmung) ab ($p = 0,090$). Gleichzeitig zeigt die Übereinstimmung über das Urteil vorliegender intrapulmonaler Metastasen eine deutliche Verbesserung von moderater zu erheblicher Übereinstimmung. Der Unterschied zwischen beiden Szenarien (0,563 vs 0,636) ist signifikant ($p = 0,048$). Bei den übrigen Übereinstimmungswerten herrscht kein signifikanter Unterschied zwischen beiden

4. Ergebnisse. Die qualitative Auswertung durch die vier Observer

Szenarien. Auffällig sind die niedrigen Übereinstimmungswerte der Beurteilung malignitätsverdächtiger Veränderungen bei vorliegenden postradiogenen Veränderungen. Diese befinden sich lediglich im Bereich leichter Übereinstimmung. Das Urteil über ein pulmonales Rezidiv liegt insgesamt für beide Szenarien im erheblichen Übereinstimmungsbereich.

Kriterium	IOV für Entscheidungen, <u>ohne</u> Kenntnis über Bestrahlungsplanung	IOV für Entscheidungen, <u>mit</u> Kenntnis über Bestrahlungsplanung
Fokale Akkumulation in 10-11R	0,553	0,511
Fokale Akkumulation in 10-11L	0,542	0,533
Zentrales Lungenparenchym	0,524	0,504
"gruppiert zentral"	0,613	0,590
Peripheres Lungenparenchym	0,440	0,388
Malignitätsverdächtiger Uptake in fibrotischem Areal	0,142	0,137
Pulmonale Metastase	0,563	0,636
Pulmonales Rezidiv	0,636	0,610

Tabelle 4.7 IOA für die Beurteilung der Lunge ohne und mit Kenntnis über die Ausgangsuntersuchung.

Um die Kausalität zwischen Bewertungskriterium und pulmonaler Rezidivbeurteilung untersuchen zu können, fand eine gruppierte Analyse des IOA statt. Dabei wurden die Kriterien zu fokalen FDG-Akkumulationen im zentralen Lungenparenchym und den Hili zusammengefasst, um den subjektiven Faktor der Ortszuordnung fokaler Akkumulationen auszuklammern. Der Wert ist in der Tabelle 4.7 als „gruppiert zentral“ bezeichnet. Davon abgeleitet, wurde im nächsten Schritt untersucht, welcher vorliegende Faktor als Risikokriterium mit dem Urteil eines pulmonalen Rezidiv in Verbindung steht. Hierzu wurden zunächst unabhängig für jedes einzelne Merkmal die Odds und das relative Risiko für das Eintreten eines positiven Urteils für ein

4. Ergebnisse. Die qualitative Auswertung durch die vier Observer

pulmonales Rezidiv berechnet. Die Tabelle 4.8 stellt diese Berechnungen unterteilt in beide Szenarien (ohne und mit Kenntnis der *Observer*) dar. Dabei stellt das Kriterium „fokal zentral“ wieder eine Zusammenfassung der Einzelkriterien „fokale Akkumulation im zentralen Lungenparenchym“ und der hilären Region „10-11R/L“ dar.

Ohne Kenntnis über die Bestrahlungsplanung

Kriterium	Odds Ratio	95%-Konfidenzintervall	Relatives Risiko
Fokale Akkumulation in 10-11R	3,186	2,062 - 4,924	1,480
Fokale Akkumulation in 10-11L	4,297	2,540 - 7,268	1,545
Zentrales Lungenparenchym	10,730	6,244 - 18,439	2,007
"gruppiert zentral"	9,021	5,842 - 13,930	3,167
Peripheres Lungenparenchym	31,818	13,637 - 74,238	2,185
Postradiogene Veränderung	0,946	0,615 - 1,453	0,979
Malignitätsverdächtiger Uptake in fibrotischem Areal	16,250	5,829 - 45,300	1,744
Pulmonale Metastase	29,412	11,690 - 73,556	2,041

Mit Kenntnis über die Bestrahlungsplanung

Kriterium	Odds Ratio	95%-Konfidenzintervall	Relatives Risiko
Fokale Akkumulation in 10-11R	3,829	2,384 - 6,148	1,520
Fokale Akkumulation in 10-11L	5,471	3,053 - 9,803	1,588
Zentrales Lungenparenchym	17,842	9,277 - 34,314	2,096
"gruppiert zentral"	11,894	7,485 - 18,899	3,115
Peripheres Lungenparenchym	15,778	7,996 - 31,132	1,972
Postradiogene Veränderung	0,885	0,566 - 1,383	0,956
Malignitätsverdächtiger Uptake in fibrotischem Areal	64,498	8,881 - 468,435	1,794
Pulmonale Metastase	51,183	15,965 - 164,096	2,075

Tabelle 4.8 Odds Ratio und relatives Risiko für die pulmonalen Risikokriterien (n=114).

In der Berechnung besitzen alle abgefragten Kriterien, bis auf eines, eine Odds Ratio von > 1 und erhöhen somit die Chance im Urteil zu einem pulmonalen Rezidiv zu gelangen. Das Kriterium einer postradiogenen pulmonalen Veränderung (unabhängig

4. Ergebnisse. Die qualitative Auswertung durch die vier Observer

von einer Wertung zur Dignität) besitzt eine Odds Ratio < 1 . Dieses Merkmal verringert also die Chance für ein pulmonales Rezidiv. Die höchsten Werte besitzen die Kriterien „intrapulmonale Metastase“ (OR = 51,183) und „malignitätssuspekter FDG-Uptake innerhalb postradiogener Veränderungen“ (OR = 64,498) jeweils nach Kenntnis der Observer über die Voruntersuchungen.

Um zu überprüfen welchen Einfluss die einzelnen Kriterien noch auf das Urteil eines pulmonalen Rezidiv haben, wenn die anderen Kriterien ebenfalls bekannt sind, wurde eine logistische Regressionsanalyse durchgeführt (Tabelle 4.9). Hier ist zu erkennen, dass alle Merkmale, wenn sie zutreffen dazu führen, dass das Urteil für das Vorliegen eines pulmonalen Rezidivs positiv ausfällt. Für die reine postradiogene Veränderung liegt die Odds Ratio auch in dem Logit Modell weiterhin < 1 und mindert somit die Chance zum Urteil „Rezidiv“ zu gelangen.

Ohne Kenntnis über die Bestrahlungsplanung

Kriterium	Odds/ Exp(B)	95%-Konfidenzintervall	Eintrittswahrscheinlichkeit
"gruppiert zentral"	15,837	7,775 - 32,259	0,941
Peripheres Lungenparenchym	43,604	15,406 - 123,407	0,978
Postradiogene Veränderung	0,907	0,465 - 1,769	0,476
Malignitätsverdächtiger Uptake in fibrotischem Areal	4,693	1,417 - 15,539	0,824
Pulmonale Metastase	40,725	13,460 - 123,218	0,976

Mit Kenntnis über die Bestrahlungsplanung

Kriterium	Odds/ Exp(B)	95%-Konfidenzintervall	Eintrittswahrscheinlichkeit
"gruppiert zentral"	34,751	15,417 - 78,327	0,972
Peripheres Lungenparenchym	30,113	10,827 - 83,756	0,968
Postradiogene Veränderung	0,664	0,320 - 1,382	0,399
Malignitätsverdächtiger Uptake in fibrotischem Areal	84,240	8,866 - 800,440	0,988
Pulmonale Metastase	119,268	30,066 - 473,125	0,992

Tabelle 4.9 Logistische Regressionsanalyse der pulmonalen Risikofaktoren (n=114).

4.2.3. Die Auswertung des Mediastinum

In der Bewertung des Mediastinums wurden im Unterschied zur Lunge keine Merkmale abgefragt. Hier mussten sich die *Observer* lediglich für oder gegen ein mediastinales Rezidiv bei einer sichtbaren FDG-Akkumulation entscheiden und die FDG-Belegung einem Lymphknotenlevel nach Mountain und Dressler zuordnen. Wie auch in der PET-Plan Studie wurden die Level 1 und 2, sowie 10 und 11 zusammengefasst abgefragt. In der Auswertung der einzelnen Lymphknotenstationen ergaben sich keine signifikanten Unterschiede zwischen den beiden bekannten Szenarien mit und ohne Kenntnis (nicht dargestellt). Zur besseren Übersicht wurden die Lymphknotenstationen als oberes (Level 1 – 3), mittleres (Level 4 – 6) und unteres (Level 7 und 8) Mediastinum zusammengefasst. Die Werte für das IOA sind in Tabelle 4.10 dargestellt. Die Kappa-Werte liegen im oberen Mediastinum im Bereich der erheblichen Übereinstimmung, im mittleren und unteren Mediastinum im Bereich der moderaten Übereinstimmung.

Lokalisation	IOV ohne Kenntnis über Baseline und Bestrahlung	IOV mit Kenntnis über Baseline und Bestrahlung
Oberes Mediastinum (1-2R, 1-2L, 3A und 3P)	0,651	0,668
Mittleres Mediastinum (4R, 4L, 5 und 6)	0,531	0,533
Unteres Mediastinum (7 und 8)	0,536	0,532

Tabelle 4.10 IOA für die Beurteilung der mediastinalen Lymphknotenstationen.

4.2.4. Urteil über Fernmetastasen

Die Kappa-Werte in der Beurteilung von extrathorakalen Fernmetastasen liegen ähnlich der Werte für pulmonale Metastasen im Bereich der erheblichen Übereinstimmung (Tabelle 4.11).

4.2.5. Die Gesamtauswertung der *Observer*

Zusammenfassend sind in Tabelle 4.11 noch einmal alle Kappa-Werte des IOA für die Lunge, das Mediastinum, Fernmetastasen und das Gesamturteil in der mittleren Spalte dargestellt. Für das IOA nach der Fragen eines pulmonalen Rezidiv und für

4. Ergebnisse. Die qualitative Auswertung durch die vier Observer

Fernmetastasen sind die Übereinstimmungswerte unabhängig vom Kenntnisstand erheblich. Für die Beurteilung eines mediastinalen Rezidiv und für die Frage ob insgesamt ein Rezidiv unabhängig von der Lokalisation vorliegt, sind die Werte im moderaten Übereinstimmungsbereich. Die Mittelwertunterschiede bezüglich des Kenntnisstandes sind nicht signifikant.

Um zu untersuchen, welchen Einfluss die subjektive Beurteilungssicherheit auf die Übereinstimmung der *Observer* nimmt, wurde zudem das IOA für sichere und unsichere Urteile bestimmt. Dazu erfolgte einmal die Aufteilung in Urteile bei denen sich die *Observer* sicher waren und Urteile bei denen sich mindestens einer der *Observer* unsicher war. Dabei lag der prozentuale Anteil der Urteile für die erste Kategorie zwischen 70 – 95 % aller ausgewerteten Fälle (nicht dargestellt). Für die Ergebnisse, bei denen sich die *Observer* sicher waren, liegt die Übereinstimmung der Beurteilung durchweg im substantiellen Bereich. Ist sich mindestens ein *Observer* im Urteil unsicher, so sinken die Werte in die Bereiche von ausreichender bis leichter Übereinstimmung.

Lokalisation	IOV für Entscheidungen, die beide Observer als sicher beurteilten	Gesamtergebnis der IOV	IOV für Entscheidungen, bei denen mind. ein Observer unsicher war
Lunge ohne Kenntnis	0,800	0,636	0,371
Lunge mit Kenntnis	0,737	0,610	0,219
Mediastinum ohne Kenntnis	0,683	0,572	0,160
Mediastinum mit Kenntnis	0,648	0,576	0,149
Urteil für Fernmetastasen	-	0,692	-
Gesamturteil ohne Kenntnis	0,664	0,559	0,304
Gesamturteil mit Kenntnis	0,658	0,574	0,235

Tabelle 4.11 Übersicht über das IOA der Gesamtauswertung der *Observer*.

4.2.6. Die subjektive Beurteilungssicherheit der *Observer*

Die subjektive Bewertungssicherheit der einzelnen Kriterien ist in Tabelle 4.12 abzulesen. Sie ist durchweg auf einem sehr hohen Niveau. Nach Kenntnis über die Bestrahlungsplanung und mit Vorliegen der Voruntersuchung steigen die Werte für die Beurteilungssicherheit der Lunge um 9,4 % und des Mediastinums um 4 % signifikant an ($p = 0,008$).

Lokalisation	Observer 1	Observer 2	Observer 3	Observer 4	Mittelwert
Lunge					
ohne Kenntnis	71,1%	77,2%	86,8%	78,9%	
mit Kenntnis	85,1%	91,2%	93,9%	81,6%	
Änderung (%)	14,0%	14,0%	7,1%	2,7%	9,5%
Mediastinum					
ohne Kenntnis	83,3%	93,9%	89,5%	88,6%	
mit Kenntnis	92,1%	95,6%	93,0%	89,5%	
Änderung (%)	8,8%	1,7%	3,5%	0,9%	3,7%

Tabelle 4.12 Beurteilungssicherheit der *Observer* (n=114).

4.3. Die quantitative Auswertung

4.3.1. Überblick

Die FDG-PET/CT-Auswertungen nach den PERCIST-Kriterien bieten sowohl Daten über die Baseline-Untersuchung, als auch über die Kontrolluntersuchungen. Die Parameter aus der Baseline-Untersuchung zum Zeitpunkt der Bestrahlungsplanung sind in Tabelle 4.13 dargestellt. Sie sind einmal zusammenfassend für das gesamte Patientenkollektiv und in Therapiearme unterteilt, abgebildet. Bemerkenswert ist, dass die SUV_{max} -Werte zwischen beiden Armen signifikant differieren und die der Patienten aus Arm A mit 17,5 deutlich über denen der Patienten aus Arm B mit 14,9 liegen ($p = 0,043$). Die Werte für MTV und TLG sind im Arm A wiederum niedriger als im Arm B. Einen Überblick über PET-Parameter zum Zeitpunkt der Bestrahlungsplanung gibt auch die Abbildung 4.14. Die Anzahl der positiven Lymphknotenstationen in der PET/CT zur Bestrahlungsplanung ist für beide Therapiearme ungefähr gleich (2,4 in Arm A vs. 2,6 in Arm B).

Tumorcharakteristika in der Planungs PET/CT*	Gesamt n= 82	Arm A konventionelle Bestrahlungsplanung n=42	Arm B PET-basierte Bestrahlungsplanung n=35
SULmean zu Therapiebeginn Mittelwert (SD)	5,6 (1,6)	5,7 (1,7)	5,4 (1,5)
SULpeak zu Therapiebeginn Mittelwert (SD)	10,1 (3,9)	10,9 (4,4)	9,4 (3,2)
SUVmax im Planungs PET/CT Mittelwert (SD)	16,4 (6,3)	17,5 (7,3)	14,9 (4,6)
Metabolic Tumor Volume zu Therapiebeginn Mittelwert (SD), [ml]	96 (86)	86 (79)	100 (92)
Total Lesion Glycolysis zu Therapiebeginn Mittelwert (SD)	575 (638)	527 (669)	609 (613)
Tumorlokalisation laut PET/CT, n (%)			
Lungenparenchym	66 (79)	31 (72)	32 (89)
Hilus	70 (83)	40 (93)	26 (72)
Mediastinum	62 (74)	31 (72)	26 (72)
Betroffene Lungenlappen laut PET/CT, n (%)			
rechter Oberlappen	26 (31,0)	16 (37)	10 (28)
rechter Mittellappen	2 (2,4)	0 (0)	1 (3)
rechter Unterlappen	9 (10,7)	3 (7)	4 (11)
linker Oberlappen	19 (22,6)	8 (18)	11 (30)
linker Unterlappen	8 (9,5)	3 (7)	5 (14)
mind. 2 Lappen ipsilateral	2 (2,4)	1 (2)	1 (3)
mind. 2 Lappen kontralateral	0 (0)	0 (0)	0 (0)
Tumor rein mediastinal	18 (21,4)	12 (28)	4 (11)
Anzahl PET-positiver Lymphknotenstationen Mittelwert (SD)	2,6 (1,6)	2,4 (1,4)	2,6 (1,7)
Haupttumorlast**, n (%)			
rechter Hemithorax	46 (55)	27 (63)	15 (42)
linker Hemithorax	35 (42)	16 (37)	19 (53)
keiner Seite zuzuordnen	3 (3)	0 (0)	2 (5)

* Schwelle lt. PERCIST über Background Leber, ** wenn Tumor Lunge+ oder Hilus+

Tabelle 4.13 Charakteristika der Baseline FDG-PET/CT.

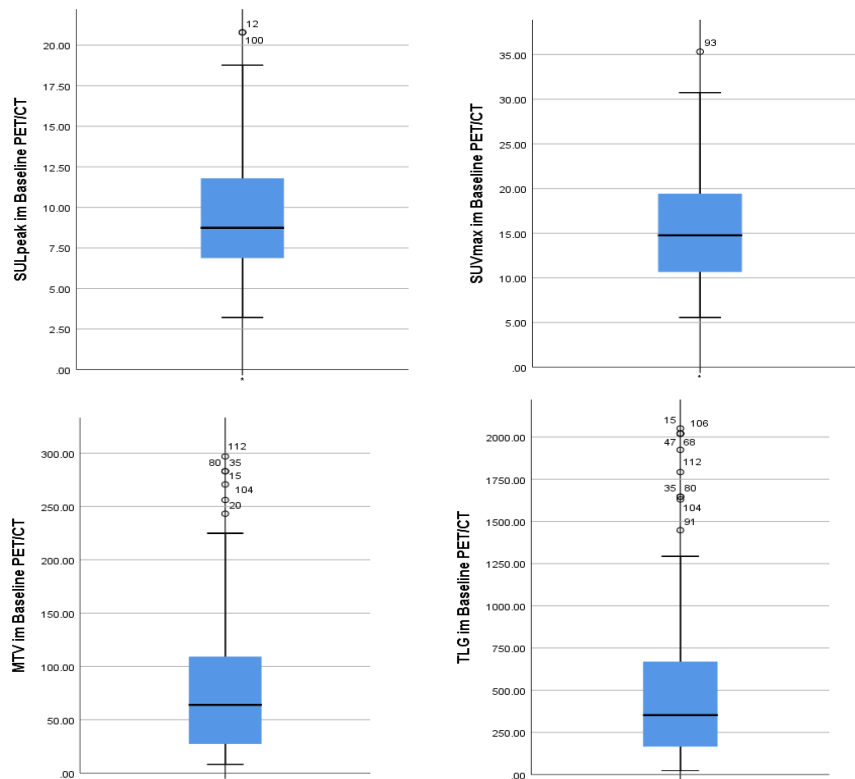


Abbildung 4.14 PET-Parameter zum Zeitpunkt der Bestrahlungsplanung.

4.3.2. Ergebnisse der PERCIST- und PREDIST-Auswertungen

Im Vergleich der PET/CT-Bilddaten zum Zeitpunkt der Bestrahlungsplanung mit denen zum Zeitpunkt des späteren Rezidivverdachts ergaben sich die in den Tabellen 4.15 und 4.16 dargestellten Ergebnisse. Die Tabelle 4.15 gibt die Häufigkeiten der 4 vergebenen Kategorien, einmal auf die Gesamtzahl der Fälle und einmal auf die Anzahl der Patienten bezogen, wider. Die Zahlen sind für den gesamten Zeitraum, sowie unterteilt in die ersten 6 Monate nach Therapiebeginn und über 6 Monate nach Therapiebeginn, aufgeschlüsselt.

Da die PERCIST-Kriterien zur Prüfung der Therapieansprache entworfen wurden, sind für deren Übertragung auf die Rezidivbeurteilung verschiedene Szenarien zur Interpretation der Kriterien als Rezidiv denkbar. Tabelle 4.16 gibt die unterschiedlichen Interpretationsmöglichkeiten der 4 Kategorien in PERCIST wider. Würde man lediglich die Kategorie PMD als Rezidiv werten, so ergäben sich in 46 von 114 Fällen ein Rezidiv. Nimmt man die Kategorie SMD als Kriterium hinzu, so ergeben sich mit 58 Fällen bei insgesamt 50,9 % ein Rezidiv. Wertet man die Kategorien PMR, SMD und PMR als Rezidiv, so käme man mit 78 von 114 Fällen bei insgesamt 68,4 % zum Vorliegen eines Rezidiv.

Zeitraum	Anzahl	CMR	PMR	SMD	PMD / Rezidiv
Alle Fälle, n (%)					
Gesamter Zeitraum	114 (100)	39 (34,2)	20 (17,5)	9 (7,9)	46 (40,4)
< 6 Monate	21 (18,4)	3 (2,6)	7 (6,1)	1 (0,9)	10 (8,8)
> 6 Monate	93 (81,6)	36 (31,6)	13 (11,4)	8 (7,0)	36 (31,6)
Letztes Follow-Up, n (%)					
Gesamter Zeitraum	82 (100)	20 (24,4)	16 (19,5)	7 (8,5)	39 (47,6)
< 6 Monate	16 (19,5)	0 (0)	6 (7,3)	1 (1,2)	9 (10,2)
> 6 Monate	66 (80,5)	20 (24,4)	10 (12,2)	6 (7,3)	30 (36,6)

Tabelle 4.15 Ergebnisse der PERCIST-Auswertung in Abhängigkeit vom Zeitpunkt.

	Anzahl	Pulmonal	Mediastinal	Fernmetastase	Gesamt-Rezidiv
Anzahl Rezidive, n (%)					
Rezidiv = PMD	114 (100)	25 (21,9)	8 (7,0)	28 (24,7)	46 (40,4)
Rezidiv = PMD, SMD	114 (100)	35 (30,7)	10 (8,8)	28 (24,7)	58 (50,9)
Rezidiv = PMD, SMD, PMR	114 (100)	54 (47,4)	12 (10,5)	28 (24,7)	78 (68,4)

Tabelle 4.16 Ergebnisse der PERCIST-Auswertung in Abhängigkeit von der Wertung als Rezidiv.

Zur besseren Vergleichbarkeit der quantitativen Auswertung mit der qualitativen Auswertung und dem Goldstandard, wurde eine modifizierte Einteilung der Auswertung nach dem PERCIST-Schema anhand der PREDIST-Kriterien vorgenommen. Diese unterteilen anhand der binären Merkmalsausprägung in CMR und *Residual Disease* (in dieser Arbeit als Rezidiv bezeichnet). Die Zuordnung der Lokalisation erfolgte zur besseren Vergleichbarkeit mit dem Goldstandard in die Kategorien lokoregionär (pulmonal und mediastinal) und/ oder Metastase (pulmonale und extrathorakale Metastasen). Die entsprechenden Ergebnisse sind in der Tabelle 4.17 abgebildet.

	Pulmonal	Mediastinal	Lokoregionär	Metastase	Gesamt-Rezidiv
Gesamter Zeitraum, n (%)					
Rezidiv	38 (33)	23 (20)	45 (40)	45 (39)	68 (60)
CMR	76 (67)	91 (80)	69 (61)	69 (61)	46 (40)
Rezidiv in Zeitraum, n					
< 6 Monate	7	7	11	12	15
> 6 Monate	31	16	34	33	53

Tabelle 4.17 Ergebnisse der Auswertung nach den PREDIST-Kriterien (n=114).

Nach PREDIST-Kriterien lag bei insgesamt 68 von 114 Fällen ein Rezidiv vor. Zu je ein Drittel lagen bei 22 von 68 Fällen nur ein lokoregionäres Rezidiv, bei 23 von 68 Fällen ein Rezidiv in Form von Metastasen und bei 23 von 68 Fällen sowohl ein lokoregionäres Rezidiv, als auch Metastasen vor. In den 21 von 114 Fällen, in denen ein Rezidivverdacht bereits in den ersten 6 Monaten nach Beginn der Radiochemotherapie zustande kam, ergab die PREDIST-Auswertung bei 15 von 21 Fällen das Vorliegen eines Rezidiv. Bei den 93 von 114 Fällen, bei denen der Rezidivverdacht mehr als 6 Monate nach Beginn der Radiochemotherapie zustande kam, lag nach PREDIST-Kriterien bei 53 Patientenfällen ein Rezidiv vor.

4.4. Vergleich der verschiedenen Auswertungen

4.4.1. Qualitative vs. quantitative Auswertung

Im Unterschied zum Auswertebogen, bei dem sich die *Observer* anhand der bewerteten Risikokriterien für oder gegen ein Rezidiv entscheiden sollten, liefert die quantitative Auswertung nach PERCIST-Kriterien keine binäre Einteilung für die Frage nach einem Rezidiv. Um die Übereinstimmung der Ergebnisse beider Auswertungen zu bestimmen, wurde das IOA bezüglich der binären Möglichkeiten ob ein Rezidiv vorliegt, oder nicht, ermittelt (Tab. 4.18). Dabei wurden drei Szenarien getestet, bei denen jeweils nur PMD oder PMD + SMD oder PMD + SMD + PMR als Rezidiv interpretiert wurden. Die Übereinstimmung wurde einmal für alle Bilddatensätze über den gesamten Zeitraum bestimmt (jeweils vorderste Spalte). Zusätzlich wurde eine zeitliche Unterteilung vorgenommen und das IOA für Fälle bestimmt, in denen der Rezidivverdacht innerhalb des ersten halben Jahres nach Therapiebeginn auftrat, sowie für den Zeitraum von mehr als 6 Monate nach Therapiebeginn. In der Beurteilung für die Lunge und das Mediastinum ergeben sich lediglich leichte bis ausreichende Übereinstimmungswerte. Die Übereinstimmungswerte verbessern sich, wenn SMD und PMR als Rezidiv gewertet werden. Das Vorliegen von Fernmetastasen, das in der PERCIST-Auswertung immer als Progress zu werten ist, zeigt hingegen erhebliche Übereinstimmungswerte.

Kriterium	Rezidiv = PMD			Rezidiv = PMD + SMD			Rezidiv = PMD + SMD + PMR		
	Gesamter Zeitraum	Zeitraum < 6 Monate	Zeitraum > 6 Monate	Gesamter Zeitraum	Zeitraum < 6 Monate	Zeitraum > 6 Monate	Gesamter Zeitraum	Zeitraum < 6 Monate	Zeitraum > 6 Monate
Pulmonales Rezidiv	0,297	0,075	0,379	0,306	-0,120	0,424	0,458	0,035	0,545
Mediastinales Rezidiv	0,324	0,208	0,364	0,305	0,108	0,389	0,375	0,108	0,486
Fernmetastasen	0,716	0,690	0,718	0,716	0,690	0,718	0,716	0,690	0,718
Gesamturteil Rezidiv	0,487	0,352	0,511	0,518	0,228	0,583	0,698	0,492	0,717

Tabelle 4.18 IOA zwischen *Observer*-Konsens und PERCIST-Auswertung.

Die niedrigen Übereinstimmungswerte in Tab. 4.18 führten zur Reevaluation der Einteilung zur Rezidivlokalisierung und zur vereinfachten Interpretation der quantitativen

Auswertung nach PREDIST-Kriterien. In der weiteren Einteilung und zum späteren Vergleich mit dem Goldstandard wurde das Kriterium „pulmonale Metastase“ in der *Observer*-Auswertung den Metastasen zugeordnet und nicht mehr zum pulmonalen Rezidiv gezählt.

Im Vergleich des IOA zwischen *Observer*-Konsens und der quantitativen Auswertung nach den PREDIST-Kriterien ergeben sich hingegen Werte im moderaten bis erheblichen Übereinstimmungsbereich (Tabelle 4.19). In der Gesamtbeurteilung bei Patientenfällen über 6 Monaten nach Therapiebeginn liegt die Übereinstimmung sogar im fast perfekten Bereich. Lediglich die Beurteilung eines pulmonalen Rezidiv innerhalb der ersten 6 Monate nach Therapiebeginn liegt weiterhin im Bereich leichter Übereinstimmung.

Kriterium	Gesamter Zeitraum	Zeitraum < 6 Monate	Zeitraum > 6 Monate
Pulmonales Rezidiv	0,458	0,108	0,558
Mediastinales Rezidiv	0,625	0,516	0,646
Metastasen	0,766	0,611	0,796
Lokoregionäres Rezidiv	0,630	0,512	0,642
Gesamturteil Rezidiv	0,775	0,588	0,810

Tabelle 4.19 IOA zwischen *Observer*-Konsens und PREDIST-Auswertung.

4.4.2. Vergleich der qualitativen Auswertung mit dem Goldstandard

Um die Güte der qualitativen Auswertung zu überprüfen, erfolgte die Bestimmung von Sensitivität und Spezifität der *Observer*-Auswertung anhand des Goldstandards. Tabelle 4.20 stellt beide Testgütekriterien einmal ohne und einmal mit Kenntnis der *Observer* über das Staging und die Bestrahlung dar. Mit entsprechenden Kenntnissen erhöht sich die Sensitivität der qualitativen Auswertung für das Erkennen eines Rezidiv insgesamt von 84,6 % auf 92,3 %. Die Spezifität des Gesamturteils steigt von 86,1 %

4. Ergebnisse. Vergleich der verschiedenen Auswertungen

auf 88,9 %. Dabei liegt in der Gesamtbeurteilung nach Kenntnis über Ausgangsbefund und Bestrahlung die Anzahl richtig-positiver Urteile bei 72, die der falsch-negativen Ergebnisse bei 6. Für die Berechnung der entsprechenden Spezifität lagen 31 richtig-negative und 4 falsch-positive Urteile vor.

Ohne Kenntnis über Bestrahlungsplanung

Kriterium	Sensitivität	Spezifität
Pulmonales Rezidiv	0,643 (27/42)	0,722 (52/72)
Mediastinales Rezidiv	0,826 (19/23)	0,835 (76/91)
Lokoregionäres Rezidiv	0,780 (39/50)	0,703 (45/64)
Metastasen	0,732 (41/56)	0,879 (51/58)
Gesamturteil Rezidiv	0,846 (66/78)	0,861 (31/36)

Mit Kenntnis über Bestrahlungsplanung

Kriterium	Sensitivität	Spezifität
Pulmonales Rezidiv	0,738 (31/42)	0,708 (51/72)
Mediastinales Rezidiv	0,826 (19/23)	0,846 (77/91)
Lokoregionäres Rezidiv	0,780 (39/50)	0,734 (47/64)
Metastasen	0,786 (44/56)	0,897 (52/58)
Gesamturteil Rezidiv	0,923 (72/78)	0,889 (32/36)

Tabelle 4.20 Sensitivität und Spezifität der qualitativen Auswertung (n=114).

Kriterium	Sensitivität	Spezifität
Pulmonales Rezidiv	0,762 (32/42)	0,917 (66/72)
Mediastinales Rezidiv	0,826 (19/23)	0,956 (87/91)
Lokoregionäres Rezidiv	0,800 (40/50)	0,922 (59/64)
Metastasen	0,786 (44/56)	0,983 (57/58)
Gesamturteil Rezidiv	0,859 (67/78)	0,972 (35/36)

Tabelle 4.21 Sensitivität und Spezifität der quantitativen Auswertung (n=114).

4.4.3. Vergleich der quantitativen Auswertung mit dem Goldstandard

Die Ergebnisse der Berechnung von Sensitivität und Spezifität für die PREDIST-Kriterien sind in Tabelle 4.21 dargestellt. Die Spezifität der Auswertung ist für alle Rezidivlokalisationen in einem durchweg sehr hohen Bereich angesiedelt und ergibt für das Gesamturteil über das Vorliegen eines Rezidiv eine Spezifität von 97,2 %. In der entsprechenden Kreuztabelle lagen hierbei lediglich ein falsch-positives und 35 richtig-negative Urteile vor. Die Sensitivität der quantitativen Rezidivbeurteilung von 85,9 % ergibt sich aus 67 richtig-positiven und 11 falsch-negativen Ergebnissen.

4.4.4. Die zeitliche Abhängigkeit der Sensitivität und Spezifität

Die Abhängigkeit der Testgütekriterien vom Zeitpunkt der Auswertung nach Therapiebeginn ist in der Tabelle 4.22 dargestellt und wieder in die Zeitspannen innerhalb und nach 6 Monaten unterteilt. Der Observer-Konsens mit Kenntnis über die Bestrahlungsplanung wird der PREDIST-Auswertung gegenübergestellt. Hervorzuheben ist, dass die Sensitivität der visuellen Auswertung durch die *Observer* sowohl für die Lunge, als auch das Lokalrezidiv und das Gesamturteil in den ersten sechs Monaten höher ist, als im Zeitraum danach. Die quantitative Auswertung erzielt im Zeitraum von mehr als 6 Monaten nach Therapiebeginn eine Spezifität von 100 % für das Vorliegen eines Rezidiv in der Gesamtbeurteilung.

Rezidivverdacht < 6 Monate nach Radiochemotherapie Rezidivverdacht > 6 Monate nach Radiochemotherapie

Kriterium	Observer		PREDIST		Kriterium	Observer		PREDIST	
	Sensitivität	Spezifität	Sensitivität	Spezifität		Sensitivität	Spezifität	Sensitivität	Spezifität
Pulmonales Rezidiv	1,000 (6/6)	0,333 (5/15)	0,667 (4/6)	0,800 (12/15)	Pulmonales Rezidiv	0,694 (25/36)	0,807 (46/57)	0,778 (28/36)	0,947 (54/57)
Mediastinales Rezidiv	0,667 (4/6)	0,600 (9/15)	0,667 (4/6)	0,800 (12/15)	Mediastinales Rezidiv	0,882 (15/17)	0,895 (68/76)	0,882 (15/17)	0,987 (75/76)
Lokoregionäres Rezidiv	0,900 (9/10)	0,364 (4/11)	0,800 (8/10)	0,727 (8/11)	Lokoregionäres Rezidiv	0,750 (30/40)	0,811 (43/53)	0,800 (32/40)	0,962 (51/53)
Metastasen	0,733 (11/15)	0,833 (5/6)	0,800 (12/15)	1,000 (6/6)	Metastasen	0,805 (33/41)	0,904 (47/52)	0,780 (32/41)	0,981 (51/52)
Gesamturteil Rezidiv	0,941 (16/17)	0,500 (2/4)	0,824 (14/17)	0,750 (3/4)	Gesamturteil Rezidiv	0,918 (56/61)	0,938 (30/32)	0,869 (53/61)	1,000 (32/32)

Tabelle 4.22 Sensitivität und Spezifität in Abhängigkeit des Zeitpunktes des Rezidivverdachts (n=114).

4.5. Evaluation der PET-Parameter

4.5.1. Einfluss der Parameter aus der Baseline PET/CT auf das Outcome

Um zu überprüfen, ob eine Prognoseabschätzung des Erkrankungsverlaufs durch die PET-Parameter in der Bestrahlungsplanung möglich ist, wurde für die entsprechenden Parameter eine Cox-Regressionsanalyse durchgeführt (Tabelle 4.23). Die Zeitspanne wurde gebildet durch den Beginn der Therapie bis zum aufkommenden Rezidivverdacht. Als Ereignis galt die Diagnose des lokoregionären Rezidiv anhand des Goldstandards.

Die Ergebnisse für SUL_{peak} und SUV_{max} sind nicht zu interpretieren, da das Konfidenzintervall den Wert 1 einschließt. Die Odds für die Parameter MTV und TLG sind verwertbar und statistisch signifikant.

PET-Parameter	Odds/ Exp(B)	95%-Konfidenzintervall	Signifikanz
SUL_{peak}	1,050	0,996 - 1,107	0,071
SUV_{max}	1,025	0,993 - 1,058	0,122
MTV	0,987	0,977 - 0,997	0,009
TLG	1,002	1,001 - 1,004	0,003

Tabelle 4.23 Cox-Regression der PET-Parameter aus der Baseline PET/CT (n=82).

Um einen eventuell vorhandenen, aber nicht linearen Einfluss der PET-Parameter auf das *Outcome* der Patienten abschätzen zu können, wurden die Parameter zum Baseline Zeitpunkt anhand ihres jeweiligen Medians im Fallkollektiv in zwei Gruppen aufgeteilt. Die eine Hälfte der Fälle besitzt einen oberhalb des Medians liegenden und die andere Hälfte einen unterhalb des Medians liegenden Parameterwert. Beide Hälften wurden in einem Kaplan-Meier-Schätzer einander gegenübergestellt und die Güte des Schätzers mittels Log Rank Test geprüft. Die Zeitspanne wurde gebildet durch den Beginn der Therapie bis zum aufkommenden Rezidivverdacht. Als Ereignis galt die Dokumentation eines lokoregionären Rezidiv im Goldstandard. Die jeweiligen Kaplan-Meier-Kurven sind in der Abbildung 4.24 zusammengefasst dargestellt. Dabei bildet die rote Kurve immer die über dem Median liegende Gruppe ab.

Für die Parameter SUV_{max} , MTV und TLG ergibt die Testung keinen signifikanten Unterschied zwischen beiden Gruppen. Ein statistisch signifikanter Unterschied zwischen den beiden Gruppen ist lediglich für den Parameter SUL_{peak} gegeben. Wobei Patienten mit einem unterhalb des Medians der Kohorte liegenden SUL_{peak} im Verlauf weniger häufig ein Gesamtrezidiv erleiden.

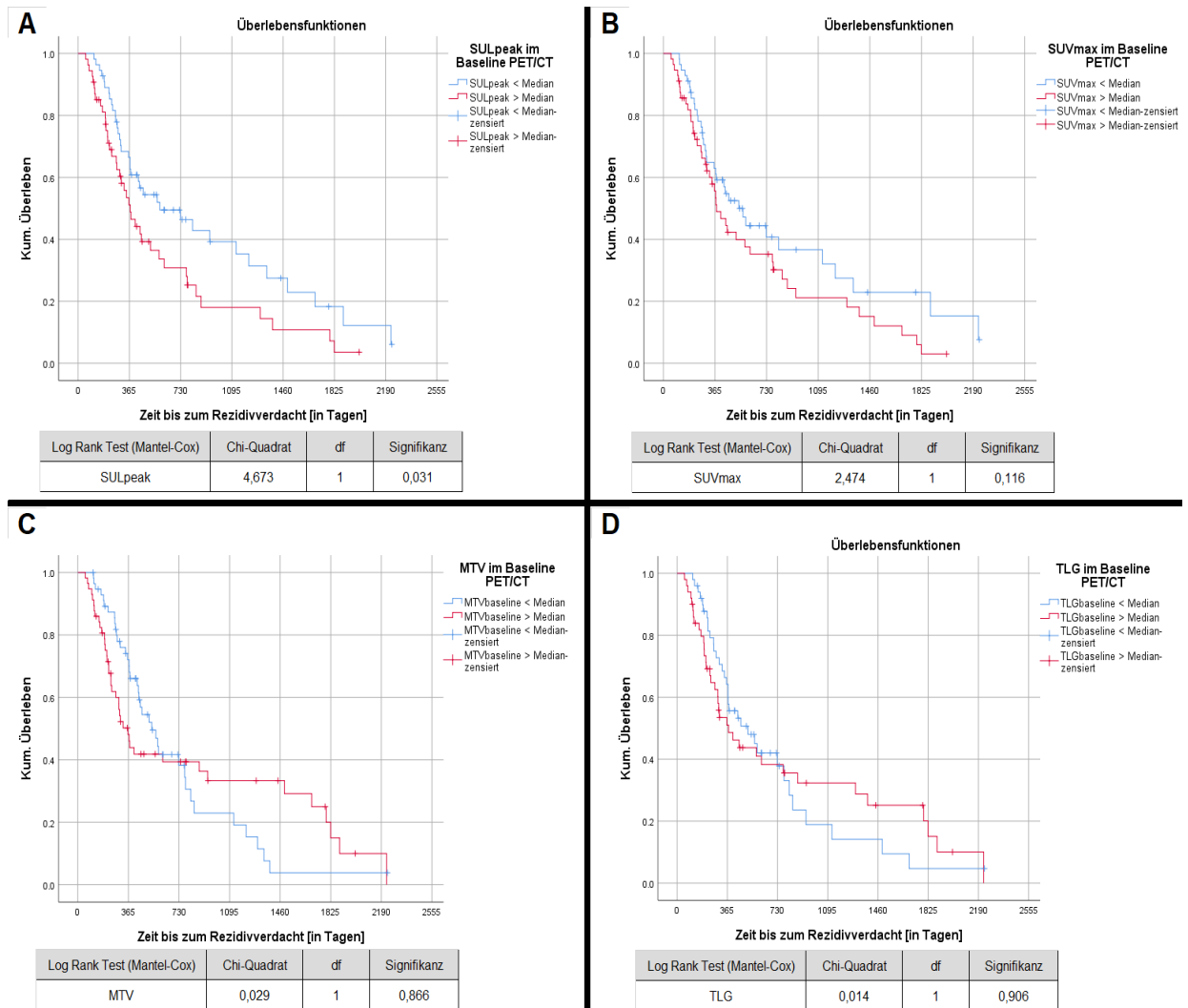


Abbildung 4.24 Prognose der PET-Parameter zum Outcome der Patienten (n=82).

Die Abbildung stellt die Kaplan-Meier-Kurven und unter der Kurve das Ergebnis der Log Rank Testung dar, unterteilt für die Baseline-Parameter SUL_{peak} (A), SUV_{max} (B), MTV (C) und TLG (D).

4.5.2. PET-Parameter in der Rezidivdiagnostik

Neben der Aussagekraft der PET-Parameter zum Baseline-Zeitpunkt wurden auch die Parameter SUL_{peak} und SUV_{max} zum Zeitpunkt des Rezidivverdachts untersucht. Die Werte für SUL_{peak} und SUV_{max} sind in der Abbildung 4.25, aufgeteilt in vorliegendes oder nicht vorliegendes Gesamtrezidiv im Follow Up, abgebildet. Zwischen beiden Gruppen besteht ein jeweils signifikanter Mittelwertunterschied ($p < 0,001$).

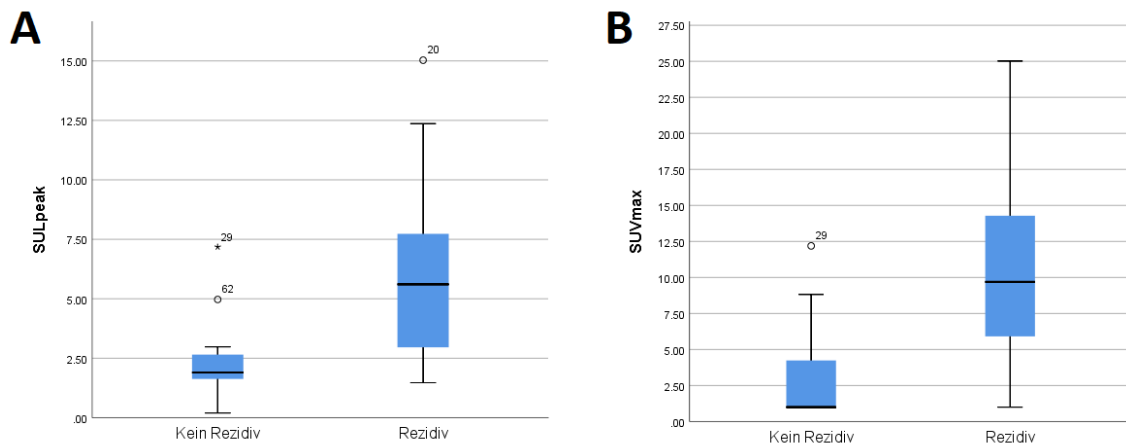


Abbildung 4.25 Die PET-Parameter SUL_{peak} (A) und SUV_{max} (B) im Follow Up (n=114).

Um zu überprüfen, welche Aussagekraft die Parameter für sich allein genommen über das Vorliegen eines Rezidivs besitzen, wurde die Sensitivität und Spezifität für bestimmte Schwellenwerte von SUL_{peak} und SUV_{max} anhand der Observermeinung und dem Goldstandard ermittelt (Tabelle 4.26). Dabei ergibt sich bei höheren Werten für SUL_{peak} und SUV_{max} auch eine höhere Spezifität in der Wertung, bei jedoch sinkender Sensitivität.

PET-Parameter	Observer-Meinung		Goldstandard	
	Sensitivität	Spezifität	Sensitivität	Spezifität
$SUL_{peak} > 2,0$	0,908 (69/76)	0,605 (23/38)	0,885 (69/78)	0,583 (21/36)
$SUL_{peak} > 3,0$	0,737 (56/76)	0,921 (35/38)	0,731 (57/78)	0,944 (34/36)
$SUV_{max} > 3,5$	0,908 (69/76)	0,658 (25/38)	0,885 (69/78)	0,639 (23/36)
$SUV_{max} > 4,5$	0,855 (65/76)	0,789 (30/38)	0,833 (65/78)	0,778 (28/36)
$SUV_{max} > 6,0$	0,776 (59/76)	0,974 (37/38)	0,731 (57/78)	0,917 (33/36)

Tabelle 4.26 Sensitivität und Spezifität von SUL_{peak} und SUV_{max} im Follow Up-PET (n=114).

5. Diskussion

5.1. Das Patientenkollektiv

Der Interpretation der Ergebnisse müssen die zwei folgenden Tatsachen zur Berücksichtigung vorangestellt werden:

1. Die Auswahl des Patientenkollektivs aus dem Kollektiv der PET-Plan Studie unterliegt nach Definition der Fragestellung dieser Arbeit bereits einer gewissen Vorselektion. So wurden nur Patienten eingeschlossen, bei denen zuvor bereits eine CT-Bildgebung erfolgte, anhand derer ein Rezidivverdacht entstand. So herrscht im hier vorliegenden Patientenkollektiv, wie auch in den meisten Studien zur Rezidivdiagnostik mittels FDG-PET/CT, eine potentiell hohe Prävalenz vor, welches einen Vorteil für den positiven Vorhersagewert der Testung darstellt. Die Untersuchung zur Wertigkeit der PET in der Rezidivdiagnostik bildet damit dennoch den klinischen Alltag und das somit am nächsten an der Wahrheit liegende Szenario ab, da die Patienten im Rahmen der Tumornachsorge ebenfalls zunächst eine CT-Untersuchung bekommen und bei Rezidivverdacht der PET/CT zugeführt werden. Die in der Tabelle 4.4 genannten absoluten Zahlen an Rezidiven sollten dennoch vorsichtig interpretiert werden, da sie demnach nicht den wahren Prozentsatz aller behandelten Patienten widerspiegeln.

2. Die Gesamtanzahl des vorliegenden Studienkollektivs, die etwas mehr als ein Viertel des Gesamtkollektivs der PET-Plan Studie beträgt, spiegelt nicht die vollständige Anzahl an Patientenrezidiven nach Radiochemotherapie wider. Berücksichtigt werden muss, dass auch einige Fälle innerhalb der Originalstudie vorlagen, die nicht im normalen Rahmen nachuntersucht werden konnten, weil sie zum Beispiel verstarben. Des Weiteren lagen zu Beginn der Auswertung nicht aus allen Zentren Bilddaten vor, sodass nicht alle Patienten der PET-Plan Studie, die ein Rezidiv erlitten, auch in diese Studie mit eingeschlossen werden konnten. Umgekehrt sind in dieser Arbeit einige Patienten eingeschlossen, die in der Nachsorge mehrfach eine PET-Bildgebung erhalten haben. Verhindert wurde hierbei zwar, dass ein lokoregionärer Progress mehrfach abgebildet wurde, da die Patienten nach lokoregionärem Progress ausgeschlossen wurden. In fünf Szenarien wiesen Patienten allerdings im Verlauf zunächst Metastasen auf, bevor bei zwei von ihnen dann ein lokoregionäres Rezidiv eintrat, welches in einer weiteren PET/CT abgebildet war. Die Metastasen wurden mittels systemischer Chemotherapie oder lokaler Radiotherapie

behandelt und trugen somit im weiteren Verlauf zur erschwerten Beurteilung bei. Obgleich auch diese Szenarien nah am tatsächlichen klinischen Alltag liegen, sollte dieser Tatsache bei der Interpretation der Ergebnisse im Gesamtkollektiv von 114 Fällen Rechnung getragen werden.

Die in Tabelle 4.1 aufgelisteten und in Therapiearme unterteilten Patientencharakteristika dieser Untersuchung spiegeln die erfolgreiche Randomisierung der Patienten innerhalb der PET-Plan Studie wider. So herrscht auch in dieser Studie eine ausgewogene Verteilung der Patienten bezogen auf die Therapiearme vor, was sich beispielsweise anhand des LBW und der Geschlechterverteilung ablesen lässt. Die mediane Zeitspanne, die zwischen Beginn der Radiochemotherapie und dem Aufkommen des Rezidivverdachts lag, betrug 12 Monate. Der Mittelwert dieser Zeitspanne lag bei 18 Monaten (Tab 4.2). Rund die Hälfte aller Untersuchungen erfolgte am Ende des ersten Jahres nach Therapiebeginn oder im darauffolgenden zweiten Jahr (Abb. 4.3). Dies deckt sich mit den Beobachtungen von Boyd *et al.*, die einen ersten Höhepunkt der Rezidivraten beim NSCLC etwa 13 Monate nach Therapieende beschreiben (54).

5.2. Die qualitative Auswertung

Bei der Übersicht über die Urteilshäufigkeiten der *Observer* (Tab 4.6) zeigt sich, dass die Zahlen, bezogen auf die einzelnen Lokalisationen für die *Observer* 1, 2 und 4 in etwa derselben Größenordnung liegen. *Observer* 3 weicht mit einer höheren Quote an Rezidivurteilen von den anderen *Observern* ab. Dies erklärt auch den Unterschied zwischen den, aus den Urteilen gebildeten Mittelwerten und dem *Observer*-Konsens. Der Konsens liegt in den absoluten Zahlenwerten immer unterhalb des *Observer*-Mittelwertes, was insgesamt für ein sehr sensibles *Reading* im Sinne der visuellen Kriterien nach Hicks *et al.* spricht (42). Einzelnen, aus Unsicherheit resultierenden potentiellen Fehlentscheidungen, wie beispielsweise einen rein entzündlichen Prozess als Rezidiv zu werten, wird in der Konsens-Wertung so entgegengewirkt.

5.2.1. Die visuelle Auswertung der Lunge

Interobserver-Agreement

Die bloße Identifikation fokaler FDG-Akkumulationen im Lungenparenchym bietet in der PET/CT, unabhängig von einer zentralen oder peripheren Lokalisation, keine

Schwierigkeiten, wie an den durchweg moderaten Übereinstimmungswerten der *Observer* in Tabelle 4.7 zu erkennen ist. Nach Zusammenfassen der Kriterien für die Abgrenzung zwischen hilärer und zentraler, parenchymaler FDG-Anreicherung als „gruppiert zentral“, verbessern sich die Übereinstimmungswerte weiter in den erheblichen Übereinstimmungsbereich. Dies spricht dafür, dass die anatomische Zuordnung nicht in allen Fällen immer eindeutig möglich ist. Hier erschwert vor allem die genaue Definition der Grenzen zwischen hilären Lymphknoten und Beginn des zentralen Lungenparenchyms eine exakte Zuordnung und einheitliche Interpretation. Die Betrachtung der Ausgangsuntersuchung zum Staging nimmt keinen Einfluss auf die Übereinstimmungswerte der *Observer* für fokale FDG-Akkumulationen im zentralen Lungenparenchym und der Hilusregion.

Im Gegensatz dazu ändert sich jedoch die Übereinstimmung für die Zuordnung fokaler FDG-Akkumulationen zwischen rein fokaler Akkumulation im peripheren Lungenparenchym und pulmonaler Metastase mit dem Hinzukommen der Kenntnisse über Voruntersuchung und Bestrahlung. Mit weiteren Informationen nimmt also die Übereinstimmung zur Zuordnung einer rein peripheren FDG-Akkumulation von moderat zu ausreichend ab, während die Übereinstimmung der Zuordnung zur pulmonalen Metastase vom moderaten in den erheblichen Bereich ansteigt. Dies spricht dafür, dass nach Kenntnis über den Ausgangsbefund ein Teil der zuvor wahrgenommenen, peripheren FDG-Akkumulationen nun sicher der Kategorie einer pulmonalen Metastase zugeordnet oder davon ausgeschlossen werden können. Zwischen beiden Szenarien besteht ein jeweils signifikanter Mittelwertunterschied, der bestätigt, dass der beobachtete Effekt tatsächlich existiert.

Auffällig sind die sehr niedrigen Übereinstimmungswerte für die Kategorie des malignitätsverdächtigen *FDG-Uptakes* in postradiogenen, fibrotischen Lungenbezirken, die in dieser Arbeit lediglich im Bereich leichter Übereinstimmung liegen. Diese Beobachtung steht in Einklang mit der Beschreibung aus den Literaturquellen, wonach vor allem die nach Radiotherapie auftretenden Lungenparenchymveränderungen die Beurteilung der FDG-PET/CT in der Rezidivdiagnostik erschweren (64). Dieser Faktor führt auch in dieser Studie zu hohen Unstimmigkeiten und erschwert die sichere Beurteilung in der Auswertung der *Observer*. Dabei liegt die Größenordnung der durch entzündliche, postradiogene Veränderungen verursachten, falsch-positiven Befunde in der Literatur bei etwa 33 % aller falsch-positiven PET-Befunde (42). In der statistischen Auswertung dieser Arbeit liegt diese Anzahl unter den in der Literatur zu findenden Zahlen. Dies dürfte ebenfalls

dadurch begünstigt sein, dass die überwiegende Mehrheit der Patienten erst in einem Zeitraum nach 6 Monaten untersucht wurde. Dann liegen meist keine, durch die Bestrahlung bedingten, entzündlichen Prozesse mit FDG-*Uptake* mehr vor. Um dieser Hypothese noch weiter nachzugehen, könnte man in einer weiteren Untersuchung zusätzlich ermitteln, in welchem Zeitraum nach Bestrahlungsbeginn, besonders niedrige Übereinstimmungswerte zustande kommen. Wobei dann auch genauer definiert werden müsste, ab wann FDG-Akkumulationen in postradiogenen Veränderungen definitiv als maligne einzustufen sind.

Ein signifikanter Unterschied in der Bewertung dieser Veränderungen ohne und mit Kenntnis über die Voruntersuchung und die Bestrahlung ist dabei nicht fest zu stellen. Die Kenntnis über Voruntersuchung und Bestrahlung bieten, im Gegensatz zur Sensitivität und Spezifität der *Observer*-Urteile (siehe Kapitel 5.2.6), keinerlei Verbesserung der Übereinstimmungswerte. Die lediglich ausreichenden Übereinstimmungswerte für die Dignitätsbeurteilung postradiogener Veränderungen wirken sich somit auch negativ auf die Übereinstimmung der Gesamtbewertung auf das Vorliegen eines pulmonalen Rezidiv aus. Wobei diese mit erheblichen Übereinstimmungswerten noch in einem sehr guten Bereich liegen.

Odds Ratio, relatives Risiko und das Logit Modell

Um den Einfluss der abgefragten Kriterien auf das Urteil eines möglichen pulmonalen Rezidiv aber genauer bewerten zu können, wurden in der Tabelle 4.8 die *Odds* und das relative Risiko für die Kriterien bestimmt. Hierfür wurde zunächst für jedes Kriterium einzeln betrachtet, inwiefern es die Chance erhöht, dass der *Observer* sich bei Vorliegen dieses Risikomerkmals für oder gegen ein pulmonales Rezidiv entscheidet. Wie zu erwarten, steigern alle Kriterien, bis auf die reinen postradiogenen Veränderung ohne Dignitätszuteilung, das Risiko für das Vorliegen eines pulmonalen Rezidiv. Sowohl die *Odds Ratio*, als auch das relative Risiko für postradiogene Veränderungen ohne Bewertung der Dignität, besitzen einen Wert, der kleiner als Eins ist und führen somit eher zur Bewertung, dass kein pulmonales Rezidiv vorliegt.

Diese Tatsache erscheint zwar logisch und zunächst recht simpel, ist aber gleichzeitig in ihrer Bedeutung für die Auswertung essentiell, bestätigt sie doch die korrekte Auslegung der vorgegebenen Kriterien durch die *Observer*. Sie steht für die korrekte Interpretation der *Observer* im Sinne der *Reading*-Kriterien nach *Hicks et al.*, wonach beispielsweise flächige FDG-Akkumulationen eher als benigne eingestuft werden

sollten (42, 53). Einschränkend muss aber darauf verwiesen werden, dass das zugehörige Konfidenzintervall die Zahl Eins einschließt und die Testung somit formal nicht verlässlich beurteilbar ist.

Der Vergleich der *Odds Ratio*, jeweils ohne und mit Kenntnis über die Bestrahlungsplanung, zeigt, dass diese für die fokale FDG-Akkumulation in den zentralen Regionen etwa gleich bleibt. Die Chance und das Risiko für den Faktor „fokale FDG-Akkumulation im peripheren Lungenparenchym“ zum Urteil eines pulmonalen Rezidiv zu führen, sinkt hingegen ab, während die jeweiligen Werte für den Risikofaktor „pulmonale Metastase“ ansteigen. Dies bestätigt die Schlussfolgerungen aus den Berechnungen zum IOA, wonach das Vorliegen der Ausgangsuntersuchung die eindeutige Identifikation einer intrapulmonalen Metastase in Abgrenzung zur bloßen peripheren FDG-Akkumulation erleichtert. Nach Kenntnis über die Ausgangsuntersuchung stellen vor allem die Möglichkeit einer intrapulmonalen Metastase oder das Vorliegen von malignitäts-verdächtigem FDG-*Uptake* innerhalb fibrotischer Areale, Risikofaktoren für das Urteil eines pulmonalen Rezidiv dar.

Um nun zu überprüfen welchen Einfluss die einzelnen Kriterien noch auf das Urteil eines pulmonalen Rezidiv haben, wenn die anderen Kriterien ebenfalls bekannt sind, wurde die in Tabelle 4.9 dargestellte logistische Regressionsanalyse durchgeführt. Da nicht alle im Fragebogen aufgelisteten Kriterien unabhängig voneinander zu bewerten sind, wurden die Hili und das zentrale Lungenparenchym im Logit Modell wieder als „gruppiert zentral“ zusammengefasst. Durch die Gruppierung wird die Tatsache der fokalen Akkumulation ortsunabhängig abgebildet und lässt somit eine Kausalität ableiten, wann welches Kriterium zum Urteil eines pulmonalen Rezidiv führt.

Auch hier ist zu erkennen, dass alle Kriterien, wenn sie als zutreffend bewertet wurden, dazu führen, dass das Urteil für das Vorliegen eines pulmonalen Rezidivs positiv ausfällt. Nur für die reine postradiogene Veränderung liegt die *Odds Ratio* auch in dem Logit Modell weiterhin unter Eins und führt somit eher zu dem Urteil, dass kein pulmonales Rezidiv vorliegt.

Mit vorliegender Kenntnis über die Ausgangsuntersuchung, liegen die *Odds* für die fokal zentrale und die fokal periphere FDG-Akkumulation beide bei etwa 30. Auch hier wird nach Kenntnis über die Voruntersuchung das Kriterium der peripheren Akkumulation durch die Differenzierung in echte pulmonale Metastasen quasi

„bereinigt“, womit die *Odds* von 44 auf 30 sinkt. Zentrale und periphere FDG-Akkumulationen führen demnach gleich wahrscheinlich zur Wertung pulmonales Rezidiv. Die höchste Eintrittswahrscheinlichkeit für die Bewertung als pulmonales Rezidiv besitzt jedoch die Detektion einer intrapulmonalen Metastase, gefolgt von der Einstufung fibrotischer Veränderungen als malignitäts-suspekt.

5.2.2. Bewertung des Mediastinum

Im Gegensatz zur Bewertung der Lunge, wurde bei der Beurteilung des Mediastinum nur eine rein anatomische Zuordnung abgefragt. Zur besseren Übersicht wurden die Lymphknotenstationen in Tabelle 4.10, gruppiert in oberes, mittleres und unteres Mediastinum dargestellt.

Die Übereinstimmungswerte zwischen den *Observern* liegen für das mittlere und untere Mediastinum im Bereich moderater Übereinstimmung und für das obere Mediastinum sogar im Bereich erheblicher Übereinstimmung. Dabei nimmt die Kenntnis der *Observer* über die Ausgangsuntersuchung keinen Einfluss auf die Übereinstimmungswerte. Diese Tatsache spricht für die PET/CT als relativ eindeutig zu interpretierende Untersuchung in der Diagnostik einer mediastinalen Beteiligung beim NSCLC. Die sehr guten Übereinstimmungswerte unterstützen bereits an dieser Stelle die Aussagen der Leitlinien, die die PET als sehr sensitive Methode für das mediastinale Staging hervorheben. Dabei lässt das IOA selbstverständlich keine Aussage über die Differenzierung eines FDG-aviden Lymphknotens in reaktiv und Rezidiv zu und erlaubt somit auch keine Aussage über die Anzahl falsch-positiver Befunde.

Das hohe IOA für das obere Mediastinum steht etwas im Gegensatz zu den Aussagen in der Literatur, die gerade das obere Mediastinum häufig als Problemzone in der Differenzierung zu supraclaviculären Lymphknoten beschreiben. In diesem Zusammenhang ist hier am ehesten davon auszugehen, dass die hohe IOA auf die genaue Definition der oberen mediastinalen Lymphknotenlevel durch den Atlas nach *Chapet et al.* zurückzuführen ist (67).

Die etwas niedriger gelegenen Übereinstimmungswerte im mittleren und unteren Mediastinum könnten beispielsweise auf schwer zu interpretierende, entzündliche Lymphknoten mit FDG-*Uptake* zurück zu führen sein (42).

5.2.3. Urteil über Fernmetastasen

Die erheblichen Übereinstimmungswerte zwischen den *Observern* in der Diagnostik extrathorakaler Fernmetastasen unterstreichen die in der Literatur beschriebene, hohe Wertigkeit der PET bei der Detektion unerwarteter Fernmetastasen (5, 6). Das es hierbei weniger um die korrekte Interpretation der PET-Information, wie beispielsweise beim Betrachten der Lunge geht, sondern die Schwierigkeit in der rein visuellen Detektion liegt, lässt sich an einem Fallbeispiel der Studie schildern, bei dem im Verlauf ein Rezidiv in Form einer Hirnmetastase auftrat. Diese sind in der PET häufig schwierig vom basalen FDG-*Uptake* des gesunden Hirngewebes zu differenzieren und können dem Betrachter entgehen (46). Das Vorliegen dieser Metastase wurde in der visuellen Auswertung nur von 2 der 4 *Observer* erkannt.

5.2.4. Gesamturteil

Insgesamt liegt der Konsens der *Observer* über das Vorliegen eines Rezidiv im Gesamturteil in 75 von 114 Fällen nahe an der Zahl, der durch den Goldstandard diagnostizierten Rezidive in 78 von 114 Fällen (Vgl. Tab. 4.4 und 4.6). Bis auf *Observer* 3 liegen alle in ihrer Einschätzung im ungefähren Bereich der tatsächlichen Anzahl an Rezidiven. Dies unterstreicht auch das Interobserver-*Agreement* für das Gesamturteil aller *Observer* in Tab. 4.7, welches sich im moderaten Bereich befindet. Nach ausklammern des dritten *Observers* befindet sich das Gesamturteil für die *Observer* 1, 2 und 4 im Bereich erheblicher Übereinstimmung (nicht dargestellt). Auch in den Einzelurteilen über pulmonale und mediastinale Rezidive liegt *Observer* 3 in der Anzahl seiner Rezidivurteile über den anderen *Observern*. Interessanterweise gibt *Observer* 3 die höchste subjektive Beurteilungssicherheit für die Lunge an und weist mit 7% Zunahme der subjektiven Sicherheit nach Kenntnis über die Bestrahlungsplanung eine weniger starke Zunahme, als die ersten beiden *Observer* auf (Tab. 4.12).

5.2.5. Einfluss der Beurteilungssicherheit und der Kenntnisse

Es liegt deshalb nahe, die angegebene Beurteilungssicherheit und den Einfluss der Kenntnisse über das Staging und die Bestrahlungsplanung in Tab. 4.11 und Tab. 4.12 genauer zu betrachten.

Das insgesamt sehr hohe Niveau der subjektiven Beurteilungssicherheit spricht sowohl für eine gut definierte und verständliche Vorgabe der auszuwertenden

Kriterien, als auch für die hohe Erfahrung der Auswerter mit der Methode der FDG-PET/CT. Auch die durchweg in einem guten Bereich angesiedelten Übereinstimmungswerte in der Auswertung unterstreichen das einheitliche Verständnis der Auswertekriterien, wie dem Gesamtergebnis der IOV in der Tab. 4.11 zu entnehmen ist. Die Zahlen der subjektiven Sicherheit liegen für die Beurteilung des Mediastinums deutlich höher als jene bei der Bewertung der Lunge, was sich am ehesten durch das Vorliegen posttherapeutischer, entzündlicher Veränderungen oder residuellen Tumorgewebes erklären lässt. Wie in der Literatur beschrieben, erschwert das Vorliegen dieser Faktoren die Bewertung der Lunge (51). Im Mediastinum ist der Effekt der posttherapeutischen Veränderungen nach Strahlentherapie weniger stark ausgeprägt und könnte einen Unterschied in der Beurteilungssicherheit somit erklären. Für beide Lokalisationen steigt die Beurteilungssicherheit mit zusätzlich vorliegenden Kenntnissen über die Ausgangsuntersuchung und die Bestrahlung erwartungsgemäß an. Dabei ist der beobachtete Mittelwertunterschied in der subjektiven Sicherheit zwischen beiden Szenarien statistisch signifikant. Die Zunahme der Häufigkeit sicherer Bewertungen ist bei der Lunge stärker ausgeprägt als für das Mediastinum. Auch diese Beobachtung erklärt sich am ehesten durch die erschwerte Beurteilung der Lunge durch postradiogene Veränderungen, deren Interpretation durch das Vorliegen weiterer Informationen, wie der Voruntersuchungen vor der Bestrahlung, erleichtert wird. Auch die sichere Einordnung eines vorher unsicheren pulmonalen Rezidiv kann mit Kenntnis über die Ausgangsuntersuchung erleichtert werden und somit zum beobachteten Effekt beitragen. Betrachtet man die Übereinstimmungswerte für die IOV in der Tab. 4.11, so wird auch umgekehrt deutlich, dass die subjektive Beurteilungssicherheit Einfluss auf die Bewertung der *Observer* nimmt. Bei als sicher markierten Urteilen liegt das IOA, unabhängig von der zu bewertenden Lokalisation, durchweg im Bereich erheblicher Übereinstimmungswerte. Wie zu erwarten, sind im Vergleich dazu die Übereinstimmungswerte bei Urteilen, bei denen sich mindestens einer der *Observer* unsicher ist, deutlich niedriger. Sie liegen für die Bewertung der Lungen und im Gesamturteil lediglich im Bereich ausreichender Übereinstimmung. Für die Beurteilung des Mediastinum liegen die Werte sogar nur im Bereich leichter Übereinstimmung. Bei der Interpretation dieses niedrigen Übereinstimmungswertes muss berücksichtigt werden, dass die Anzahl der unsicheren mediastinalen Urteile im Durchschnitt nur etwa 10 – 15 von 114 Fällen betraf und die Fallzahl für die Berechnung der IOV in diesem Zusammenhang sehr niedrig war. Dennoch lassen die schlechten Übereinstimmungswerte in diesem Zusammenhang vermuten, dass es für

die Bewertung des mediastinalen Rezidiv einige wenige, strittige und schwierig zu interpretierende Fälle gibt. Bei diesen könnte beispielsweise die sichere Zuordnung zum Mediastinum in Abgrenzung zu zervikalen, perigastralen oder pulmonalen Lymphknoten eine Rolle spielen. In einem der Fälle, in dem ein Rezidiv in Form einer Lymphknotenmetastase am gastroösophagealen Übergang vorlag, entschieden sich beispielsweise 2 der *Observer* für die Zuordnung zur Lymphknotenstation 8 und somit für ein mediastinales Rezidiv (Abb. 5.1). Die beiden anderen entschieden sich in diesem Fall für das Vorliegen einer abdominalen Fernmetastase und somit gegen ein mediastinales Rezidiv. Alle vier bewerteten ihre Beurteilung des Mediastinums bei diesem Fall jedoch als unsicher. In Fällen wie in diesem genannten Beispiel tragen einheitliche Kriterien zur Lokalisationszuordnung dazu bei, die Interpretation der PET zu vereinfachen.

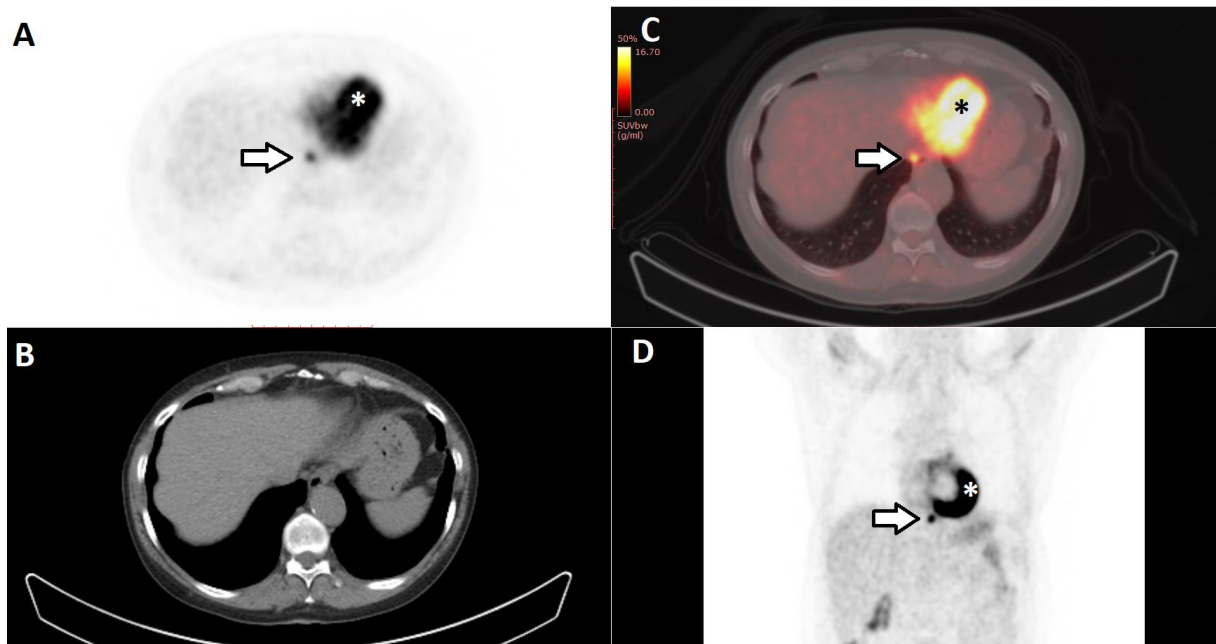


Abbildung 5.1 Beispiel für die unsichere Zuordnung einer Läsion.

Das Bildbeispiel zeigt das schwächungskorrigierte PET (A), die axiale CT-Schicht (B), das resultierende Fusionsbild (C) und die entsprechende coronare Ebene der PET (D) eines Patienten mit einer Lymphknotenmetastase im Bereich des gastroösophagealen Übergangs (Pfeilmarkierung). Erschwert wird die Detektion der Metastase zusätzlich durch das kräftige FDG-Signal des Herzmuskels (Sternmarkierung).

Vergleicht man die Übereinstimmungswerte der IOV in Tab 4.11 für die Beurteilungen ohne und mit Kenntnis über die Ausgangsuntersuchung und die Bestrahlung, so unterscheiden sich diese zwischen den beiden Szenarien nicht: Sie liegen stets in derselben Kategorie der Übereinstimmung und zeigen auch in der t-Testung keinen signifikanten Unterschied zwischen ihren Mittelwerten. So könnte man zunächst (fälschlicherweise) daraus schließen, dass das Vorliegen zusätzlicher Informationen

lediglich die subjektive Sicherheit verbessert, aber gar keinen Effekt auf die Rezidivbewertung nimmt. Dass der Zusammenhang jedoch nicht so einfach zu interpretieren ist, zeigt die nachfolgende Überlegung. In einem Fall, in dem sich alle *Observer* ohne Vorkenntnisse für ein falsch-negatives Urteil entschieden hätten, mit Vorkenntnissen aber dann das richtig-positive Ergebnis favorisierten, würde sich das IOA nicht ändern, da alle immer einstimmig geantwortet hätten. Das Vorliegen von Vorkenntnissen hätte dennoch einen Effekt auf die Beurteilung. Dieser lässt sich jedoch nicht anhand der IOA, sondern mit dem Vergleich der Sensitivität und Spezifität zwischen beiden Szenarien aufzeigen.

5.2.6. Sensitivität und Spezifität anhand des Goldstandards

Die in dieser Arbeit ermittelten und in der Tabelle 4.20 dargestellten Werte für die Sensitivität und Spezifität der visuellen Interpretation der PET/CT liegen in einem, für die klinische Routine, insgesamt guten Bereich. Die Sensitivität liegt, wie zu erwarten, für die Bewertung des Mediastinum über der der Lunge. Dabei decken sich die in dieser Arbeit ermittelten Werte für die Sensitivität und Spezifität für das Mediastinum mit 82,6 % und 84,6 % mit denen, der in der Literatur beschriebenen Werte, welche mit 75 – 85 % für die Sensitivität und 85 – 90 % für die Spezifität angegeben werden (46). Die bereits diskutierten, erschwerenden Faktoren in der Beurteilung der Lunge, wirken sich auch auf die Sensitivität und Spezifität aus.

Die Detektion von Metastasen liegt mit Sensitivitäten von 73,2 %, bzw. 78,6 % etwas hinter den Erwartungen zurück, zeigt jedoch mit bis zu 89,7 % eine hohe Spezifität. Ein möglicher Einflussfaktor auf die, im Vergleich zur Literatur, niedrige Sensitivität in der Detektion von Metastasen, könnte die hier ursprünglich andere Zuordnung der intrapulmonalen Metastasen im Fragebogen sein. Diese wurden von den *Observern* zunächst mit in die Kategorie des pulmonalen Rezidiv gezählt und erst später in der Auswertung zur besseren Vergleichbarkeit mit dem Goldstandard den Metastasen zugerechnet. Die nachträglichen Berechnungen beruhen zwar exakt auf den ausgewerteten Fragebögen, geben aber nicht in jedem Fall eine differenzierte Interpretation der *Observer* zwischen kontralateraler Satellitenmetastase (Metastase) und neuem Rundherd mit lokalem Bezug (pulmonales Rezidiv) wieder.

In der optimalen Befundkonstellation, in der der visuell Beurteilende sämtliche Kenntnisse über den Ausgangsbefund und die Bestrahlungsplanung besitzt, erreicht die qualitative Auswertung durch die *Observer* im Gesamturteil eine Sensitivität von

92,3 % und eine Spezifität von 88,9 %. Die Werte entsprechen den in der Literatur genannten Zahlen von über 90 % für die Sensitivität und 89 % für die Spezifität der PET/CT beim NSCLC (4, 46).

Dass die Werte für die Sensitivität über denen der Spezifität liegen, spricht für eine von Vorsicht geprägte Interpretation der *Observer*, bei denen sich diese im Zweifel lieber falsch-positiv als falsch negativ-entscheiden. Eine Erhöhung der Anzahl der falsch-positiven Bewertungen senkt dabei gleichzeitig die Spezifität, wirkt sich aber nicht auf die Sensitivität der Beurteilung aus, da sich diese aus richtig-positiven und falsch-negativen Urteilen zusammensetzt.

Vergleicht man die, anhand des Goldstandards ermittelten Werte der Sensitivität und Spezifität zwischen den beiden Szenarien, ohne und mit Kenntnissen über die Ausgangsuntersuchung und die Bestrahlung, zeigt sich nun ein Effekt, der zuvor anhand der IOA nicht abzuschätzen war. In den meisten Fällen nehmen sowohl Sensitivität, als auch Spezifität in der Bewertung zu, was die Wichtigkeit über vorliegende klinische Informationen und Voruntersuchungen hinsichtlich der korrekten Befundinterpretation betont. Der Effekt ist, entsprechend *Iravani et al.*, auch in dieser Arbeit am stärksten bei der Bewertung der Lunge zu sehen (51). Hier verbessert sich die Sensitivität der Bewertung von 64,3 % auf 73,8 %, während die Spezifität annähernd gleichbleibt. Im Gesamturteil sorgen die entsprechenden Kenntnisse der *Observer* dafür, dass sich die Sensitivität in der Beurteilung von 84,6 % zu 92,3 % erhöht. Die Spezifität steigt von 86,1 % auf 88,9 % ebenfalls leicht an.

Es ist bemerkenswert, dass dieser Einfluss der Kenntnisse über die Voruntersuchung für die Bewertung des Mediastinums und des lokoregionären Rezidiv insgesamt (pulmonal und mediastinal) scheinbar nicht zu beobachten ist. Hier nimmt lediglich die Spezifität in der Beurteilung zu. Die Tatsache, dass die Sensitivität in beiden Szenarien gleich bleibt, spricht dafür, dass die Beurteilung des Mediastinum ohne Kenntnisse über die Voruntersuchung einfacher fällt, als die Beurteilung der Lunge. In den Fällen, in denen sich die Beurteiler nach Vorliegen der Kenntnisse korrekterweise für das Vorliegen eines pulmonalen Rezidiv entschieden, lag ebenfalls ein lokoregionäres mediastinales Rezidiv vor. Deswegen erhöht sich zwar die Anzahl richtig-positiver Urteile für die Lunge zwischen beiden Szenarien (von 27/42 auf 31/42), die Anzahl richtig-positiver lokoregionärer Rezidive bleibt aber gleich (39/50). Die nun richtig-positiv gewerteten pulmonalen Rezidive fehlen jedoch in der Konsequenz bei den falsch-negativen lokoregionären Befunde und erhöhen somit die Spezifität (von 45/64 auf 47/64).

Interessanterweise liegt das Interobserver-*Agreement* für die Lunge insgesamt über dem des Mediastinums, während Sensitivität und Spezifität umgekehrt für das Mediastinum besser sind, als für die Bewertung der Lunge. Die *Observer* sind sich also in der Bewertung der Lunge untereinander eher einig, liegen damit aber auch häufiger falsch.

Nachdem der Einfluss der Kenntnisse über die Ausgangsuntersuchung untersucht und die Richtung des Effekts klar war, wurde für die weiteren Vergleiche zwischen der qualitativen Auswertung und der quantitativen Auswertung in Kapitel 5.4 nur noch das optimale und klinisch wünschenswerte Szenario mit entsprechenden Kenntnissen berücksichtigt.

5.2.7. Das Interobserver-*Agreement* im Vergleich zur Literatur

Es gibt hinsichtlich der Rezidivbeurteilung in der PET/CT nur wenige Vergleichsstudien. Viele der gängigen Übereinstimmungswerte in der FDG-PET/CT berufen sich vor allem auf Untersuchungen zur Bewertung einer Therapieansprache und beschäftigen sich weniger implizit und meist als Nebeneffekt mit der Rezidivdiagnostik. Die in dieser Arbeit ermittelten, guten Übereinstimmungswerte in der visuellen Auswertung der *Observer*, finden sich so dennoch in den aktuellen Literaturquellen wieder und liegen in einzelnen Aspekten sogar leicht über den gängigen Beschreibungen. Dabei nehmen, wie bereits gezeigt werden konnte, vor allem die Erfahrung des Bewertenden und die vorher einheitlich definierte Interpretationsweise der PET Einfluss auf die Übereinstimmung zwischen den *Observern*. In Untersuchungen mit ähnlicher Methodik, wie durch *Turgeon et al.*, finden sich ähnlich hohe Übereinstimmungswerte für die visuelle Form der Auswertung durch zwei erfahrene Nuklearmediziner (64). Der hier ermittelte Kappa-Wert von $\kappa = 0,95$ für die Rezidivbeurteilung in der PET erscheint für die klinische Realität dennoch etwas zu hoch gegriffen und ist in der Routine des klinischen Alltags und zwischen mehr als zwei *Observern* sicherlich nur schwer zu erzielen. Maßgeblichen Einfluss hierauf dürfte wohl auch die tägliche Zusammenarbeit der *Observer* in derselben Klinik nehmen. So ließen sich die höchsten Übereinstimmungswerte zwischen zwei *Observern* in dieser Arbeit auch zwischen den beiden, in derselben Klinik tätigen Ärzten messen. Diese Werte lagen, wie im Szenario von *Turgeon et al.*, durchgehend im Bereich nahezu perfekter Übereinstimmung (nicht separat dargestellt). Insgesamt betrachtet, lag der Übereinstimmungswert der *Observer* in dieser Arbeit, für das Gesamturteil über ein

Rezidiv bei $\kappa = 0,574$ und deckt sich so mit den von *Fledelius et al.* berichteten Werten (65). Einschränkend hinzugefügt werden, muss an dieser Stelle, dass *Fledelius et al.* den Übereinstimmungswert von $\kappa = 0,6$, zwar für 8 *Observer* berichten, die Komplexität der abgefragten Kriterien zum Vorliegen eines Rezidiv jedoch deutlich einfacher gehalten war, als in dieser Arbeit. Die von *Wahl et al.* berichteten Übereinstimmungswerte für das mediastinale Staging in der PET entsprechen mit $\kappa = 0,65$ auch ungefähr denen für die mediastinale Rezidivbeurteilung in dieser Arbeit (8). Letztere lagen für sichere Beurteilungen bei $\kappa = 0,65$ und insgesamt bei $\kappa = 0,57$. Sehr erfreulich sind vor allem die erheblichen Übereinstimmungswerte für die Beurteilung der Lunge, die in der aktuellen Literatur so nicht zu finden sind und die die Notwendigkeit einheitlich definierter Kriterien zur pulmonalen Rezidivbeurteilung beim NSCLC nach Radiochemotherapie unterstreichen.

An die erheblichen Übereinstimmungswerte des Harmonisierungsprozesses innerhalb der PET-Plan Studie ließ sich mit dieser Arbeit nur bedingt anknüpfen. Gleichzeitig lagen den *Observern* in dieser Arbeit, im Unterschied zur lückenlosen Dokumentation innerhalb der PET-Plan Studie, keine vollständigen klinischen Angaben und Daten der Patienten vor. Auch von der Dauer und dem Zeitpunkt der Bestrahlung hatten die *Observer* in dieser Arbeit keine Kenntnis. Zwischenuntersuchungen in Form einer CT flossen ebenfalls nicht in die Beurteilung dieser Arbeit mit ein. So ist es nicht allzu überraschend, dass der Übereinstimmungswert für ein Gesamtrezidiv in dieser Studie mit $\kappa = 0,574$ unterhalb des, nach dem Harmonisierungsprozess berichteten Wertes von $\kappa = 0,66$ liegt. Könnte man aber die Bewertung der, als unsicher bewerteten Fälle durch ergänzende Informationen verbessern, so käme man auch in dieser Arbeit in den Bereich der als sicher bewerteten Urteile mit $\kappa = 0,658$ (siehe Tab. 4.11). Aus dieser Tatsache lassen sich zwei weitere wichtige Hypothesen generieren. Zum einen ist ein Harmonisierungsprozess zeitlich gebunden und muss immer wieder reevaluiert werden, um den Effekt aufrecht zu erhalten. Zum anderen sorgen auch hier ergänzende klinische Informationen für bessere Übereinstimmungswerte zwischen den *Observern*.

Insgesamt sind die, in dieser Arbeit erzielten Werte des Interobserver-*Agreements* in der Rezidivbeurteilung jedoch als sehr gut einzustufen.

5.3. Die quantitative Auswertung

Wie in Kapitel 2 erläutert, gibt es für die Rezidivbeurteilung in der FDG-PET/CT bislang keine evaluierten, objektivierbaren, quantitativen Kriterien. Einteilungen, wie die PERCIST- oder die PREDIST-Kriterien, dienen primär der Beurteilung einer Therapieansprache und sind nur für kurze Zeiträume von Wochen bis wenigen Monaten nach Therapiebeginn evaluiert (8). Um der visuellen und somit subjektiv geprägten Auswertung durch die Observer dennoch eine objektivierbare Methode gegenüber stellen zu können, wurden die Bilddaten nach PERCIST-Schema ausgewertet und zur Rezidivbeurteilung anhand der PREDIST-Kriterien als komplette metabolische Remission (*CMR*) oder Rezidiv eingestuft.

5.3.1. Die Ergebnisse der PERCIST- und der PREDIST-Kriterien

Bei einem Vergleich der Ergebnisse nach der PERCIST-Auswertung in Tabelle 4.15 mit den Ergebnissen des Goldstandards (Tab. 4.4) wird schnell deutlich, dass die Einteilung in vier Kategorien, mit den beiden Möglichkeiten *PMR* und *SMD*, die Anzahl absoluter Rezidive unterschätzt. Nach PERCIST-Klassifikation liegen in der Gesamtauswertung der 114 Fälle im Verlauf lediglich 46 Fälle mit einem Progress bzw. einem Rezidiv (*PMD*) vor. Im Vergleich dazu gibt der Goldstandard ein Gesamtrezidiv bei 78 von 114 Fällen vor. Hauptgrund hierfür ist sicherlich, dass ein sinkendes FDG-*Uptake* des Primärtumors zum Zeitpunkt unmittelbar nach Bestrahlung noch zweifelsfrei für eine teilweise metabolische Remission und einen Rückgang der Erkrankung steht. Zeigt der Ort der Primärtumorlokalisation jedoch nach beispielsweise 3 Jahren wieder einen erneuten FDG-*Uptake*, so ist dies vermutlich auf ein Lokalrezidiv zurückzuführen. Liegt der FDG-*Uptake* in dieser Situation allerdings noch unterhalb des Ausgangsniveaus aus dem Staging und fand zum Zeitpunkt der kompletten metabolischen Remission zwischenzeitlich keine Bildgebung statt, so würde der Befund fälschlicherweise in die Kategorie *PMR* eingeordnet werden, obwohl eigentlich ein metabolisches Rezidiv vorläge. In der Mehrheit der Fälle in dieser Arbeit liegt ein Zeitraum von über 6 Monaten nach Therapiebeginn vor.

Wertet man unabhängig vom Zeitraum neben der Kategorie *PMD* auch die Kategorien *SMD* und *PMR* als Rezidiv, so steigt die Anzahl der Rezidivbewertungen auf insgesamt 78 von 114 Fällen an (Tab. 4.16). Dies entspricht formal zwar der Anzahl der anhand des Goldstandards diagnostizierten Rezidive, stimmt jedoch nicht mit jenen überein. Die Werte für die Sensitivität (82 %) und Spezifität (75 %) liegen deutlich hinter jenen

der PREDIST-Kriterien zurück (nicht gesondert dargestellt). Auch ein Vergleich der Übereinstimmungswerte zwischen dem Konsens der *Observer*-Auswertung und dem PERCIST-Kriterium PMD als Rezidiv in Tabelle 4.18 macht deutlich, dass die Kategorisierung nach PERCIST ungeeignet für die Rezidivbeurteilung erscheint. Die Werte liegen für die pulmonale und mediastinale Rezidivbeurteilung lediglich im Bereich leichter Übereinstimmung. Auch das Gesamturteil liegt lediglich im unteren Bereich der moderaten Übereinstimmung. Vergleicht man das IOA für die beiden in der Tabelle aufgelisteten Zeiträume, so würde man zunächst annehmen, dass die Übereinstimmungswerte im ersten halben Jahr höher sein müssten, als in dem Zeitraum danach, da die PERCIST-Kriterien ja gewöhnlich in diesem Zeitraum noch Anwendung finden. Dass die Übereinstimmungswerte hier aber noch stärker abweichen, als zu den späteren Zeitpunkten, macht deutlich, dass die quantitative Beurteilung der Therapieansprache nicht mit einer qualitativen Rezidivbeurteilung vergleichbar sind. Dies beweist auch die Beurteilung der Fernmetastasen, deren Neuauftreten, unabhängig vom Lokalbefund, auch in PERCIST immer als Rezidiv gewertet werden. Die Werte für eine neu aufgetretene Fernmetastasierung, die sich in beiden Fällen ja rein auf ein Rezidiv fokussieren, liegen durchweg im Bereich erheblicher Übereinstimmung.

In Tabelle 4.18 wurden zusätzlich Übereinstimmungswerte für die Szenarien ermittelt, in denen PMD und SMD oder PMD, SMD und PMR als Rezidiv gewertet wurden. Die Übereinstimmungswerte für die Bewertung eines Gesamtrezidiv verbessern sich zwar, jedoch nur für die Bewertungen im Zeitraum von mehr als 6 Monaten nach Therapiebeginn.

Da also die bloße Umsetzung der Kategorie *PMD* als Rezidiv bereits in der ersten Phase der Evaluation scheiterte, war die naheliegendste Überlegung, die Kategorien *SMD* und *PMR* ebenfalls als Rezidiv zu werten. Dies hätte die Anzahl der Rezidive für den anfänglichen Zeitraum wiederum überschätzt, da in dieser Phase nach Bestrahlung noch Tumorresiduen vorliegen können, die FDG akkumulieren (56). So hätte man überlegen können, die Wertung der Kategorien *SMD* und *PMR* zeitabhängig zu gestalten. Allerdings hätte man auch hierfür zunächst gültige Zeitspannen evaluieren müssen, da weder *Wahl et al.* noch andere Literaturquellen diese definieren (8, 35). Die andererseits notwendige Einzelfallprüfung für jede Bildgebung widerspräche dem Gedanken einer automatischen quantitativen Auswertung.

So erschien es sinnvoller den Evaluierungsprozess und die Kriterienvorgabe möglichst simpel zu halten und die Einteilung anhand der binären Merkmalsausprägungen der PREDIST-Kriterien vorzunehmen, die sich im Gegensatz zu den PERCIST-Kriterien auch einer anderen Schwellenwertdefinition im *Follow Up* bedienen (Siehe Kapitel 5.3.3). Die Tabelle 4.17 gibt die Ergebnisse der Auswertung nach PREDIST-Schema wieder und zeigt die nun im Bereich des Goldstandards liegende Anzahl an Rezidivurteilen, wobei auch die Anzahl kompletter metabolischer Remissionen, durch die veränderte Schwellenwertdefinition etwas höher liegt als in der Auswertung nach PERCIST. Die Anzahl der Metastasen erhöht sich ebenfalls, da beschlossen wurde, für die PREDIST-Kriterien nur noch zwischen lokoregionärem Rezidiv und Metastase zu unterscheiden. Im Vergleich mit den absoluten Zahlen des Goldstandard (Tab. 4.4) liegt auch die Zuordnung des lokoregionären Progress (in pulmonal und/ oder mediastinal) in der gleichen Größenordnung.

Der Vergleich des IOA mit der qualitativen Auswertung in Tabelle 4.19 offenbart hervorragende Übereinstimmungswerte im erheblichen Bereich für die beiden Formen der Auswertung. Wie zu erwarten fällt die Zuordnung eines pulmonalen Rezidiv etwas weniger gut aus. Der Wert liegt im Bereich moderater, also noch immer guter, Übereinstimmung. Hierbei muss jedoch angemerkt werden, dass auch in dieser semi-automatischen Form der Auswertung nach den PREDIST-Kriterien, offensichtlich entzündlich bedingte Veränderungen, wie beispielsweise ein homogener FDG-*Uptake* der gesamten Pleura, als pulmonales Rezidiv abgelehnt wurde.

Im Vergleich zwischen dem frühen und dem späteren Zeitpunkt des Rezidivverdachts, zeigt sich nun eine bessere Übereinstimmung für den späteren Zeitraum, was sich dadurch erklären ließe, dass zu einem früheren Zeitraum noch Tumorresiduen und entzündliche Veränderungen nach Bestrahlung die Beurteilung erschweren und so einer visuellen Interpretation bedürfen. Diese, die Bewertung erschwerenden Veränderungen liegen zum späteren Zeitpunkt nicht mehr vor. Gestützt wird diese Theorie durch den schlechtesten Übereinstimmungswert zwischen beiden Formen der Auswertung, der bei der Bewertung eines pulmonalen Rezidiv innerhalb der ersten sechs Monate nach Therapie vorliegt. Bemerkenswert ist jedoch die im fast perfekten Übereinstimmungsbereich liegende Gesamtbeurteilung im Zeitraum von über sechs Monaten nach Therapiebeginn, die eine weitere Überprüfung der Sensitivität und Spezifität der PREDIST-Kriterien in diesem Zusammenhang unterstützt.

5.3.2. Sensitivität und Spezifität im Vergleich zum Goldstandard

Bei der Überprüfung der Testgütekriterien für die PREDIST-Auswertung bestehen diese vor allen Dingen durch ihre hohe Spezifität bezüglich der korrekten Diagnose (Tab. 4.21). In der Gesamtbeurteilung auf das Vorliegen eines Rezidiv liegt innerhalb der PREDIST-Auswertung lediglich ein Fall eines falsch-positiven Urteils vor. Der Testung kommt hier sicherlich die Vorselektion des Patientenkollektivs durch die vorher stattgefundenene CT-Bildgebung zugute, die Fälle mit bereits sicherer Remissionen für diese Arbeit ausschließt. Die Spezifität liegt sowohl für die Beurteilung eines lokoregionären Rezidiv, als auch für Metastasen in einem sehr guten Bereich von über 90 %. Die Sensitivität ist, wie zu erwarten, für die Beurteilung der Lunge am niedrigsten und liegt hier bei dennoch akzeptablen 76,2 %. Die Sensitivität von 85,9 % für die Detektion eines Gesamtrezidivs ist für ein automatisiertes und Schwellenwert-basiertes Verfahren zunächst einmal als zufriedenstellend einzuordnen. Sie macht aber deutlich, dass die Anzahl falsch-negativer Ergebnisse für einen Einsatz dieses Verfahrens als alleiniges Diagnostikum noch zu hoch liegt.

Die hohen Werte in der Spezifität der Auswertung könnte man beispielsweise als zweiten, bestätigenden diagnostischen Test nutzen. So ist es in der klinischen Testung häufig üblich, in einem ersten Schritt zunächst ein Verfahren mit hoher Sensitivität anzuwenden, um möglichst viele richtig-negative Ergebnisse zu identifizieren. Gleichzeitig bekommt man mit dieser Testung eine gewisse Anzahl an falsch-positiven Befunden. In einem zweiten Schritt benötigt man zur Bestätigung der Diagnose ein Verfahren mit hoher Spezifität, um die richtig-positiven Ergebnissen unter den getesteten herauszufiltern. Eine konkretes Anwendungsbeispiel für den Einsatz der PREDIST-Kriterien in der Rezidivdiagnostik wäre die Unterstützung von noch unerfahrenen Assistenzärzten in der Weiterbildung.

5.3.3. Die Bedeutung der Schwellenwert-Definition

Bei der Definition eines bestimmten Schwellenwertes, ab dem eine Läsion als Rezidiv gelten soll, muss man die individuell unterschiedliche Stoffwechselsituation jedes Untersuchten berücksichtigen. Zwar werden bei der Akquisition der PET-Bilddaten bereits einheitliche Vorkehrungen, bezüglich der Nüchternheit des Patienten und dem Zeitpunkt der FDG-Injektion getroffen. Dennoch gibt es bei jedem Patienten kleine, individuelle Unterschiede in der Aktivität des Glukosemetabolismus. Die PERCIST-Kriterien orientieren sich deshalb am Basisstoffwechsel der Leber und ziehen diesen heran, um einen möglichst individuellen Schwellenwert definieren zu können. Die

Formel zur Berechnung der Schwelle ist wiederum für jeden Patienten gleich. Diese Schwellenwertdefinition gilt bei den PERCIST-Kriterien aber lediglich in der Ausgangsuntersuchung. In der Kontrolle der Therapieansprache gibt es keine harte Schwellenwertdefinition mehr, sondern hier werden die höchsten SUL_{peak} -Werte des Tumors aus beiden Untersuchungen verglichen. Damit beide Werte aber miteinander verglichen werden dürfen, fordern die PERCIST-Kriterien zusätzlich, dass der SUL_{mean} der Leber zu beiden Zeitpunkten sich nicht um mehr als 0,2 oder insgesamt 20 % voneinander unterscheiden darf. Der Einsatz verschiedener Geräte in multizentrischen Studien erschwert dies häufig.

Es ergibt also Sinn, abweichend von den PERCIST-Kriterien, eine Schwellenwertdefinition zu wählen, für deren Berechnung man lediglich die vorliegende Untersuchung benötigt – wie bei den PREDIST-Kriterien. Den Einsatz, der in der *Baseline* genutzten Schwellenwertdefinition durch den FDG-*Uptake* der Leber, zur Diagnostik im *Follow Up* hatten bereits *Wahl et al.* in der Originalpublikation der PERCIST-Kriterien diskutiert und als mögliche Alternative vorgeschlagen (8). Erst *Maffione et al.* hatten diese Definition in die PREDIST-Kriterien implementiert und das Verhalten des Tumors in der FDG-PET/CT mit der Tumorphistologie korreliert (43). Das Ergebnis der Korrelation sprach dafür, dass diese Schwellenwertdefinition sich hervorragend zur Kontrolle solider Tumoren nach Radiochemotherapie eignete. Daher schien es naheliegend, dieselbe Schwellenwertdefinition auch für diese Arbeit zu wählen. Die hohen Übereinstimmungswerte mit den *Observern* und die Spezifität der Schwelle in dieser Arbeit festigten zunächst die Entscheidung einer individuellen, am Leberstoffwechsel orientierten Schwelle im *Follow Up*. Um dies weiter zu überprüfen wurden Schwellenwerte für den SUL_{peak} und den SUV_{max} im *Follow Up* definiert, ab denen ein Rezidiv vorliegen sollte (Tab. 4.26). Für diese Werte wurden dann die Sensitivität und Spezifität anhand der *Observer*-Meinung und des Goldstandards bestimmt. Bei einem niedrigen Schwellenwert weisen beide Parameter eine Sensitivität um die 90 % auf, besitzen jedoch eine schlechte Spezifität. Hebt man den Schwellenwert an, so steigt auch die Spezifität. Im Gegenzug sinkt jedoch die Sensitivität der Parameter. Insgesamt findet sich keine Schwellenwertdefinition, welche dieselbe Kombination aus guter Sensitivität und hervorragender Spezifität wie die PREDIST-Einteilung aufweist.

5.4. Vergleich zwischen quantitativer und qualitativer Auswertung

Die Testgütekriterien befinden sich unter den hier gegebenen Voraussetzungen, mit lediglich eingeschränkten klinischen Patienteninformationen, für beide Verfahren in einem guten Bereich. Dies pflichtet der Aussage von *Turgeon et al.* bei, dass beide Verfahren zur Beurteilung der FDG-PET/CT eine vergleichbare Aussagekraft besitzen (64). Bei genauerer Differenzierung besitzt jedes Verfahren aber seine individuellen Vor- und Nachteile. So weist die visuelle, qualitative Form der Auswertung durch die Observer eine wesentlich höhere Sensitivität in der Rezidivbeurteilung auf. Ein Vorteil, der von *Iravani et al.* für die Kontrolle der Therapieansprache bereits ähnlich beschrieben wird (51). Die höhere Sensitivität der *Observer*-Auswertung im Vergleich zur quantitativen Auswertung spricht dafür, dass hier eine differenzierte Entscheidung in Abgrenzung zu nicht malignen Veränderungen notwendig ist und diese auch weiterhin ganz klar die Domäne des erfahrenen Nuklearmediziners darstellt. Der Unterschied zwischen visueller und automatisierter Auswertung wird besonders in den Fällen innerhalb der ersten 6 Monate nach Therapiebeginn deutlich. In diesem Zeitraum, in dem vor allem entzündliche Veränderungen nach Radiatio die Beurteilung der PET erschweren, erfordert es die Interpretation der PET durch einen erfahrenen Befunder. Hingegen gelingt eine Differenzierung zwischen entzündlichen und malignen Prozessen durch eine reine Quantifizierung der SUV-Werte nicht, wie die Übereinstimmungswerte in den Tabellen 4.18 und 4.19 zeigen.

Die differenziertere Betrachtung zur Zuordnung der Lokalisation des Rezidivs ist nur sehr eingeschränkt beurteilbar und kann nur vorsichtig interpretiert werden, da die Unterscheidung in Lokal- und Fernrezidiv in den Fragebögen nicht gemacht wurde. Die Zuordnung der pulmonalen Metastase als pulmonales Rezidiv wurde zwar für die Berechnung der Sensitivität und Spezifität berücksichtigt, dennoch liegt hier eine potentielle Fehlerquelle vor, die die niedrigere Sensitivität der *Observer* im Vergleich zu PREDIST in diesem Aspekt erklären dürfte.

Erstaunlich ist, dass die Sensitivität beider Verfahren für die Beurteilung des Mediastinums und das Urteil über Metastasen gleichwertig ist. Die jeweils etwas höhere Spezifität der PREDIST-Kriterien ist sehr vielversprechend, hier muss allerdings ebenfalls berücksichtigt werden, dass die retrospektive Zuordnung der intrapulmonalen Metastase zu den extrathorakalen Metastasen zu potentiell falsch niedrigen Werten der *Observer* führen könnte.

Am verlässlichsten und von der Umstrukturierung unberührt, bleiben die Werte über das Gesamturteil. Die hier höhere Sensitivität der Observer bestätigt den Vorteil einer dezidierten Beurteilung durch einen erfahrenen Nuklearmediziner. Die Unterschiede zwischen den Szenarien ohne und mit Kenntnis über die Bestrahlungsplanung bestätigen darüber hinaus, dass zusätzliche klinische Informationen das Urteil des Nuklearmediziners verbessern. Die Kombination der visuellen Beurteilung mit dem quantitativen Verfahren könnte die Spezifität der Auswertung zusätzlich erhöhen und die Beurteilung durch den Nuklearmediziner damit verbessern, ähnlich der Feststellung von *Fledelius et al.* (65).

5.5. Der Zeitpunkt der Rezidivdiagnostik

Der Zeitpunkt der Rezidivdiagnostik ist von entscheidender Wichtigkeit für die Anwendung der richtigen Kriterien zur Rezidivdiagnostik. So wurde bereits erläutert, dass es in den ersten Monaten eine breite Überlappung zwischen der Beurteilung der Therapieansprache und der Diagnostik eines Frührezidiv gibt. Außerdem sorgen Tumorrestgewebe und Therapie-assoziierte entzündliche Prozesse für eine erschwerte Beurteilbarkeit der FDG-PET/CT im Vergleich zu einem späteren Zeitpunkt. Da der zeitliche Rahmen in der Literatur mit etwa drei bis sechs Monaten beziffert wird, wurde für den zeitlichen Vergleich der Testgütekriterien beider Auswertungen in Tabelle 4.22 wieder ein halbes Jahr ausgewählt (42, 51, 53). Nach Betrachtung des IOA zwischen der *Observer*-Meinung und den PREDIST-Kriterien in Tabelle 4.19 würde man erwarten, dass die Beurteilung zum späteren Zeitpunkt einfacher ist und somit auch Sensitivität und Spezifität beider Verfahren ansteigen. Dies trifft in der Gesamtbeurteilung auch im Groben zu. Beachtenswert ist jedoch, dass die Beurteilung eines pulmonalen Rezidiv durch die *Observer* zu einem frühen Zeitpunkt mit 6/6 Fällen sehr viel sensitiver ist, als zu einem späteren Zeitpunkt. Gerade in diesem Zeitraum ist die visuelle Auswertung der PET einer (semi-)automatischen Auswertung deutlich überlegen. Bemängeln lässt sich jedoch die sehr niedrige Spezifität des *Observer*-Urteils mit lediglich 33 %, was für eine zu vorsichtige Interpretation der Befunde spricht, bei dem die Bewerter im Zweifel lieber falsch-positiv entscheiden. Zu sehen ist das auch in der statistisch sehr niedrigen Beurteilungssicherheit der *Observer* bei den Urteilen über ein pulmonales Rezidiv im ersten halben Jahr (nicht dargestellt).

5. Diskussion. Abschätzung der Prognose anhand der PET-Parameter

Wie bei der Bewertung des pulmonalen Rezidiv, ist die Sensitivität für die Bewertung des lokoregionären Rezidiv und für die Gesamtbeurteilung innerhalb des ersten halben Jahres nach Therapiebeginn beim Observer-Urteil höher, als für die PREDIST-Kriterien. Die intelligente und korrekte Interpretationsfähigkeit der qualitativen Auswertung, stellt also einen Vorteil bei der Beurteilung eines frühen Rezidivverdachts dar. Das im Verhältnis hierzu relativ schlechtere Abschneiden der PREDIST-Kriterien überrascht etwas, da die Kriterien ja ursprünglich für diesen früheren Zeitraum konzipiert waren. Einschränkend muss man allerdings hervorheben, dass die Fallzahl im ersten halben Jahr relativ niedrig ist und statistisch sichere Aussagen hierdurch erschwert sind.

Die Spezifität des Observer-Urteils in der Gesamtbewertung liegt für den Zeitraum von mehr als sechs Monaten bei 93,8 % und verschlechtert sich für den gesamten Zeitraum maßgeblich durch die niedrige Spezifität für die Frührezidive.

Für die Auswertung nach den PREDIST-Kriterien verbessern sich sowohl Sensitivität als auch Spezifität im späteren Zeitraum. Die Spezifität für die Beurteilung nach über sechs Monaten liegt für alle Bewertung bei 95 % oder höher und erreicht im Gesamturteil sogar einen Wert von 100 % mit 32/32 richtig-positiven und keinen falsch-negativen Urteilen.

Daraus lässt sich schließen, dass die PREDIST-Kriterien durchaus ein hohes Potential zur Anwendung in der Rezidivbeurteilung besitzen und in klinischen Studien mit erhoben und weiter evaluiert werden sollten. Die hohe Sensitivität der visuellen Beurteilung anhand der visuellen Auswertung lässt sich hierdurch jedoch nicht einfach ersetzen.

5.6. Abschätzung der Prognose anhand der PET-Parameter

Wie in der PET-Plan Studie und auch in anderen Studien gezeigt werden konnte, steht das Gesamttumorvolumen zum Zeitpunkt der Bestrahlungsplanung in linearem Zusammenhang mit der Eintrittswahrscheinlichkeit eines lokoregionären Rezidiv beim lokal fortgeschrittenen NSCLC nach Radiochemotherapie (22, 70). Um zu überprüfen, ob dieser Zusammenhang auch für die PET-Parameter SUL_{peak} , SUV_{max} , MTV und TLG zum Zeitpunkt der Bestrahlungsplanung besteht, wurde eine Cox-Regressionsanalyse durchgeführt, bei der die Diagnose des lokoregionären Rezidiv im Goldstandard als eintretendes Ereignis galt (Tab. 4.23).

5. Diskussion. Abschätzung der Prognose anhand der PET-Parameter

Dabei sind die Ergebnisse für die Parameter SUL_{peak} und SUV_{max} in diesem Zusammenhang nicht verwertbar, da die jeweiligen Konfidenzintervalle die Zahl Eins einschließen. Die Ergebnisse für die Parameter MTV und TLG sind hingegen verwertbar und statistisch signifikant. Ihre *Odds* liegen jedoch so nahe an der Zahl 1, dass sie für keinen größeren Effekt sprechen. Die *Odds* für das MTV liegt bei 0,987, was bedeuten würde, dass sich die Chance ein lokoregionäres Rezidiv zu erleiden mit linearer Zunahme des Tumolvolumens sinken würde. Eine Erklärung für diese Beobachtung könnte sein, dass das Tumolvolumen im Arm B der PET-Plan Studie trotz Randomisierung etwas größer war, als in Arm A. Diese These würde durch die minimale Chancenerhöhung für den Parameter TLG (*Odds* = 1,002), welcher als virtuelles Maß für die tatsächliche Tumorzelllast und -aggressivität steht, unterstützt werden. Die Effektausprägung ist letztendlich aber zu klein und unsicher, um ihr diese Bedeutung sicher beimessen zu können. Zudem gilt es auch in diesem Zusammenhang zu bedenken, dass in dieser Arbeit Patienten, die sicher kein Rezidiv erlitten, vorab ausgeschlossen wurden.

Um dennoch einschätzen zu können, ob ein Zusammenhang zwischen PET-Parameter und dem Auftreten eines Lokalrezidiv besteht, der aber im vorselektionierten Patientenkollektiv dieser Arbeit nicht in einem linearen Zusammenhang steht, wurden die Parameter anhand ihres Medians geteilt und beide Gruppen in einem Kaplan-Meier-Schätzer miteinander verglichen (Abb. 4.24). Die Kurven für die Parameter SUV_{max} , MTV und TLG weisen keinen signifikanten Unterschied auf. Ein signifikanter Unterschied in der Log Rank Testung ergibt sich jedoch zwischen beiden Gruppen für den Parameter SUL_{peak} . Demnach erleiden die Patienten in der Gruppe mit den höheren SUL_{peak} -Werten im Baseline-PET häufiger ein Lokalrezidiv, als die Patienten der Gruppe mit den niedrigen SUL_{peak} -Werten (Abb. 4.24, A). Wie bereits erwähnt, kann dieses Ergebnis aus dem vorliegenden, vorselektionierten Patientenkollektiv nicht auf die Gesamtheit aller behandelten Patienten mit lokal fortgeschrittenem Lungenkarzinom übertragen werden. Es sollte aber Anlass geben, eine weitere Untersuchung dieser Parameter im Gesamtkollektiv der Originalstudie durchzuführen. Denn wie in einigen Studien bereits gezeigt werden konnte, könnten diese PET-Parameter beim NSCLC als prognostische Marker für den Verlauf der Erkrankung nach Radiochemotherapie dienen (28, 29).

5.7. Limitationen der Studie

Wenngleich sich die retrospektive Auswertung dieser Arbeit einem Teilkollektiv einer großen multizentrischen Studie bedient, liegen auch in dieser Arbeit einige Limitationen vor.

5.7.1. Datenerhebung

In der Dokumentation der Patienten sind keine Angaben zur Tumorgenetik erfasst. Der Einfluss einer möglichen Immuntherapie wird in dieser Arbeit nicht gesondert erfasst und der Unterschied zur konventionellen Chemotherapie nicht untersucht.

In der klinischen Dokumentation zum Outcome der Patienten wurde nicht sicher zwischen Fernmetastase und Zweitmalignom differenziert.

Die bereits erwähnte Vorselektion der Patienten durch die CT-Bildgebung beeinflusste wesentlich die Selektion der Patienten für diese Arbeit. Wurde ein Rezidiv in der CT-Bildgebung fälschlicherweise sicher ausgeschlossen, erhielt der Patient konsekutiv auch keine PET-Bildgebung und wurde nicht in diese Arbeit eingeschlossen.

5.7.2. Qualitative Auswertung

Eine mögliche Limitation in der Aussagekraft der Arbeit stellt die Erfassung der pulmonalen Metastase im Fragebogen der Auswerter als Teil des pulmonalen Rezidiv dar.

5.7.3. Quantitative Auswertung

Der Einsatz der PERCIST-Kriterien und der PREDIST-Kriterien bleibt bislang größtenteils Studien vorbehalten und dient zur Evaluation der Therapieansprache nach Radiochemotherapie. Ein Einsatz der Kriterien in der Rezidivdiagnostik wurde, soweit dem Autor bekannt, bislang in keiner Studie getestet.

5.7.4. Goldstandard

Die Erhebung des Goldstandards der PET-Plan Studie, der hier ebenfalls zum Vergleich dient, erfolgte zum Teil anhand der vorliegenden und erneut ausgewerteten PET-Bilddatensätze.

6. Zusammenfassung

In Zeiten stetig besser werdender, kombinierter und gezielter Therapiestrategien beim lokal fortgeschrittenen Lungenkarzinom, steigen auch die Ansprüche an die Bildgebung in der Rezidivdiagnostik. Die PET/CT stellt dabei mit ihrer Kombination aus anatomischer und metabolischer Bildgebung das wichtigste Verfahren zur Beurteilung und Differenzierung zwischen therapie-assoziierten Veränderungen und einem Rezidiv dar. Hierfür ist es zwingend notwendig, standardisierte, verlässliche und gut reproduzierbare Beurteilungskriterien zu definieren, die ein Rezidiv sicher erkennen oder ausschließen können. Dabei stehen in der Anwendung sowohl qualitativ visuelle, als auch quantitativ (semi-) automatische Auswertestrategien zur Auswahl:

In der visuellen Auswertung der *Observer* nehmen die Kenntnisse über die Ausgangsuntersuchung und die Therapie zwar keinen Einfluss auf die Übereinstimmungswerte zwischen den Beurteilenden. Sie erhöhen aber die subjektive Sicherheit des gefällten Urteils über das Vorliegen eines Rezidiv und erhöhen die Sensitivität und Spezifität der Beurteilung der PET/CT bezüglich der korrekten Diagnose. Die visuelle Rezidivbeurteilung der Lunge in der PET/CT erweist sich als schwieriger, als die des Mediastinum. Die Werte für Sensitivität und Spezifität, sowie die subjektive Beurteilungssicherheit liegen unterhalb derer des Mediastinum. Die höheren Übereinstimmungswerte in der Rezidivbeurteilung der Lunge, im Vergleich zum Mediastinum, bestätigen jedoch die hohe Wertigkeit einheitlicher Auswertekriterien, die für die Bewertung der Lunge in dieser Arbeit klarer definiert waren als für das Mediastinum. Hieraus ist abzuleiten, dass es für eine visuelle Beurteilung, auch durch erfahrene Auswerter, Sinn ergibt einheitliche Studienkriterien zu definieren und diese in einem Harmonisierungsprozess durch mehrmalige Reevaluation anzupassen und zu verbessern.

Die Definition eines Rezidiv, rein durch einen quantitativen, automatischen Auswertalgorithmus, liegt in seiner Sensitivität hinter der visuellen Auswertung durch einen erfahrenen Auswerter zurück. Mit seiner hohen Spezifität könnte er jedoch eine wertvolle Stütze in der Auswertung der PET durch unerfahrene Ärzte oder bei unsicheren Beurteilungen darstellen. Außerdem könnte die quantitative Form der Auswertung als zweiter Schritt zur Bestätigung nach der visuellen Auswertung eines erfahrenen Arztes dienen. So ergäbe sich für die kombinierte Testung in zwei Schritten eine optimale Kombination aus hoher Sensitivität und Spezifität. Die zweite Stufe zur

Bestätigung könnte dabei Software-gestützt und automatisch parallel zur zeitaufwändigeren, visuellen Auswertung ablaufen und müsste vom befundenen Arzt am Ende nur noch gegengeprüft werden.

Die visuelle Auswertung und Patienten-individuelle Befundinterpretation durch einen erfahrenen Nuklearmediziner stellt weiterhin den Goldstandard der FDG-PET/CT in der Rezidivbeurteilung dar. Sie ist durch ihre hohe Sensitivität der quantitativen Analyse überlegen - gerade innerhalb der ersten 6 Monate nach Radiatio, in denen therapie-assoziierte, entzündliche Veränderungen die Auswertung und Unterscheidung zu einem Rezidiv erschweren.

Für die quantitative, semi-automatisierte Form der Rezidivbeurteilung in der FDG-PET/CT sind in der Literatur bislang weder Kriterien evaluiert noch definiert. Die eigentlich zur Therapieansprache beim lokal fortgeschrittenen Rektumkarzinom entwickelten PREDIST-Kriterien stellen aber einen vielversprechenden Ansatz in der Lösung dieses Problems dar. In dieser Arbeit zeigen sie eine akzeptable Sensitivität und eine sehr hohe Spezifität in der Diagnose eines Rezidiv nach Radiochemotherapie. Die gilt vor allem für den Zeitraum von mehr als 6 Monaten nach Therapie. Im Vergleich zu herkömmlichen Parametern wie dem SUV_{max} zeigen sie eine günstigere Konstellation an Sensitivität und Spezifität in der Bewertung eines Rezidiv. Sie sollten an einem größeren und nicht vorselektionierten Patientenkollektiv weiter evaluiert werden.

Den metabolischen PET-Parametern kommt im Staging, in der Abschätzung der Prognose und der Beurteilung der Therapieansprache eine zunehmend wichtigere Rolle zu, da sie den Effekt neuer zytostatischer Therapiestrategien (Angiogenese- und Wachstumsinhibition, Immunmodulation) besser widerspiegeln als die konventionelle, rein morphologische Bildgebung. In diesem Zusammenhang könnten vor allem die Parameter SUL_{peak} und TLG wichtige Prognoseparameter zur Abschätzung der Wahrscheinlichkeit eines Lokalrezidivs darstellen und in Zukunft vielleicht sogar über Dauer der Radiotherapie und Höhe der lokalen Dosis entscheiden. Sie sollten in Zukunft in großen, multizentrischen Studien mit erhoben und ihre Wertigkeit weiter überprüft werden.

7. Literaturverzeichnis

1. GLOBOCAN 2018: Estimated Cancer Incidence, Mortality and Prevalence 2018. WHO, International Agency for Research on Cancer (Abgerufen am 23.04.2020). 2018.
2. DESTATIS, RKI: Sterbefälle (absolut, Sterbeziffer, Ränge, Anteile) für die 10/20/50/100 häufigsten Todesursachen (ab 1998). Gliederungsmerkmale: Jahre, Region, Alter, Geschlecht, ICD-10. Gesundheitsberichterstattung des Bundes (Abgerufen am 23.04.2020).
3. Griesinger F, Eberhardt W, Früh M, Gautschi O, Hilbe W, Hoffmann H, et al. Leitlinie Lungenkarzinom, nicht-kleinzellig (NSCLC). Deutsche Gesellschaft für Hämatologie und Medizinische Onkologie (DGHO). 2019.
4. Leitlinienprogramm Onkologie (Deutsche Krebsgesellschaft, Deutsche Krebshilfe, AWMF): Prävention, Diagnostik, Therapie und Nachsorge des Lungenkarzinoms, Langversion 1.0, 2018, AWMF-Registernummer: 020/007OL.
5. Vansteenkiste JF, Stroobants SG, De Leyn PR, Dupont PJ, Bogaert J, Maes A, et al. Lymph node staging in non-small-cell lung cancer with FDG-PET scan: a prospective study on 690 lymph node stations from 68 patients. *J Clin Oncol.* 1998;16(6):2142-9.
6. Dwamena BA, Sonnad SS, Angobaldo JO, Wahl RL. Metastases from non-small cell lung cancer: mediastinal staging in the 1990s--meta-analytic comparison of PET and CT. *Radiology.* 1999;213(2):530-6.
7. Mac Manus MP, Hicks RJ, Matthews JP, McKenzie A, Rischin D, Salminen EK, et al. 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.* 2003;21(7):1285-92.
8. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. *J Nucl Med.* 2009;50 Suppl 1:122S-50S.
9. Pinker K, Riedl C, Weber WA. Evaluating tumor response with FDG PET: updates on PERCIST, comparison with EORTC criteria and clues to future developments. *Eur J Nucl Med Mol Imaging.* 2017;44(Suppl 1):55-66.
10. Lardinois D, Weder W, Hany TF, Kamel EM, Korom S, Seifert B, et al. Staging of non-small-cell lung cancer with integrated positron-emission tomography and computed tomography. *N Engl J Med.* 2003;348(25):2500-7.
11. Wittekind C. TNM-Klassifikation maligner Tumoren. 8 ed: Wiley; 2020.
12. Rami-Porta R, Bolejack V, Crowley J, Ball D, Kim J, Lyons G, et al. The IASLC Lung Cancer Staging Project: Proposals for the Revisions of the T Descriptors in the Forthcoming Eighth Edition of the TNM Classification for Lung Cancer. *J Thorac Oncol.* 2015;10(7):990-1003.
13. Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WE, et al. The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. *J Thorac Oncol.* 2016;11(1):39-51.
14. Postmus PE, Kerr KM, Oudkerk M, Senan S, Waller DA, Vansteenkiste J, et al. Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol.* 2017;28 Suppl 4:iv1-iv21.
15. Group NM-aC, Arriagada R, Auperin A, Burdett S, Higgins JP, Johnson DH, et al. Adjuvant chemotherapy, with or without postoperative radiotherapy, in operable non-small-cell lung cancer: two meta-analyses of individual patient data. *Lancet.* 2010;375(9722):1267-77.

16. Douillard JY, Tribodet H, Aubert D, Shepherd FA, Rosell R, Ding K, et al. Adjuvant cisplatin and vinorelbine for completely resected non-small cell lung cancer: subgroup analysis of the Lung Adjuvant Cisplatin Evaluation. *J Thorac Oncol.* 2010;5(2):220-8.
17. Artal Cortes A, Calera Urquizu L, Hernando Cubero J. Adjuvant chemotherapy in non-small cell lung cancer: state-of-the-art. *Transl Lung Cancer Res.* 2015;4(2):191-7.
18. Chang JY, Bezjak A, Mornex F, Committee IART. Stereotactic ablative radiotherapy for centrally located early stage non-small-cell lung cancer: what we have learned. *J Thorac Oncol.* 2015;10(4):577-85.
19. Chang JY, Senan S, Paul MA, Mehran RJ, Louie AV, Balter P, et al. Stereotactic ablative radiotherapy versus lobectomy for operable stage I non-small-cell lung cancer: a pooled analysis of two randomised trials. *Lancet Oncol.* 2015;16(6):630-7.
20. O'Rourke N, Roque IFM, Farre Bernado N, Macbeth F. Concurrent chemoradiotherapy in non-small cell lung cancer. *Cochrane Database Syst Rev.* 2010(6):CD002140.
21. Auperin A, Le Pechoux C, Rolland E, Curran WJ, Furuse K, Fournel P, et al. Meta-analysis of concomitant versus sequential radiochemotherapy in locally advanced non-small-cell lung cancer. *J Clin Oncol.* 2010;28(13):2181-90.
22. Nestle U, Schimek-Jasch T, Kremp S, Schaefer-Schuler A, Mix M, Kusters A, et al. Imaging-based target volume reduction in chemoradiotherapy for locally advanced non-small-cell lung cancer (PET-Plan): a multicentre, open-label, randomised, controlled trial. *Lancet Oncol.* 2020;21(4):581-92.
23. Yuan S, Sun X, Li M, Yu J, Ren R, Yu Y, et al. A randomized study of involved-field irradiation versus elective nodal irradiation in combination with concurrent chemotherapy for inoperable stage III nonsmall cell lung cancer. *Am J Clin Oncol.* 2007;30(3):239-44.
24. Antoch G, Stattaus J, Nemat AT, Marnitz S, Beyer T, Kuehl H, et al. Non-small cell lung cancer: dual-modality PET/CT in preoperative staging. *Radiology.* 2003;229(2):526-33.
25. van Tinteren H, Hoekstra OS, Smit EF, van den Bergh JH, Schreurs AJ, Stallaert RA, et al. Effectiveness of positron emission tomography in the preoperative assessment of patients with suspected non-small-cell lung cancer: the PLUS multicentre randomised trial. *Lancet.* 2002;359(9315):1388-93.
26. Giraud P, De Rycke Y, Lavole A, Milleron B, Cosset JM, Rosenzweig KE. Probability of mediastinal involvement in non-small-cell lung cancer: a statistical definition of the clinical target volume for 3-dimensional conformal radiotherapy? *Int J Radiat Oncol Biol Phys.* 2006;64(1):127-35.
27. MacManus MP, Hicks RJ, Matthews JP, Hogg A, McKenzie AF, Wirth A, et al. High rate of detection of unsuspected distant metastases by pet in apparent stage III non-small-cell lung cancer: implications for radical radiation therapy. *Int J Radiat Oncol Biol Phys.* 2001;50(2):287-93.
28. Wen W, Xuan D, Hu Y, Li X, Liu L, Xu D. Prognostic value of maximum standard uptake value, metabolic tumor volume, and total lesion glycolysis of positron emission tomography/computed tomography in patients with breast cancer: A systematic review and meta-analysis. *PLoS One.* 2019;14(12):e0225959.
29. Park SY, Cho A, Yu WS, Lee CY, Lee JG, Kim DJ, et al. Prognostic value of total lesion glycolysis by 18F-FDG PET/CT in surgically resected stage IA non-small cell lung cancer. *J Nucl Med.* 2015;56(1):45-9.
30. Fledelius J, Winther-Larsen A, Khalil AA, Bylov CM, Hjorthaug K, Bertelsen A, et al. (18)F-FDG PET/CT for Very Early Response Evaluation Predicts CT Response

- in Erlotinib-Treated Non-Small Cell Lung Cancer Patients: A Comparison of Assessment Methods. *J Nucl Med.* 2017;58(12):1931-7.
31. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer.* 2009;45(2):228-47.
32. Seymour L, Bogaerts J, Perrone A, Ford R, Schwartz LH, Mandrekar S, et al. iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. *Lancet Oncol.* 2017;18(3):e143-e52.
33. Choi NC, Fischman AJ, Niemierko A, Ryu JS, Lynch T, Wain J, et al. Dose-response relationship between probability of pathologic tumor control and glucose metabolic rate measured with FDG PET after preoperative chemoradiotherapy in locally advanced non-small-cell lung cancer. *Int J Radiat Oncol Biol Phys.* 2002;54(4):1024-35.
34. Young H, Baum R, Cremerius U, Herholz K, Hoekstra O, Lammertsma AA, et al. Measurement of clinical and subclinical tumour response using [18F]-fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations. European Organization for Research and Treatment of Cancer (EORTC) PET Study Group. *Eur J Cancer.* 1999;35(13):1773-82.
35. O JH, Lodge MA, Wahl RL. Practical PERCIST: A Simplified Guide to PET Response Criteria in Solid Tumors 1.0. *Radiology.* 2016;280(2):576-84.
36. Min SJ, Jang HJ, Kim JH. Comparison of the RECIST and PERCIST criteria in solid tumors: a pooled analysis and review. *Oncotarget.* 2016;7(19):27848-54.
37. Ordu C, Selcuk NA, Akosman C, Eren OO, Altunok EC, Toklu T, et al. Comparison of metabolic and anatomic response to chemotherapy based on PERCIST and RECIST in patients with advanced stage non-small cell lung cancer. *Asian Pac J Cancer Prev.* 2015;16(1):321-6.
38. Kim HD, Kim BJ, Kim HS, Kim JH. Comparison of the morphologic criteria (RECIST) and metabolic criteria (EORTC and PERCIST) in tumor response assessments: a pooled analysis. *Korean J Intern Med.* 2019;34(3):608-17.
39. Beer L, Hochmair M, Haug AR, Schwabel B, Kifjak D, Wadsak W, et al. Comparison of RECIST, iRECIST, and PERCIST for the Evaluation of Response to PD-1/PD-L1 Blockade Therapy in Patients With Non-Small Cell Lung Cancer. *Clin Nucl Med.* 2019;44(7):535-43.
40. Shang J, Ling X, Zhang L, Tang Y, Xiao Z, Cheng Y, et al. Comparison of RECIST, EORTC criteria and PERCIST for evaluation of early response to chemotherapy in patients with non-small-cell lung cancer. *Eur J Nucl Med Mol Imaging.* 2016;43(11):1945-53.
41. Kim JH. Comparison of the EORTC criteria and PERCIST in solid tumors: a pooled analysis and review. *Oncotarget.* 2016;7(36):58105-10.
42. Hicks RJ, Kalff V, MacManus MP, Ware RE, McKenzie AF, Matthews JP, et al. The utility of (18)F-FDG PET for suspected recurrent non-small cell lung cancer after potentially curative therapy: impact on management and prognostic stratification. *J Nucl Med.* 2001;42(11):1605-13.
43. Maffione AM, Ferretti A, Chondrogiannis S, Rampin L, Marzola MC, Grassetto G, et al. Proposal of a new 18F-FDG PET/CT predictor of response in rectal cancer treated by neoadjuvant chemoradiation therapy and comparison with PERCIST criteria. *Clin Nucl Med.* 2013;38(10):795-7.
44. Maffione AM, Marzola MC, Grassetto G, Chondrogiannis S, Giammarile F, Rubello D. Are PREDIST criteria better than PERCIST criteria as a PET predictor of preoperative treatment response in rectal cancer? *Nucl Med Commun.* 2014;35(8):890-2.

45. Caulo A, Mirsadraee S, Maggi F, Leccisotti L, van Beek EJ, Bonomo L. Integrated imaging of non-small cell lung cancer recurrence: CT and PET-CT findings, possible pitfalls and risk of recurrence criteria. *Eur Radiol.* 2012;22(3):588-606.
46. Kandathil A, Sibley RC, III, Subramaniam RM. Lung Cancer Recurrence: (18)F-FDG PET/CT in Clinical Practice. *AJR Am J Roentgenol.* 2019;213(5):1136-44.
47. Jimenez-Bonilla JF, Quirce R, Martinez-Rodriguez I, De Arcocha-Torres M, Carril JM, Banzo I. The Role of PET/CT Molecular Imaging in the Diagnosis of Recurrence and Surveillance of Patients Treated for Non-Small Cell Lung Cancer. *Diagnostics (Basel).* 2016;6(4).
48. Mattonen SA, Palma DA, Haasbeek CJ, Senan S, Ward AD. Early prediction of tumor recurrence based on CT texture changes after stereotactic ablative radiotherapy (SABR) for lung cancer. *Med Phys.* 2014;41(3):033502.
49. Huang K, Senthil S, Palma DA, Spoelstra FO, Warner A, Slotman BJ, et al. High-risk CT features for detection of local recurrence after stereotactic ablative radiotherapy for lung cancer. *Radiother Oncol.* 2013;109(1):51-7.
50. Tirkes T, Hollar MA, Tann M, Kohli MD, Akisik F, Sandrasegaran K. Response criteria in oncologic imaging: review of traditional and new criteria. *Radiographics.* 2013;33(5):1323-41.
51. Iravani A, Turgeon GA, Akhurst T, Callahan JW, Bressel M, Everitt SJ, et al. PET-detected pneumonitis following curative-intent chemoradiation in non-small cell lung cancer (NSCLC): recognizing patterns and assessing the impact on the predictive ability of FDG-PET/CT response assessment. *Eur J Nucl Med Mol Imaging.* 2019;46(9):1869-77.
52. Mattonen SA, Huang K, Ward AD, Senan S, Palma DA. New techniques for assessing response after hypofractionated radiotherapy for lung cancer. *J Thorac Dis.* 2014;6(4):375-86.
53. Hicks RJ, Mac Manus MP, Matthews JP, Hogg A, Binns D, Rischin D, et al. Early FDG-PET imaging after radical radiotherapy for non-small-cell lung cancer: inflammatory changes in normal tissues correlate with tumor response and do not confound therapeutic response evaluation. *Int J Radiat Oncol Biol Phys.* 2004;60(2):412-8.
54. Boyd JA, Hubbs JL, Kim DW, Hollis D, Marks LB, Kelsey CR. Timing of local and distant failure in resected lung cancer: implications for reported rates of local failure. *J Thorac Oncol.* 2010;5(2):211-4.
55. Consonni D, Pierobon M, Gail MH, Rubagotti M, Rotunno M, Goldstein A, et al. Lung cancer prognosis before and after recurrence in a population-based setting. *J Natl Cancer Inst.* 2015;107(6):djv059.
56. Reddy JP, Tang C, Shih T, Kim B, Kim C, Nguyen QN, et al. Influence of Surveillance PET/CT on Detection of Early Recurrence After Definitive Radiation in Stage III Non-small-cell Lung Cancer. *Clin Lung Cancer.* 2017;18(2):141-8.
57. Biggi A, Gallamini A, Chauvie S, Hutchings M, Kostakoglu L, Gregianin M, et al. International validation study for interim PET in ABVD-treated, advanced-stage hodgkin lymphoma: interpretation criteria and concordance rate among reviewers. *J Nucl Med.* 2013;54(5):683-90.
58. Barrington SF, Qian W, Somer EJ, Franceschetto A, Bagni B, Brun E, et al. Concordance between four European centres of PET reporting criteria designed for use in multicentre trials in Hodgkin lymphoma. *Eur J Nucl Med Mol Imaging.* 2010;37(10):1824-33.
59. Thie JA. Understanding the standardized uptake value, its methods, and implications for usage. *J Nucl Med.* 2004;45(9):1431-4.

60. Cerfolio RJ, Bryant AS, Winokur TS, Ohja B, Bartolucci AA. Repeat FDG-PET after neoadjuvant therapy is a predictor of pathologic response in patients with non-small cell lung cancer. *Ann Thorac Surg.* 2004;78(6):1903-9; discussion 9.
61. Arnett AL, Packard AT, Mara K, Mansfield AS, Wigle DA, Haddock MG, et al. FDG-PET parameters as predictors of pathologic response and nodal clearance in patients with stage III non-small cell lung cancer receiving neoadjuvant chemoradiation and surgery. *Pract Radiat Oncol.* 2017;7(6):e531-e41.
62. Huang K, Dahele M, Senan S, Guckenberger M, Rodrigues GB, Ward A, et al. Radiographic changes after lung stereotactic ablative radiotherapy (SABR)--can we distinguish recurrence from fibrosis? A systematic review of the literature. *Radiother Oncol.* 2012;102(3):335-42.
63. James WP. Research on obesity. *Nutr Bull.* 1977;4(3):187-90.
64. Turgeon GA, Iravani A, Akhurst T, Beaulieu A, Callahan JW, Bressel M, et al. What (18)F-FDG PET Response-Assessment Method Best Predicts Survival After Curative-Intent Chemoradiation in Non-Small Cell Lung Cancer: EORTC, PERCIST, Peter Mac Criteria, or Deauville Criteria? *J Nucl Med.* 2019;60(3):328-34.
65. Fledelius J, Khalil A, Hjorthaug K, Frokiaer J. Inter-observer agreement improves with PERCIST 1.0 as opposed to qualitative evaluation in non-small cell lung cancer patients evaluated with F-18-FDG PET/CT early in the course of chemoradiotherapy. *EJNMMI Res.* 2016;6(1):71.
66. Nestle U, Rischke HC, Eschmann SM, Holl G, Tosch M, Miederer M, et al. Improved inter-observer agreement of an expert review panel in an oncology treatment trial--Insights from a structured interventional process. *Eur J Cancer.* 2015;51(17):2525-33.
67. Chapet O, Kong FM, Quint LE, Chang AC, Ten Haken RK, Eisbruch A, et al. CT-based definition of thoracic lymph node stations: an atlas from the University of Michigan. *Int J Radiat Oncol Biol Phys.* 2005;63(1):170-8.
68. Schaefer A, Nestle U, Kremp S, Hellwig D, Grgic A, Buchholz HG, et al. Multi-centre calibration of an adaptive thresholding method for PET-based delineation of tumour volumes in radiotherapy planning of lung cancer. *Nuklearmedizin.* 2012;51(3):101-10.
69. Schimek-Jasch T, Troost EG, Rucker G, Prokic V, Avlar M, Duncker-Rohr V, et al. A teaching intervention in a contouring dummy run improved target volume delineation in locally advanced non-small cell lung cancer: Reducing the interobserver variability in multicentre clinical studies. *Strahlenther Onkol.* 2015;191(6):525-33.
70. Yu Y, Guan H, Xing LG, Xiang YB. Role of gross tumor volume in the prognosis of non-small cell lung cancer treated with 3D conformal radiotherapy: a meta-analysis. *Clin Ther.* 2015;37(10):2256-66.
71. Plathow C, Weber WA. Tumor cell metabolism imaging. *J Nucl Med.* 2008;49 Suppl 2:43S-63S.
72. Warburg O, Wind F, Negelein E. The Metabolism of Tumors in the Body. *J Gen Physiol.* 1927;8(6):519-30.
73. Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement.* 1960;20(1):687-99.
74. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159-74.
75. Scott W. Reliability of content analysis: The case of nominal scale coding. *Public Opinion Quarterly.* 1955;17(3):321-5.
76. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychological Bulletin.* 1971;76(5):378-82.

Danksagung

Der Einstieg in die Welt des wissenschaftlichen Arbeitens ist als Arzt nicht zwangsläufig einfach und bietet am Anfang viele kleine und große Herausforderungen, für deren Bezwingung der Rat eines erfahrenen Mentors ungeheuer wertvoll und wichtig ist. Ich bin zutiefst dankbar einen Doktorvater gefunden zu haben, der mich stets motivierte, meine Arbeit wertschätzte und einen unerschöpflichen Quell an fachlichen Ideen für dieses Projekt darstellte. Vielen Dank für all die Zeit, Kraft und Mühe, die Sie mit in diese Arbeit investiert haben. Ich hätte mir für diese Arbeit keine bessere Betreuung wünschen können.

Bei der Leitung der PET Plan Studie möchte ich mich herzlich für die Bereitstellung der Bild- und Patientendaten und ebenfalls für die fachlichen Ideen und freundlichen Ratschläge zu dieser Arbeit bedanken. In diesem Zusammenhang möchte ich auch allen Beteiligten für die Hilfe beim Sichten und Sortieren der Bilddaten danken.

Großer Dank gebührt natürlich ebenfalls den Auswertern der Bilddaten für die viele Zeit, die sie in die Arbeit bei der Auswertung der Bilddaten investiert haben und für das Feedback zu den Ergebnissen dieser Arbeit.

Auch ohne die Hilfe des IMBEI in Mainz wäre die Umsetzung dieser Arbeit nicht möglich gewesen, weshalb ich mich auch für die Bereitstellung und die Hilfe beim Umgang mit der Datenbank, sowie für die Ratschläge zu den statistischen Auswertungen bedanken möchte.

Dem Direktor der Klinik und Poliklinik für Nuklearmedizin möchte ich für die herzliche Aufnahme in die Abteilung der Nuklearmedizin und für seine fachlichen Ratschläge danken. Zudem danke ich den Physikern der Abteilung sehr herzlich für ihre Unterstützung im Umgang mit den Programmen zur Bilddatenauswertung und der Mithilfe beim Lösen zahlreicher kleiner technischer und administrativer Probleme mit den Bilddatensätzen.

Last but not least möchte ich meiner Familie und meinen Freunden für ihre stetige Unterstützung im Rahmen dieser Arbeit danken.

Vielen Dank!