

Aus der Abteilung  
für Nuklearmedizin  
Prof. Dr. Dirk Hellwig  
der Fakultät für Medizin  
der Universität Regensburg

Metabolisches Tumorvolumen in der FDG-PET/CT nach Induktionstherapie bei  
aggressiven Non-Hodgkin-Lymphomen

Inaugural – Dissertation  
zur Erlangung des Doktorgrades  
der Medizin

der  
Fakultät für Medizin  
der Universität Regensburg

vorgelegt von  
Anna Kirchmayr

2024



Aus der Abteilung  
für Nuklearmedizin  
Prof. Dr. Dirk Hellwig  
der Fakultät für Medizin  
der Universität Regensburg

Metabolisches Tumrvolumen in der FDG-PET/CT nach Induktionstherapie bei  
aggressiven Non-Hodgkin-Lymphomen

Inaugural – Dissertation  
zur Erlangung des Doktorgrades  
der Medizin

der  
Fakultät für Medizin  
der Universität Regensburg

vorgelegt von  
Anna Kirchmayer

2024

Dekan: Prof. Dr. Dirk Hellwig  
1. Berichterstatter: Prof. Dr. Dirk Hellwig  
2. Berichterstatter: Prof. Dr. Tobias Pukrop  
Tag der mündlichen Prüfung: 23.09.2025

## Inhaltsverzeichnis

1. Einleitung .....	7
1.1 Non-Hodgkin-Lymphome .....	7
1.1.1 Definition und Klassifikation .....	7
1.1.2 Epidemiologie und Ätiologie .....	7
1.1.3 Diagnostik .....	8
1.1.4 Therapie .....	9
1.2 Positronenemissionstomographie (PET) .....	10
1.2.1 F-18-2-Fluoro-2-desoxy-D-glucose (FDG) als Tracer .....	10
1.2.2 Funktionsweise der FDG-PET .....	11
1.2.3 Schwächungskorrektur .....	12
1.2.4 PET/CT .....	12
1.2.5 Quantifizierung der PET-Daten .....	12
1.2.6 Tumor-Sink-Effekt .....	15
1.3 Zielsetzung und Fragestellungen .....	15
2. Material und Methoden .....	16
2.1 OPTIMAL>60-Studie .....	16
2.1.1 Studiendesign .....	16
2.1.2 Definition der PET-Positivität nach Induktionschemotherapie .....	16
2.1.3 Datenembargo für die Analyse der OPTIMAL>60-Studie .....	17
2.1.4 Patientenkollektiv .....	17
2.2 Konturierungsverfahren .....	19
2.3 Inter-Observer-Vergleich .....	20
2.4 Statistische Analyse .....	21
3. Ergebnisse .....	25
3.1 Gegenüberstellung der visuellen Analysemethode und des Deauville-Scores .....	25
3.2 Deskriptive Statistik .....	26
3.2.1 Berechnungen für Läsionen mit einem Mindestvolumen von 1 ml .....	26
3.2.2 Berechnungen für Läsionen aller Volumina .....	27
3.3 Vergleich der MTV- und TLG-Ergebnisse unter Berücksichtigung der verschiedenen Volumengrenzen .....	29
3.4 Inter-Observer-Vergleich .....	29
3.4.1 Analyse nach Bland und Altman .....	29
3.4.2 Analyse der Verteilung und Korrelation mithilfe von Scatterplots, Histogrammen und Kendall's Tau Korrelationskoeffizienten .....	35
3.4.3 Interklassenkorrelationsanalyse .....	43

3.5	Tumor-Sink-Effekt .....	43
4.	Diskussion.....	47
4.1	Überblick über die Kernaussagen der Arbeit .....	47
4.2	Vergleich mit Patientenkollektiven anderer Studien .....	47
4.3	Vergleich mit Methodiken anderer Studien.....	48
4.4	Vergleich mit Ergebnissen anderer Studien .....	50
4.4.1	PET-Bewertungsverfahren .....	50
4.4.2	MTV-Messwerte .....	52
4.4.3	ROI-Größen.....	52
4.4.4	Inter-Observer-Vergleich .....	53
4.4.5	Tumor-Sink-Effekt .....	54
4.5	Limitationen.....	55
4.6	Ausblick.....	56
5.	Zusammenfassung.....	57
6.	Anhang .....	58
a.	Tabellenverzeichnis.....	58
b.	Abbildungsverzeichnis .....	59
c.	Abkürzungsverzeichnis .....	61
7.	Literaturverzeichnis .....	63
8.	Danksagung.....	71

# 1. Einleitung

Im Folgenden werden die Themen ‘Non-Hodgkin-Lymphome‘ und ‘Positronenemissionstomographie (PET)‘ näher ausgeführt, sowie die Zielsetzung der Arbeit zusammen mit den Fragestellungen erläutert.

## 1.1 Non-Hodgkin-Lymphome

Dieser Abschnitt gibt Auskunft über die Definition, Klassifikation, Epidemiologie, Ätiologie, Diagnostik und Therapie der Non-Hodgkin-Lymphome.

### 1.1.1 Definition und Klassifikation

Non-Hodgkin-Lymphome (NHL) sind Hämoblastosen, bestehend aus maligne entarteten Zellen des Immunsystems (1). Es handelt es sich um Neoplasien der B-, T-Zellen und/oder natürlichen Killerzellen. Typischerweise infiltrieren sie lymphatisches und/oder hämatopoisches Gewebe. Zusätzlich können aber auch andere Organe befallen werden. Die Erkrankung äußert sich folglich häufig als - meist schmerzlose - Lymphadenopathie (nodaler Befall) oder mit dem Auftreten von extranodalen Lymphomherden (2).

NHL werden nach der Klassifikation der World Health Organization (WHO) von 2016 in über 50 Subtypen unterteilt (3). Dabei lassen sich die Non-Hodgkin-Lymphome in high-grade bzw. aggressive NHL und low-grade bzw. indolente NHL einteilen (1). Zu den häufigsten Subtypen zählen das indolente folliculäre Lymphom und das aggressive diffuse großzellige B-Zell-Lymphom (DLBCL) (4). Im Jahr 2022 wurde mit der fünften Ausgabe der WHO-Klassifikation hämatologischer und lymphatischer Neoplasien eine aktualisierte Darstellung veröffentlicht. Diese ermöglicht unter anderem eine genauere Einteilung in einige Lymphomsubtypen und legt größeren Fokus auf genetische und molekulare Diagnostik und Marker. In der klinischen Routine hat sich die aktualisierte Klassifikation jedoch bisher noch nicht durchgesetzt (5). Als klinischer Handlungsleitfaden zur Diagnostik, Therapie und Nachsorge von Patienten mit diffusem großzelligen B-Zell-Lymphom dient die dazugehörige internationale, evidenzbasierte S3-Leitlinie (6), die als erste ihrer Art im Jahr 2022 veröffentlicht wurde (7).

### 1.1.2 Epidemiologie und Ätiologie

Non-Hodgkin-Lymphome stellen weltweit die häufigste maligne hämatologische Erkrankung dar (4). Den neuesten GLOBOCAN Daten zufolge machten sie 2022 etwa 2.8 % der Krebsneuerkrankungen weltweit aus (8). Für Deutschland ergaben die aktuellsten Daten des Robert Koch-Instituts (RKI) für das Jahr 2020 eine Inzidenz von 8230 unter den Frauen und 10090 unter den Männern (9).

Die Ätiologie der Non-Hodgkin-Lymphome ist noch nicht abschließend geklärt. Eine bedeutende Rolle in der Entstehung der NHL spielen die Anhäufung von modifizierter DNA und onkogenetische Mutationen. Bei einigen Lymphomsubtypen sind auch einzelne Mutationen bekannt, die mit der Lymphomentstehung assoziiert sind, zum Beispiel die Translokation des bcl-2-Onkogens beim follikulären Lymphom (10).

In der Literatur werden einige Faktoren genannt, die die Entstehung eines Non-Hodgkin-Lymphoms möglicherweise begünstigen. Dazu zählen einerseits Viren. Das Ebstein-Barr-Virus (EBV) soll beispielsweise einen Beitrag zur Entstehung des Burkitt-Lymphoms leisten (11). Andererseits spielen womöglich auch vorausgegangene Chemo- und Radiotherapie, Immunschwäche und Umweltgifte eine begünstigende Rolle in der Genese von NHL (2, 9).

### 1.1.3 Diagnostik

Zur Diagnosesicherung ist eine Biopsie erforderlich. Am besten eignet sich eine Exzisionsbiopsie. Anschließend muss diese von erfahrenem ärztlichen Personal der Pathologie untersucht werden (6).

Das Ausmaß und die Lokalisation der Tumormasse wird im Rahmen des Stagings bestimmt. Die Stadieneinteilung basiert auf der anfänglich für Hodgkin-Lymphome entwickelten Ann Arbor-Klassifikation. Die letzten Jahrzehnte war die Computertomographie (CT) das führende bildgebende Verfahren. Dieses wurde mittlerweile von der Kombination aus Positronen-Emissions-Tomographie (PET) und Computertomographie, kurz PET/CT, mit dem Radiopharmakon F-18-Fluorodesoxyglukose (FDG) abgelöst. Nur bei nicht FDG-avid Lymphomen, die mit FDG nicht sicher nachgewiesen werden können, wird weiterhin auf die alleinige CT zurückgegriffen (12).

Für die Beurteilung des Therapieansprechens der Non-Hodgkin-Lymphome wird die Lugano-Klassifikation von 2014 benutzt (13). Zuvor erfolgte die Beurteilung anhand visueller Kriterien des International Harmonization Project (IHP) mit PET-Positivität bei Mehrspeicherung einer Region mit einem Mindestdurchmesser von 2 cm im Vergleich zum mediastinalen Blutpool. Kleinere Läsionen wurden als positiv gewertet, wenn ihr FDG-Speicherverhalten visuell das des umliegenden Hintergrunds übersteigt (14).

Mit Einführung der Lugano-Klassifikation wurde auch eine 5-Punkte-Skala, der Deauville-Score (DS) (Tabelle 1), festgelegt, um das Therapieansprechen der Non-Hodgkin-Lymphome mittels PET/CT zu beurteilen. Dazu wird die Anreicherungsintensität in Herden im Verhältnis zur FDG-Speicherung im mediastinalen Blutpool und der Leber betrachtet (13). Um die Abgrenzung zwischen DS 4 und DS 5 zu objektivieren, hat man inzwischen das Doppelte der maximalen FDG-Speicherintensität in der Leberreferenzregion als Unterscheidungskriterium eingeführt (15).

---

**DS 1** Keine FDG-Speicherung

---

**DS 2** FDG-Speicherung (Tumor)  $\leq$  FDG-Speicherung (Mediastinaler Blutpool)

---

**DS 3** FDG-Speicherung (Tumor)  $>$  FDG-Speicherung (Mediastinaler Blutpool)  $\leq$  FDG-Speicherung (Leber)

---

**DS 4** FDG-Speicherung (Tumor) ist mäßig höher als FDG-Speicherung (Leber)

---

**DS 5** FDG-Speicherung (Tumor) ist deutlich höher als FDG-Speicherung (Leber) und/oder neue Tumorherde

---

**DS X** Neue FDG-Speicherungen, die mit geringer Wahrscheinlichkeit mit vorausgegangenem Tumor assoziiert sind

---

*Tabelle 1: Deauville-Score zur Beurteilung des Therapieansprechens der Non-Hodgkin-Lymphome*

Liegen nach Induktionstherapie FDG-Speicherintensitäten von DS 1 bis DS 3 vor, spricht man von *complete metabolic response*. Das bedeutet, dass keine bzw. kaum mehr Stoffwechselaktivität nachweisbar ist und die noch vorhandenen Lymphomreste in der PET/CT als nicht vital angesehen werden. Beim Vorhandensein der Stufen DS 4 und DS 5 nach Therapie geht man von einer *residual metabolic disease* aus. Das ausgewählte Lymphomareal ist noch metabolisch aktiv, weshalb man von residuellem, aktiven Tumorgewebe ausgeht. Steigt bzw. bleibt die Stoffwechselaktivität im bekannten Lymphomherd auf Stufe DS 5 und/oder finden sich neue FDG-avide Lymphomherde, die mit dem bereits vorhandenen Lymphom einhergehen, klassifiziert man den Befund als Therapieversagen und/oder Progression (13).

#### 1.1.4 Therapie

Die Auswahl der Therapie ist unter anderem vom Subtyp und Stadium der Erkrankung abhängig. Die Standardtherapie des diffusen großzelligen B-Zell-Lymphoms, dem häufigsten Subtyp der aggressiven NHL, beinhaltet eine Kombination aus Chemotherapie und Immuntherapie. Dabei zählt das R-CHOP-Regime, bestehend aus Rituximab, Cyclophosphamid, Doxorubicin, Vincristin und Prednison, zur Standardtherapie. Patienten mit

therapieresistenten Lymphomherden können von aufstrebenden immunmodulatorischen Therapien wie der Chimären Antigenrezeptor-T-Zell-Therapie (CAR-T-Zell-Therapie) oder bispezifischen Antikörpern (BiTEs) profitieren (16). Neben der Immunochemotherapie zählt auch die Strahlentherapie zu den wichtigen Therapiesäulen der NHL, insbesondere bei Patienten mit residual metabolic disease in der End-of-Treatment-PET (EoT-PET) (17).

Um die Prognose von Patienten mit aggressiven Non-Hodgkin-Lymphomen einzuschätzen und damit auch die Therapieplanung zu steuern, wurde der International Prognostic Index (IPI) eingeführt. Der IPI ist ein klinischer Score, der Patienten in vier Risikogruppen, basierend auf der erreichten Gesamtpunktzahl in den untenstehenden Kategorien, einteilt: 0 bis 1 Punkt  $\triangleq$  geringes Risiko, 2 Punkte  $\triangleq$  niedrig-intermediäres Risiko, 3 Punkte  $\triangleq$  hoch-intermediäres Risiko und 4 bis 5 Punkte  $\triangleq$  hohes Risiko. Das Zutreffen der folgenden Kriterien wird jeweils mit einem Punkt bewertet: Alter  $> 60$  Jahre, Laktatdehydrogenase (LDH) höher als der obere Normwert, Eastern Cooperative Oncology Group (ECOG)-Status  $\geq 2$ , mehr als eine extralymphatische Lymphommanifestation und Ann-Arbour-Stadium III oder IV (18).

Seit der Veröffentlichung des IPI in den 1990er Jahren, wurde der Score fortlaufend weiterentwickelt. Mit der Optimierung des Therapieschemas für NHL durch den monoklonalen Antikörper Rituximab wurde auch der IPI weiterentwickelt und der revised-International Prognostic Index (r-IPI) speziell für Patienten mit neu diagnostiziertem diffusen großzelligen B-Zell-Lymphom, behandelt mit R-CHOP, erstellt (19).

## 1.2 Positronenemissionstomographie (PET)

Das sich anschließende Kapitel beinhaltet Erläuterungen zum radioaktiven Tracer FDG, zur Durchführung und Funktionsweise der PET, der Schwächungskorrektur, der PET/CT, der Quantifizierung von PET-Daten und zum Tumor-Sink-Effekt.

### 1.2.1 F-18-2-Fluoro-2-desoxy-D-glucose (FDG) als Tracer

Die PET nutzt einen radioaktiven Tracer, welcher dem Patienten intravenös verabreicht wird. Zur Detektion von Non-Hodgkin-Lymphomen eignet sich F-18-2-Fluoro-2-desoxy-D-glucose, kurz FDG, ein radioaktiv markierter Zucker (20). Chemisch gesehen entspricht FDG einem D-Glucose-Molekül ( $C_6H_{12}O_6$ ), dessen Hydroxygruppe (OH-Gruppe) am C<sub>2</sub>-Atom gegen ein radioaktives Fluor-Atom getauscht worden ist (21). FDG folgt dann dem gleichen Weg, nämlich über Glukosetransporter, in die Zellen wie Glukose, die mit der Nahrung oral zugeführt wird. Folglich gelangt FDG vorrangig in Körpergewebe, die einen hohen Glukoseumsatz haben (20). In den 1920er Jahren postulierte Otto Warburg, dass Tumorzellen im Vergleich

zum umgebenden Gewebe vermehrt Zucker aufnehmen (22), was auf eine Überexpression der Glukosetransporter GLUT-1 und GLUT-3 zurückgeführt wird (23). Der Warburg-Effekt besagt, dass maligne Zellen Glucose auch in Präsenz von ausreichenden Sauerstoffmengen verstärkt zu Laktat verstoffwechseln (22). Der erhöhte Glukosebedarf der stoffwechselaktiven Tumorzellen wird durch die Präferenz der ineffizienten anaeroben Glykolyse gegenüber der aeroben Glykolyse verstärkt. Die gesteigerte Glykolyse wird als Strategie der Tumorzellen, Überleben und Wachstum in Umgebungen mit niedriger Sauerstoffkonzentration zu sichern, gesehen (23).

Die FDG-Anreicherung in den Tumorzellen lässt sich wie folgt erklären: FDG wird nach Aufnahme in die Zelle durch das Enzym Hexokinase am C<sub>6</sub>-Atom phosphoryliert. Dabei entsteht FDG-6-Phosphat, welches, anders als D-Glucose, in der Zelle akkumuliert, da es nicht weiter verstoffwechselt werden kann (Trapping-Mechanismus) (24).

### 1.2.2 Funktionsweise der FDG-PET

Beim Zerfall des radioaktiven <sup>18</sup>F der FDG entstehen Positronen, die Antiteilchen der Elektronen. Das Positron wird durch Streuung an umgebenden Atomhüllen im Gewebe der Untersuchten abgebremst. Sobald das Positron langsam genug ist, bildet es mit einem Elektron für kurze Zeit ein Positronium. Im Anschluss folgt der Prozess der Annihilation. Dabei wandeln sich die Massen des Positrons und des Elektrons in zwei Photonen mit jeweils 511 keV Energie um (24).

Diese zwei Photonen entfernen sich mit Lichtgeschwindigkeit von ihrem Entstehungsort in entgegengesetzter Richtung, was einem Winkel von 180° entspricht. Da Elektron und Positron bei diesem Vorgang vernichtet werden, wird die dabei entstehende Photonenstrahlung auch Vernichtungsstrahlung oder Annihilationsstrahlung genannt. Die statischen ringförmig angeordneten Detektoren des PET-Scanners registrieren die emittierten Photonen. Die Linie zwischen den beiden gegenüberliegenden Detektoren, die die Photonen registriert haben, nennt man *Line of Response* (LOR). Diese LOR repräsentieren die stattgefundenen Annihilationsereignisse und können graphisch in einem Sinogramm dargestellt werden, wobei jede Zeile einen bestimmten Winkel der LOR repräsentiert und jede Spalte einer bestimmten Detektorposition entspricht (25). Aus dem Sinogramm kann mithilfe mathematischer Methoden ein Bild rekonstruiert werden (26).

Da FDG-6-Phosphat mittels Glucose-6-Phosphatase in FDG zurückverwandelt werden und die Zelle als solches wieder verlassen kann, sollten die PET-Aufnahmen innerhalb der ersten

beiden Stunden nach Verabreichung des Tracers stattfinden. In dieser Zeit kann die Rückreaktion noch vernachlässigt werden (24).

### 1.2.3 Schwächungskorrektur

Die Schwächungskorrektur ist ein essenzieller Prozess, um die korrekte Radioaktivitätsverteilung im Körper darzustellen. Die emittierten Photonen werden im Gewebe des Patienten zum Teil absorbiert oder gestreut, bevor sie die Detektoren erreichen. Dadurch werden falsch-niedrige Radioaktivitätskonzentrationen dargestellt, welche zu Ungenauigkeiten in den PET-Bildern führen (24). Um dem entgegenzuwirken, besitzt jeder PET-Scanner, außer die PET/MRT, die Möglichkeit, eine Transmissionsmessung durchzuführen. Dazu dient heutzutage bei der PET/CT die CT-Komponente. Anhand der in der CT ermittelten lokalen Schwächung kann berechnet werden, wie stark die Strahlung von dem dazwischenliegenden Patienten abgeschwächt wird (25).

### 1.2.4 PET/CT

Kombinierte PET/CT-Geräte ermöglichen die aufeinanderfolgende Anfertigung von Computertomographien und Positronenemissionstomographien am selben Ort, ohne dass der Patient an verschiedene Gerätstandorte transportiert werden muss. Dazu befinden sich PET-Systeme und CT-Systeme meist hintereinander in einem gemeinsamen Außengehäuse (Gantry). Nach der Bildakquisition können die Datensätze überlagert werden, sodass Fusionsbilder entstehen (25). Zu den größten Vorteilen der Hybrid-Bildgebung PET/CT zählt zum einen die Möglichkeit, mithilfe der CT das exakte anatomische Korrelat der Mehrspeicherungen in der PET zu identifizieren, zum anderen die CT-basierte Schwächungskorrektur (23).

### 1.2.5 Quantifizierung der PET-Daten

Um einen Therapieerfolg bzw. ein mögliches Therapieversagen objektivieren und möglicherweise prognostizieren zu können, werden unter anderem folgende PET-Bild-basierte Biomarker verwendet: SUV<sub>max</sub>, MTV und TLG (27, 28).

Der *Standardized Uptake Value* (SUV) ermöglicht die Quantifizierung der Anreicherung von radioaktiven Tracern im Gewebe. In der onkologischen Bildgebung beschreibt der SUV den Glucosestoffwechsel in Tumorherden und Normalgewebe. Der Wert beschreibt das Verhältnis zwischen in der PET gemessenen Radioaktivitätskonzentration im Gewebe und einer hypothetischen Gleichverteilung der applizierten Radioaktivitätsmenge in der gesamten Körpermasse des Patienten. Der SUV errechnet sich zusammen über die kalibrierten PET-

Bilddaten aus der verabreichten Dosis des Radiopharmakons und dem Gewicht des Patienten mithilfe folgender Formel (29):

$$SUV = \frac{\text{Radioaktivitätskonzentration[kBq/ml]}}{\text{Applizierte Aktivität[MBq] / Körpergewicht[kg]}}$$

Die Messung des SUV wird unter anderem von biologischen Faktoren wie dem Körpergewicht, dem Blutzuckerspiegel und der Dauer der post-FDG-Injektionszeit beeinflusst. Die Einflussnahme dieser biologischen Faktoren erklärt sich wie folgt: Fettgewebe weist eine geringere FDG-Aufnahme als Muskulatur auf, wodurch Patienten mit hohem Körperfettanteil höhere SUVs verzeichnen können. Hohe Blutzuckerspiegel hemmen kompetitiv die Aufnahme von FDG, wodurch abhängig vom Blutglucosespiegel geringere SUVs gemessen werden können. F-18-2-Fluoro-2-desoxy-D-glucose wird aus normalem, gesunden Gewebe schneller entfernt als aus malignem Gewebe, weshalb eine längere Aufnahmezeit ab einem gewissen Punkt zu höheren SUVs führen kann (29).

$SUV_{\max}$  und  $SUV_{\text{mean}}$  bezeichnen den maximalen bzw. mittleren SUV in der untersuchten Tumorregion.  $SUV_{\text{peak}}$  beschreibt den durchschnittlichen SUV in einem  $1 \text{ cm}^3$  großen Bereich des Tumors mit der höchsten Stoffwechselaktivität (30).

MTV, *Metabolic Tumor Volume*, ist definiert als die FDG-PET-positive Gesamttumormasse. Die TLG, *Tumor Lesion Glycolysis*, vereint die volumetrische und metabolische Komponente der oben definierten Größen. Sie errechnet sich als Produkt aus MTV und  $SUV_{\text{mean}}$  (31):

$$TLG = SUV_{\text{mean}} \times MTV$$

Um die oben beschriebenen Werte messen zu können, werden ROIs benutzt. Eine *Region of Interest*, ROI, bezeichnet ein definiertes Areal innerhalb eines Bildes, das für Analysen verwendet werden kann. Eine ROI kann um einen Tumorherd gezeichnet werden, um diesen zu markieren und die Tracer-Aufnahme dort zu messen (32).

#### 1.2.5.1 Ermittlung des MTV

Da die Ermittlung des MTV bisher nicht abschließend standardisiert ist, kommt es aufgrund nachfolgend genannter Faktoren zu variablen Ergebnissen. Dazu zählt zum einen die Bildqualität, welche von der Scannerleistung des PET/CT-Systems und den Rekonstruktionsalgorithmen abhängig ist (33). Zum anderen sind für die Messung des MTV auch die Expertise der Beobachter, die die Dignität des mit FDG angereicherten Gewebes einschätzen, und die Wahl der Segmentierungsmethode entscheidend (34).

Die Segmentierung in der medizinischen Bildverarbeitung bezeichnet den Prozess der Identifikation und Abgrenzung eines Tumors. Dabei unterscheidet man vor allem zwischen Schwellenwert-basierten und Algorithmus-basierten Verfahren. Erstere nutzen einen festgelegten Wert, um das Bild in Tumor und Hintergrund zu unterteilen, beispielsweise einen SUV von 2.5. Regionen in der PET, deren SUV größer als 2.5 ist, werden als Tumor gewertet, Areale mit niedrigerem SUV zählen zum Hintergrund (35). Neben dem häufig verwendeten absoluten Schwellenwert SUV von 2.5 existieren auch relative Schwellenwerte, wie der von der European Association of Nuclear Medicine (EANM) empfohlene Grenzwert von 41 % (36). Hintergrund-basierte Schwellenwerte berechnen sich aus den SUVs der Referenzregionen Leber bzw. Aorta oder der unmittelbaren Tumorumgebung und sind damit patienten- und scanspezifisch (35).

Algorithmus-basierte Segmentierungsverfahren beinhalten unter anderem Gradienten-basierte Methoden, die die Tumorgrenze anhand des Gradienten zwischen hohem SUV in Tumorzellen und niedrigem SUV in benachbartem Gewebe definieren, und Clustering. Letztere teilt Voxel, dreidimensionale Pixel, anhand bestimmter Eigenschaften wie SUVs, in Gruppen ein. Nach der Gruppierung können die Cluster als verschiedene Regionen im Bild identifiziert werden, zum Beispiel Tumor oder Hintergrund. Bei maschinellem Lernen wird ein Algorithmus mit PET-Datensätzen mit präzise gelegten ROIs gespeist, um das Vorhersagen der optimalen ROIs für später eingeschleuste PET-Bilder zu lernen (35).

Vor allem die Segmentierung kleiner Läsionen wird durch technische Limitationen erschwert, wodurch Herde mit geringem Volumen aus praktischen Gründen oft von der Analyse ausgeschlossen werden. Eben genannte Schwierigkeiten beruhen zum einen auf der niedrigen räumlichen Auflösung der PET. Diese führt zum sogenannten Partialvolumeneffekt. Dabei erscheinen kleine Läsionen volumetrisch größer, ihre tatsächliche Tracer-Aufnahme jedoch geringer, da die Aktivität auf benachbarte Areale ausläuft, was auch spill-out-effect genannt wird (37). Stark speichernde Läsionen erscheinen folglich größer aber weniger intensiv (38). Zum anderen beeinflusst auch ein niedriges Signal-Rausch-Verhältnis die Analyse von kleinen Herden negativ, wenn das Signal der Läsionen gegenüber dem Hintergrundrauschen zu klein ist. Kleine Läsionen sind mit ihrem geringen Radioaktivitätssignal in diesem Fall nur wenig vom Hintergrund zu unterscheiden und folglich schwer zu erkennen (37).

Technische Fortschritte in der PET-Diagnostik mit neuen Rekonstruktionsverfahren können das Hintergrundrauschen reduzieren, jedoch ist die Sichtbarkeit einer kleinen Läsion auch

zusätzlich von dem Radioaktivitätssignal der umliegenden Organe abhängig. Bisher gibt es keine klaren Empfehlungen, inwiefern Läsionsgrößen in die Bewertung von Lymphomen gemäß der Lugano-Klassifikation einfließen sollen (30).

### 1.2.6 Tumor-Sink-Effekt

Der Tumor-Sink-Effekt beschreibt ein Phänomen in der PET, bei welchem die Traceranreicherung im gesunden Gewebe durch Vorhandensein von großen, metabolisch aktiven Tumormassen reduziert wird (39). Diese Erscheinung lässt sich möglicherweise wie folgt erklären: Wenn Gewebe eine erhöhte Radionuklidspeicherung zeigt, wie beispielsweise ein Tumorherd, und damit den radioaktiven Tracer ‘wegfängt’, steht weniger FDG für gesundes Gewebe zur Verfügung (40). In gleicher Weise gibt es den Effekt nach Chemotherapie, wenn nach Gabe von G-CSF, granulocyte-colony stimulating factor, durch das stimuliertes Knochenmark FDG ‘sequestriert’ wird, welches dann zu verminderter Anreicherung des Tracers im gesunden Gewebe, aber auch in Lymphommanifestationen führt (41).

## 1.3 Zielsetzung und Fragestellungen

Zu den Zielen der vorliegenden Arbeit zählen die Analyse der PET-Positivität und des metabolischen Tumorvolumens der End-of-Treatment-PET bei Patienten mit aggressiven Non-Hodgkin-Lymphomen anhand des Patientenkollektivs der OPTIMAL>60-Studie. Dazu werden folgende Hauptforschungsfragen bearbeitet:

Zum einen wird ein Vergleich zweier Analysemethoden, die Beurteilung nach visuellen Kriterien und die Einschätzung der Tumormasse anhand des quantitativen SUV-basierten Deauville-Scores, durchgeführt, soweit dies aktuell ohne Vorliegen von Überlebensdaten möglich ist. Zum anderen wird auch untersucht, inwiefern sich eine Volumengrenze von 1 ml bei der Auswahl der ROIs auf das MTV und die TLG auswirkt. Der Deauville-Score dient auch als Beurteilungsmethode der Wahl im Inter-Observer-Vergleich. Dabei wird anhand eines Subkollektivs von 83 Patienten unter anderem die Reproduzierbarkeit der erhobenen Parameter MTV, TLG,  $SUV_{max}$ , Anzahl der ROIs und SUVs der Referenzregionen Aorta und Leber betrachtet. Zusätzlich wird analysiert, inwiefern die FDG-Aufnahme der Leber und der Aorta nach Induktionstherapie mit dem Tumorvolumen der Patienten in Zusammenhang steht, ob also ein Tumor-Sink-Effekt zu beobachten ist.

## 2. Material und Methoden

Nachfolgend werden die OPTIMAL>60-Studie, das analysierte Patientenkollektiv, die angewandten Konturierungsverfahren, der durchgeführte Inter-Observer-Vergleich und die statistische Analyse näher beschrieben.

### 2.1 OPTIMAL>60-Studie

#### 2.1.1 Studiendesign

Analysiert werden FDG-PET-Aufnahmen von Patienten der OPTIMAL>60-Studie (NCT01478542), einer deutschlandweiten, multizentrischen, prospektiven, randomisierten Studie mit positivem Ethikvotum, durchgeführt zwischen November 2011 und Januar 2024. Die OPTIMAL>60-Studie schloss insgesamt 1152 über 60-jährige Patienten mit neu diagnostizierten, CD20-positiven, aggressiven Non-Hodgkin-Lymphomen ein, für welche die Abteilung für Nuklearmedizin am Universitätsklinikum Regensburg die zentrale Referenzbefundung der FDG-PET-Untersuchungen durchführt.

Die OPTIMAL>60-Studie wurde entworfen, um das progressionsfreie Überleben der 61- bis 80-jährigen Patienten mit CD20-positiven, aggressiven B-Zell-Non-Hodgkin-Lymphomen zu verbessern und die Toxizität der standardmäßig eingesetzten Immuno-Chemotherapie-Medikation zu verringern. Letzteres beinhaltet ein optimiertes Verabreichungsschema des monoklonalen Antikörpers Rituximab, den Ersatz von konventionellem Vincristin durch liposomales Vincristin und die PET-gestützte Therapiereduktion in Kombination mit Vitamin D-Substitution. Anhand der prä-therapeutischen, initial durchgeführten FDG-PET und des International Prognostic Index (IPI) wurde das Patientenkollektiv in zwei Kohorten eingeteilt, eine mit günstiger Prognose und eine mit ungünstiger Prognose. Anschließend wurden diese in die verschiedenen Therapiearme *Favorable A und B* für Patienten mit günstiger Prognose und *Less Favourable A, B, C und D* für Patienten mit ungünstiger Prognose randomisiert.

#### 2.1.2 Definition der PET-Positivität nach Induktionschemotherapie

20 Tage nach der letzten Induktionschemotherapie wurde unter anderem eine „definitive Restaging“-FDG-PET (entspricht einer End-of-Treatment-PET) durchgeführt und von zwei unabhängigen, erfahrenen Nuklearmedizinern anhand visueller Kriterien beurteilt, die noch vor Einführung der Lugano-Kriterien (siehe 1.1.3) für diese Studie vorgegeben wurden. Vergleichbare Studien aus der Zeit der Konzeption der OPTIMAL>60-Studie, wie die Untersuchung von Gallamini et al. (42), haben die gleichen Kriterien zur Bewertung herangezogen. Ein positiver Scan wurde dabei definiert als eine fokale oder diffuse FDG-

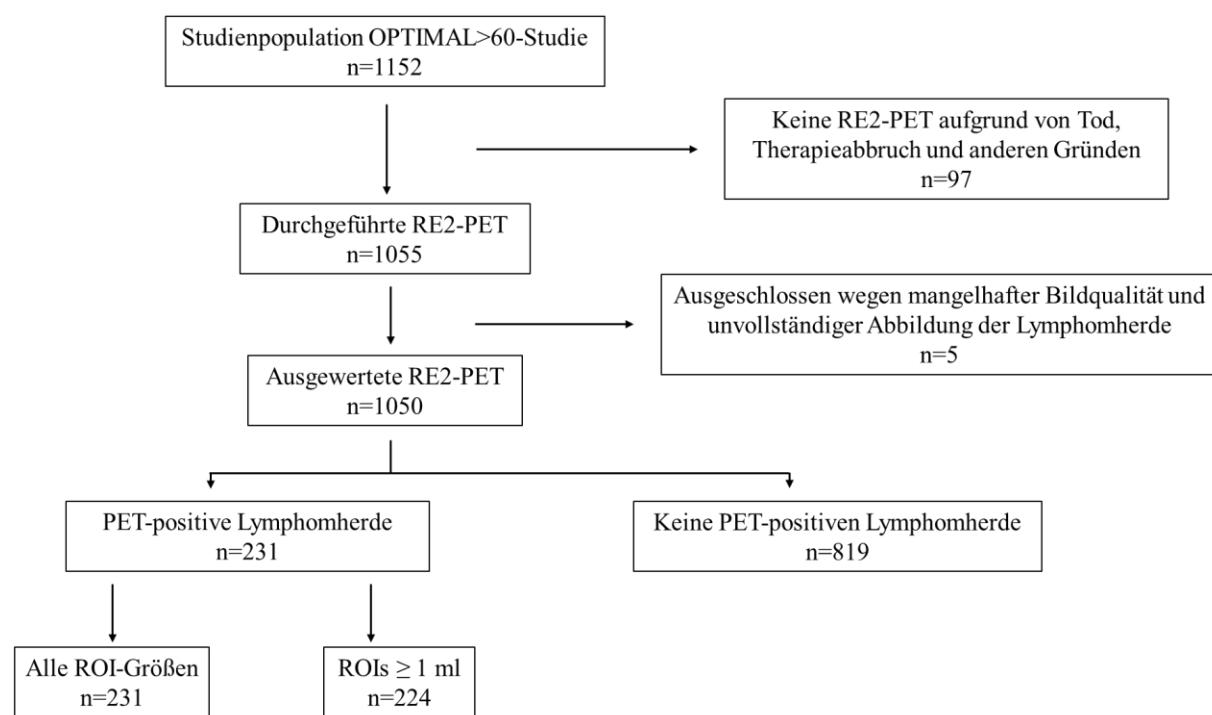
Aufnahme größer als der Hintergrund in einer Lokalisation, welche nicht mit einer physiologischen FDG-Anreicherung vereinbar ist, ohne spezifische SUV-Schwelle. Diese Bewertung entspricht den IHP-Kriterien (14). Abhängig vom PET-Befund folgte in der OPTIMAL>60-Studie die entsprechende Konsolidierungstherapie gemäß den Schemata der jeweiligen Therapiearme.

### 2.1.3 Datenembargo für die Analyse der OPTIMAL>60-Studie

Nach mehrjähriger Nachbeobachtung werden die aktuell noch unter Embargo stehenden Überlebensdaten nach Freigabe durch die Studienzentrale erst zu Beginn des Jahres 2025 verfügbar sein, welche aufgrund der zeitlich verzögerten Verfügbarkeit nicht in die nachstehende Analyse einfließen können.

### 2.1.4 Patientenkollektiv

1050 Patienten mit neu diagnostizierten, CD20-positiven, aggressiven Non-Hodgkin-Lymphomen der OPTIMAL>60-Studie bilden das zu analysierende Patientenkollektiv der vorliegenden Arbeit. Die Auswahl der eingeschlossenen und der Ausschluss der übrigen Patienten wird durch Abbildung 1 verdeutlicht.



**Abbildung 1:** STARD-Flowchart zur Darstellung des analysierten Patientenkollektivs. RE2-PET: Restaging-PET, STARD: Standards for Reporting of Diagnostic Accuracy

Von den 1152 in die OPTIMAL>60-Studie eingeschlossenen Patienten erhielten 1055 Personen eine End-of-Treatment-PET. Bei 97 der 1152 Patienten standen aufgrund von Tod, Therapieabbruch oder anderen Gründen keine PET-Aufnahmen zum Zeitpunkt des Restaging

für die Analyse zur Verfügung. Von den 1055 bereitgestellten PET-Bilddaten mussten fünf Fälle aufgrund von mangelhafter Bildqualität oder unvollständiger Abbildung der Lymphomherde ausgeschlossen werden. Die übrigen 1050 PET-Datensätze definieren das untersuchte Patientenkollektiv der vorliegenden Arbeit. In Hinblick auf residuelles vitales Lymphomgewebe zum Zeitpunkt des Restaging zeigten 231 der 1050 untersuchten Patienten PET-positive Herde in der EoT-PET und 819 Personen keine PET-positiven Lymphomherde.

Zur näheren Charakterisierung des Subkollektivs der 231 Patienten mit PET-positiven Lymphomherden in der EoT-PET eignen sich folgende klinische Parameter, welche prätherapeutisch erhoben und in den Randomisationsunterlagen der Studie dokumentiert wurden: Alter der Patienten, Höhe der Laktatdehydrogenase (LDH), ECOG, Anzahl der extralymphatischen Manifestationen, Ann-Arbour-Stadium, Vorhandensein eines Bulk ( $\geq 7.5$  cm) und der IPI. Einen Überblick über die erhobenen Werte ermöglicht die Tabelle 2.

Variable	Kategorie	N	%
Alter [Jahre]	$\geq 60$	231	100
	< 60	0	0
LDH [U/l]	erhöht	137	59.3
	normwertig	94	40.7
ECOG	0	112	48.5
	1	103	44.6
	2	14	6.1
	3	1	0.4
	4	0	0
	5	0	0
extralymphatische Manifestationen	fehlend	1	0.4
	$> 1$	92	39.8
Ann-Arbour-Stadium	$\leq 1$	139	60.2
	I	38	16.5
IPI	II	52	22.5
	III	48	20.8
	IV	93	40.3
Bulk ( $\geq 7.5$ cm)	1	48	20.8
	2	50	21.6
	3	71	30.7
	4	54	23.4
	5	8	3.5
Patienten mit Bulk	Patienten mit Bulk	111	48.1
	Patienten ohne Bulk	120	51.9

*Tabelle 2: Klinische Charakteristika des Subkollektivs mit PET-positivem Lymphomgewebe in der End-of-Treatment-PET*

100 % der Patienten des Subkollektivs waren älter als 60 Jahre, da die OPTIMAL>60-Studie nur Patienten über 60 Jahre einschließt. Der Mittelwert der LDH (n=205) betrug 348.72 U/l und lag bei mehr als der Hälfte der Patienten oberhalb des Normwertes. Der ECOG im ausgewählten

Patientenkollektiv reichte von 0 bis 3. Von den 230 dokumentierten ECOG-Werten wies die Mehrzahl der Patienten (n=112) einen ECOG von 0 auf. Bei weniger als 40.0 % der Patienten wurden extralymphatische Manifestationen in der initialen Bildgebung festgestellt. Diese fanden sich am häufigsten im knöchernen Skelett (13.2 %), im Muskel-, Fett- oder Bindegewebe (12.4 %) und im Magen (6.6 %). Die Einteilung in die vier Krankheitsstadien erfolgte nach Ann-Arbour, wobei sich die meisten Patienten (n=93) im Stadium IV befanden. Zur Abschätzung der Prognose wurde der IPI berechnet, welcher zwischen 0 und 5 Punkten liegen kann und sich aus den eben genannten fünf Kriterien zusammensetzt. Für die Mehrzahl der Patienten (30.7 %) lag der IPI bei 3 Punkten. Zusätzlich wurden auch Lymphomherde in Form eines Bulks, der mindestens 7.5 cm misst, betrachtet. Bei weniger als 50.0 % konnte initial ein Bulk nachgewiesen werden. Zu den häufigsten Lokalisationen zählten die paraaortalen Lymphknoten (10.1 %) und die mesenterialen Lymphknoten (7.8 %) bzw. das Muskel-, Fett- oder Bindegewebe (7.8 %).

## 2.2 Konturierungsverfahren

Für die im Rahmen dieser Dissertation durchgeführte Analyse wurden die PET-Aufnahmen von 1050 der 1152 eingeschlossenen Patienten der OPTIMAL>60-Studie anhand des Deauville-Scores mit dem Analysewerkzeug *Multifokale Segmentierung* aus dem Software-Paket *syngo.via* von Siemens (43) hinsichtlich ihrer Tumorlast, gemessen als MTV, analysiert. Neben dem MTV wurden auch die TLG, das  $SUV_{max}$  und die Anzahl der ROIs der Lymphommanifestationen, sowie der  $SUV_{mean}$ , die zugehörige Standardabweichung und das  $SUV_{max}$  der Referenzregionen Leber und Aorta ermittelt.

Die semiautomatische Analyse erfolgte nach einem festgelegten Algorithmus: Radionuklidspeichernde Areale wurden nur ausgewählt, wenn ihr  $SUV_{max}$  über der sogenannten PERCIST-Schwelle (Positron Emission Response Criteria in Solid Tumors) lag, definiert als (44):

$$PERCIST = 1.5 \times SUV_{mean}^{Leber} + 2 \times SUV_{SD}^{Leber}$$

Dazu wurden, in den meisten Fällen automatisch, zuerst die Referenzregionen in der Aorta und der Leber festgelegt. Bei den Datensätzen, bei welchen aufgrund technischer Fehler die automatische Auswahl der Referenzregionen nicht möglich oder die automatische Platzierung jener dezentral oder nicht im jeweiligen Organ erfolgte, wurde eine manuelle Korrektur bzw. Platzierung durchgeführt.

Danach wurden die Areale, deren  $SUV_{max}$  über der PERCIST-Schwelle gelegen hatte, von der Software konturiert, indem gemäß der Empfehlungen der Europäischen Gesellschaft für Nuklearmedizin (EANM) ein Schwellenwert von 41 % des läsionalen SUV-Maximums zur Konturierung des Lymphomherdes diente (36). Da auch gesunde Areale, die physiologisch FDG speichern, durch den Algorithmus markiert wurden, mussten diese nach geschulter Bewertung manuell wieder gelöscht werden, um eine Verfälschung des Ergebnisses durch falsch-positive Befunde zu vermeiden. Vom Algorithmus nicht erkannte oder (neben physiologisch speichernden Arealen) fehlkonturierte Herde mussten entsprechend ergänzend konturiert werden. Dazu diente das Werkzeug *Isocontour VOI* mit einer Konturierung anhand von 41 % des läsionalen  $SUV_{max}$ .

Neben der PERCIST-Schwelle war auch das Mindestvolumen der Lymphommanifestationen für die Analyse von Bedeutung. Lymphomherde wurden als PET-positiv eingeordnet, sofern ihr Volumen mindestens 1 ml betragen hatte. Zusätzlich erfolgte pro Patient auch eine Analyse unter Einschluss aller Läsionen, deren  $SUV_{max}$  oberhalb der PERCIST-Schwelle gelegen hatte, unabhängig von dem Volumengrenzwert von 1 ml.

Die Dokumentation der ausgegebenen Zahlenwerte erfolgte handschriftlich auf Case Report Forms und wurde durch Bild-Ausdrucke der entsprechenden Lymphomherde ergänzt. Zusätzlich wurden die Werte in Form von Excel-Tabellen digitalisiert.

Bei 20 PET-Datensätzen, die aus technischen Gründen durch das Programm Siemens syngo.via nicht zu verarbeiten waren, erfolgte mit den gleichen Konturierungskriterien die Messung der genannten Parameter mittels der Programme ROVER (Version 3.0.58h, Firma ABX GmbH, Radeberg) oder Osirix MD (Version 6.0, Firma pixmeo, Bernex, Schweiz).

### 2.3 Inter-Observer-Vergleich

Ein Subkollektiv von 83 der 1050 in die Analyse eingeschlossenen Patienten der OPTIMAL>60-Studie wurden zur Beurteilung der Inter-Observer-Reliabilität einer Interrater-Reliabilitätsanalyse zugeführt und von zwei unabhängigen Betrachtern analysiert. Die ersten 40 der 83 Patienten wurden zufällig ausgewählt. Die einzige Bedingung war, dass zusätzlich ein initialer Staging-PET-Datensatz vorliegen musste. Da von diesen zuerst ausgewählten Patienten nur weniger als ein Viertel PET-positive Lymphomherde aufgewiesen hatten, wurde das Subkollektiv um 43 zufällig nach folgenden Kriterien ausgewählte Patienten erweitert:

1. PET-Positivität in der Referenzbefundung der Restaging-PET. Als PET-positiv wurden dabei all jene Lymphomherde gezählt, die visuell mindestens eine höhere Tracer-Aufnahme als im Blut der thorakalen Aorta zeigten.
2. Die Bewertung der PET-Bilder durch die Autorin dieser Doktorarbeit lautete positiv für vitales Lymphomgewebe. Unter Anwendung des quantitativen SUV-basierten Deauville-Scores wurden all jene Läsionen als positiv gewertet, deren Radionuklidspeicherung höher war als die der Leber, entspricht DS 4 und DS 5.
3. Der Datensatz der initial durchgeführten Staging-PET war verfügbar.

Die Analyse der selektierten Patienten erfolgte nach dem gleichen bereits weiter oben beschriebenen semi-automatischen Algorithmus mithilfe des Analysewerkzeugs *Multifokale Segmentierung* aus dem Software-Paket *syngo.via* von Siemens (43) unter Berücksichtigung der PERCIST-Schwelle, der Konturierung nach 41 % des läsionalen  $SUV_{max}$  der Lymphommanifestationen, der notwendigen manuellen Konturierung oder Löschung von physiologisch speichernden Herden und einer Volumengrenze von 1 ml für vitales Lymphomgewebe. Zusätzlich erhob jeder Observer pro Patient alle Läsionen, deren  $SUV_{max}$  oberhalb der PERCIST-Schwelle gelegen hatte, unabhängig von dem Volumengrenzwert von 1 ml. Folgende Parameter wurden zur Beurteilung der Tumorlast ermittelt: MTV, TLG,  $SUV_{max}$ , Anzahl der ROIs und SUVs der Referenzregionen Aorta und Leber.

Eine handschriftliche Dokumentation der ermittelten Zahlenwerte erfolgte auf Case Report Forms und wurde durch Bild-Ausdrucke der entsprechenden Lymphomherde vervollständigt. Auch digital wurden die Werte in Form von Excel-Tabellen festgehalten.

#### 2.4 Statistische Analyse

Die statistischen Analysen wurden unter Verwendung der Programme IBM SPSS Statistics Version 29.0, RStudio (Version 2023.06.2 + 561) und R (Version 4.3.1) durchgeführt. Zur Erhebung eines möglicherweise signifikanten Unterschieds in der Beurteilung des PET-Status unter Anwendung der visuellen Analysemethode, eingesetzt in den Referenzbefunden der OPTIMAL>60-Studie, und der Analyse mithilfe des quantitativen SUV-basierten Deauville-Scores wurde der McNemar-Test eingesetzt. Der McNemar-Test eignet sich, um eine zweifach gestufte Variable in einer Messwiederholung zu überprüfen (45).

Anschließend wurden deskriptive Statistiken berechnet, um die Verteilung der untersuchten Variablen MTV, TLG, Anzahl der ROIs und  $SUV_{max}$  der Lymphommanifestation darzustellen. Dabei wurde hinsichtlich der Volumina der Lymphomherde unterschieden und es wurden

sowohl Berechnungen für Läsionen größer oder gleich 1 ml als auch Analysen unter Einschluss aller Läsionen unabhängig von den Volumina durchgeführt. Berechnet wurden Mittelwerte, Standardabweichungen, Mediane und Häufigkeitsverteilungen der oben genannten Variablen. Zusätzlich wurde die Verteilung der Ergebnisse für Läsionen ohne Volumengrenze mithilfe von Histogrammen dargestellt und hinsichtlich Normalverteilung untersucht. Außerdem wurde die Positivitätsrate ermittelt. Diese ergibt sich aus dem Anteil an Patienten mit PET-positiven Lymphomherden geteilt durch die Gesamtheit der analysierten Patienten. Die Positivitätsrate wurde für Läsionen größer oder gleich 1 ml als auch für die Gesamtheit an Läsionen unabhängig von den Volumina berechnet.

Zur Überprüfung auf einen möglicherweise statistisch signifikanten Unterschied der MTV- bzw. TLG-Ergebnisse unter Berücksichtigung der Volumengrenze von 1 ml bzw. dem Einschluss von Lymphomläsionen jeglicher Volumina wurde nach Ausschluss der Normalverteilung der Differenzen der genannten Variablen durch den Shapiro-Wilk-Test der Wilcoxon Signed-Rank Test durchgeführt. Letzterer dient der Überprüfung von zwei abhängigen Stichproben mit nicht-normalverteilten Variablen auf Ordinalskalenniveau (45).

Zur Beurteilung der Übereinstimmung der im Rahmen des Inter-Observer-Vergleichs erhobenen Messergebnisse für das MTV erfolgte eine Analyse nach Bland und Altman. Diese ermöglicht die Darstellung systematischer Fehler der MTV-Messung und die Einschätzung der Streuung der Differenzen, um die Zuverlässigkeit der MTV-Messungen der Beobachter beurteilen zu können. Im Rahmen der Bland-Altman-Analyse wurden die Differenzen zwischen den MTV-Messungen jedes Beobachters für jedes Messpaar berechnet und die Berechnung der Mittelwerte durchgeführt. Im Bland-Altman-Plot wurden die Differenzen gegen die Mittelwerte aufgetragen, sowie die mittlere Differenz, auch Bias genannt, und die 95 %-Übereinstimmungsgrenzen dargestellt. Der Bias zeigt die systematische Abweichung zwischen den Beobachtern an, während die Übereinstimmungsgrenzen den Bereich definieren, in dem 95 % der Differenzen liegen sollten (46).

Die genannten Bland-Altman-Analysen wurden unter anderem für das MTV unter Einschluss aller Lymphomläsionen unabhängig der erhobenen Volumengröße durchgeführt. Auch die TLG-Ergebnisse für ROIs ohne Volumenschwelle, die  $SUV_{max}$  der Lymphomherde und die SUVs der Referenzorgane Aorta und Leber der beiden Observer wurden mittels Bland-Altman-Analyse ausgewertet.

Zusätzlich zu den Analysen nach Bland und Altman wurden Pairs Plots, unter Einschluss von Histogrammen, Scatterplots und der Berechnung des Korrelationkoeffizienten Kendall's Tau eingesetzt. Diese dienten der Analyse von MTV, TLG,  $SUV_{max}$  der Lymphomläsionen und  $SUV_{max}$  und  $SUV_{mean}$  der Referenzregionen Leber und Aorta, sowie der Untersuchung der PERCIST-Schwelle, Anzahl der ROIs und Konturierungszeiten der beiden Beobachter. Die Histogramme zeigen die graphische Verteilung der Messungen jedes Beobachters und ermöglichen es, Ähnlichkeiten bzw. Abweichungen in der Verteilung darzustellen. Die Scatterplots dienen der Visualisierung der Beziehung zwischen den Messungen der Beobachter. Der Korrelationskoeffizient Kendall's Tau, besonders geeignet für ordinal skalierte Variablen und robust gegen Ausreißer, quantifiziert die Stärke der Übereinstimmung zwischen den Messergebnissen der zwei Beobachter (45).

Daran wurde eine Interklassenkorrelationsanalyse angeschlossen, um die Inter-Observer-Reliabilität zu überprüfen, welche sowohl die Korrelation als auch die Übereinstimmung zwischen den Messergebnissen des MTV und der TLG sowie des  $SUV_{max}$  der Lymphomläsionen der beiden Beobachter bei Läsionen ohne Volumengrenze analysiert. Dazu wurden der Interklassenkorrelationskoeffizient (ICC(3,1)) und das 95 %-Konfidenzintervall berechnet, basierend auf einem zweifaktoriellen Mischmodell (two-way mixed effects model), Einzelwerten (single measurement) als Datengrundlage und der Überprüfung auf Übereinstimmung (absolute agreement) (47).

Zur Untersuchung der Beziehung der FDG-Anreicherung in der Leber und des MTV wurde nach Ausschluss der Normalverteilung der genannten Variablen durch den Shapiro-Wilk-Test der Kendall's Tau Korrelationskoeffizient bestimmt, um die Stärke und Richtung des Zusammenhangs zu überprüfen. Zusätzlich wurde eine lineare Regression durchgeführt, um das Ausmaß der Beziehung zu quantifizieren und mögliche Vorhersagen zuzulassen (45). Zur Überprüfung der Annahmen der linearen Regression wurde unter anderem der Durbin-Watson-Test durchgeführt, um die Unabhängigkeit der Residuen, der Abweichungen der tatsächlichen Werte von den durch das Modell vorhergesagten Werten, zu überprüfen. Die Unabhängigkeit der Residuen bedeutet, dass die Abweichungen nicht korreliert sind, also die Abweichung eines bestimmten Datenpunkts keine systematische Beziehung zu einem anderen Datenpunkt hat. Eine Verletzung dieser Annahme kann zu verzerrten Schätzungen der Standardfehler führen, was die Signifikanztests ungültig machen kann (48). Daraufhin wurde die Homoskedastizität untersucht. Bei Homoskedastizität spricht man von einer konstanten Varianz der Residuen für alle Werte der unabhängigen Variablen. Variiert die Varianz der Residuen, Heteroskedastizität

genannt, kann dies die Schätzung und die Signifikanz des Regressionskoeffizienten negativ beeinflussen. Die Homoskedastizität wurde mithilfe des Breusch-Pagan-Tests kontrolliert. Zur Überprüfung der Normalverteilung der Residuen wurde ein Histogramm mit Normalverteilungskurve erstellt. Die Normalverteilung der Residuen ist eine Voraussetzung für zuverlässige Signifikanztests. Die Berechnung der Cook's Distance dient der Identifizierung potentiell einflussreicher Datenpunkte, die die Regressionsanalyse stark verzerren, indem sie die Schätzungen des Regressionskoeffizienten unverhältnismäßig beeinflussen (49).

### 3. Ergebnisse

Dieser Abschnitt beinhaltet die Präsentation der Ergebnisse der Gegenüberstellung der visuellen Analysemethode und des quantitativen SUV-basierten Deauville-Scores, sowie die Ergebnisdarstellung der deskriptiven Statistik und des Vergleichs der Volumengrenzen. Darauffolgend werden die Auswertung des Inter-Observer-Vergleichs und die Resultate der Untersuchung hinsichtlich eines Tumor-Sink-Effekts beschrieben.

#### 3.1 Gegenüberstellung der visuellen Analysemethode und des Deauville-Scores

Der McNemar-Test (n=1049) zeigte einen statistisch signifikanten Unterschied ( $p<0.001$ ) in der Beurteilung des PET-Status bei Anwendung des quantitativen SUV-basierten Deauville-Scores unter Einschluss von Läsionen aller Volumina bzw. der visuellen Analysemethode, eingesetzt in den Referenzbefunden der OPTIMAL>60-Studie. Ein Patient musste ausgeschlossen werden, da in der Referenzbefundung der PET-Status jenes Patienten als *unbekannt* definiert worden war. Somit war die Dichotomie (*positiv, negativ*) der Variablen nicht mehr gegeben, und der McNemar-Test war auf diesen Patienten nicht anwendbar. Mithilfe des Deauville-Scores wurden 818 Patienten als PET-negativ und 231 als PET-positiv bewertet. Unter Anwendung des visuellen Analyseverfahrens waren es 623 Patienten mit PET-Bewertung *negativ* und 426 Patienten mit PET-Bewertung *positiv*. Eine ausführliche Darstellung der PET-Bewertungen mithilfe der beiden genannten Verfahren ermöglicht Tabelle 3.

		PET-Bewertung visuell		
		negativ	positiv	Gesamt
PET-Bewertung	negativ	619	199	818
	positiv	4	227	231
Deauville	Gesamt	623	426	1049

**Tabelle 3:** Vergleich der PET-Bewertungen, gewonnen nach visuellem Analyseverfahren und mithilfe des Deauville-Scores (DS). PET-Bewertung Deauville, basierend auf Vergleich läsionales SUV-Maximum gegenüber Referenzorganen Blutpool bzw. Leber: positiv bei DS 4 und 5, negativ bei DS 1, 2 und 3; PET-Bewertung visuell: Einschätzung im Vergleich zum mediastinalen Blutpool. Positiv entspricht läsionaler Speicherintensität über, negativ unter mediastinalem Blutpool.

Von den 818 Patienten, die mithilfe des Deauville-Scores als PET-negativ bewertet worden waren, wurden nach visueller Analysemethode etwa drei Viertel der gleichen Patienten auch als PET-negativ angesehen. Die übrigen 199 Patienten wurden als PET-positiv dokumentiert. Bei Betrachtung der als PET-positiv gewerteten Patienten zeigte sich hinsichtlich der beiden Analysemethoden, visuell und Deauville-Score, eine Übereinstimmung von über 50 %. Vier Patienten, die mithilfe des Deauville-Scores als PET-positiv bewertet wurden, erhielten nach visueller Analysemethode die PET-Bewertung *negativ*.

Auffallend sind vier Patientenfälle, welche unter Berücksichtigung des Deauville-Scores mit der PET-Bewertung *positiv* eingestuft wurden, unter Anwendung des visuellen Auswerteverfahren jedoch als *negativ* bewertet wurden. Der erste jener vier Patienten zeigte eine abgrenzbare FDG-Anreicherung im Lungengewebe, welche unter Berücksichtigung des Deauville-Scores als PET-positiv und in den Referenzbefunden unter Einsatz des visuellen Bewertungsverfahren als entzündlich und somit PET-negativ gewertet wurde. Der zweite Patient zeigte nach Bewertung mithilfe des Deauville-Scores eine PET-positive Radionuklidspeicherung im Darm, welche im Rahmen der visuellen Befundung als PET-negativ mit Empfehlung einer Verlaufskontrolle eingestuft wurde. Bei einem weiteren Patienten ergab sich eine differierende Bewertung des PET-Status. Dieser wurde unter Berücksichtigung des Deauville-Scores mit einer FDG-speichernden Läsion mediastinal als *positiv* gewertet, wohingegen der Patient nach Einsatz des visuellen Befundungsverfahrens die PET-Bewertung *negativ* erhielt. Der letzte der vier abweichend beurteilten Patienten verzeichnete nach Beurteilung des Deauville-Scores eine PET-positive Läsion an mesenterialer Lokalisation, welche nach visuellem Auswerteverfahren als PET-negativ gewertet wurde.

### 3.2 Deskriptive Statistik

Die Positivitätsrate unter Einschluss von Läsionen mit einem Volumen von mindestens 1 ml lautete 21.3 %, unter Berücksichtigung von allen Lymphomherden unabhängig der Läsionsgröße betrug sie 22.0 %.

#### 3.2.1 Berechnungen für Läsionen mit einem Mindestvolumen von 1 ml

Einen Überblick über die Ergebnisse der MTV-, TLG-, SUV<sub>max</sub>- und ROI-Analyse unter Beachtung der Läsionen mit mindestens 1 ml Volumen gibt Tabelle 4.

	N	Mittelwert $\pm$ SD	Median (Min-Max)
MTV-Last [ml]	224	31.4 $\pm$ 53.5	12.8 (1.0-356.6)
TLG-Last	224	156.7 $\pm$ 417.3	48.3 (2.9-3595.1)
SUV <sub>max</sub>	224	8.3 $\pm$ 6.3	6.0 (2.3-36.4)
ROIs	224	2.6 $\pm$ 4.7	1.0 (1.0-60.0)

*Tabelle 4: Deskriptive Statistik für MTV, TLG, SUV<sub>max</sub> und die Anzahl der ROIs unter Berücksichtigung der Läsionen mit einem Mindestvolumen von 1 ml. SD: Standardabweichung*

### 3.2.2 Berechnungen für Läsionen aller Volumina

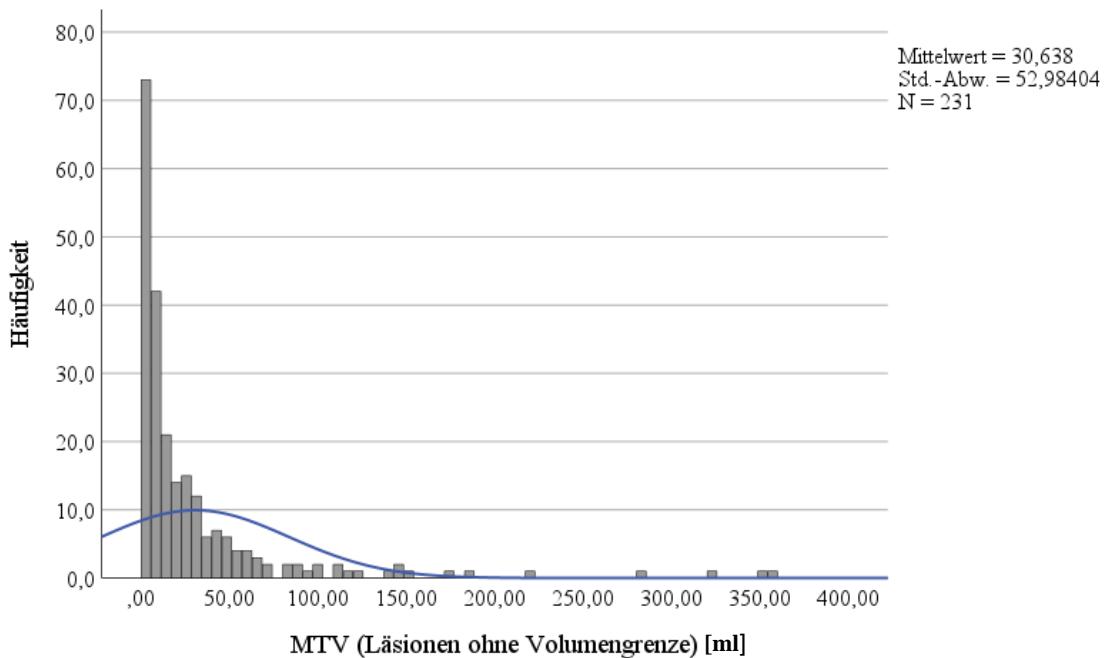
Eine tabellarische Übersicht über die Ergebnisse der MTV-, TLG-, SUV<sub>max</sub>- und ROI-Analyse unter Beachtung von Läsionen aller Volumina ermöglicht Tabelle 5.

	N	Mittelwert $\pm$ SD	Median (Min-Max)
MTV-Last [ml]	231	30.6 $\pm$ 53.0	11.8 (0.3-356.6)
TLG-Last	231	153.4 $\pm$ 412.8	48.0 (1.0-3595.1)
SUV <sub>max</sub>	231	8.7 $\pm$ 7.0	6.1 (2.3-37.5)
ROIs	231	2.8 $\pm$ 5.1	1.0 (1.0-60.0)

*Tabelle 5: Deskriptive Statistik für MTV, TLG, SUV<sub>max</sub> und die Anzahl der ROIs unter Berücksichtigung von Läsionen aller Volumina. SD: Standardabweichung*

Die Normalverteilung der logarithmierten MTV- und TLG-Werte wird nachfolgend graphisch dargestellt und auch durch den Shapiro-Wilk-Test bestätigt (MTV: p=0.575, TLG: p=0.111).

Die Häufigkeitsverteilung der MTV-Werte wird im untenstehenden Histogramm (Abbildung 2) veranschaulicht.



*Abbildung 2: Häufigkeitsverteilung der MTV-Werte bei Läsionen aller Volumina*

Nach dem Logarithmieren der MTV-Werte zeigte sich eine Normalverteilung der Ergebnisse, dargestellt in Abbildung 3.

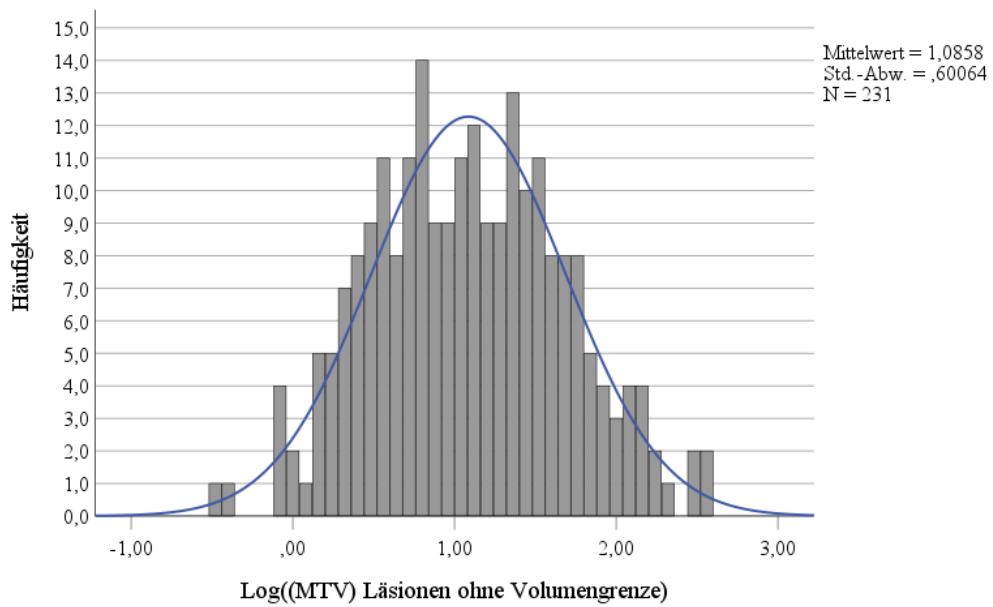


Abbildung 3: Häufigkeitsverteilung der logarithmierten MTV-Werte bei Läsionen aller Volumina

Das Histogramm (Abbildung 4) stellt die Verteilung der TLG-Werte für Läsionen ohne Volumengrenze graphisch dar.

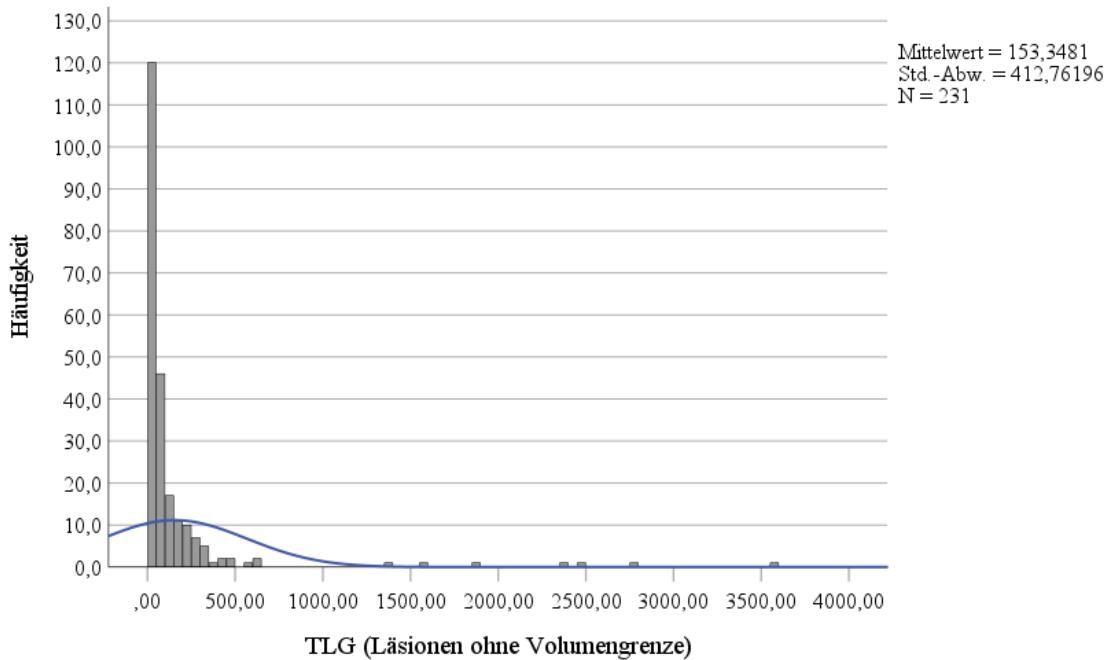


Abbildung 4: Häufigkeitsverteilung der TLG-Werte bei Läsionen aller Volumina

Nach Log-Transformation der TLG-Ergebnisse ergab sich eine Normalverteilung der Messwerte (Abbildung 5).

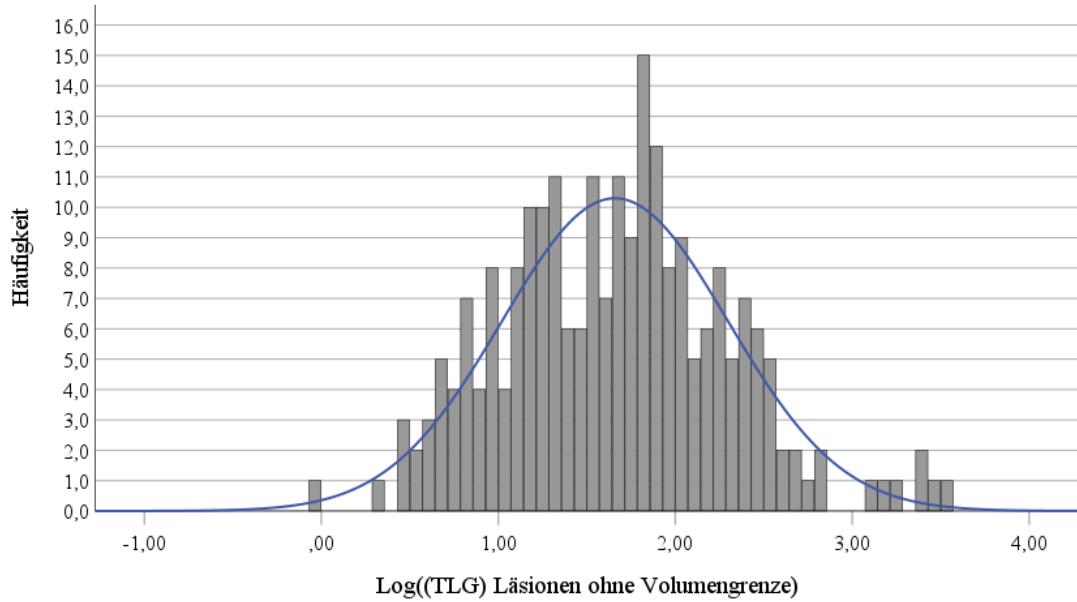


Abbildung 5: Häufigkeitsverteilung der logarithmierten TLG-Werte bei Läsionen aller Volumina

### 3.3 Vergleich der MTV- und TLG-Ergebnisse unter Berücksichtigung der verschiedenen Volumengrenzen

Der Shapiro-Wilk-Test ergab, dass die Differenzen zwischen den Ergebnissen der MTV-Analyse für 1 ml und alle Läsionsgrößen nicht normalverteilt sind ( $p<0.001$ ). Der angeschlossene Wilcoxon Signed-Rank Test zeigte, dass sich die gepaarten Werte der MTV-Erhebung (Läsionen aller Volumina gegenüber Läsionen mit einer Mindestgröße von 1 ml;  $n=1050$ ) statistisch signifikant voneinander unterscheiden ( $p<0.001$ ). Gleiches wurde für die statistische Auswertung der erhobenen TLG-Werte bewiesen. Nach Ausschluss der Normalverteilung der Differenzen zwischen den TLG-Werten für 1 ml und alle Läsionsgrößen ( $p<0.001$ ) verdeutlichte der Wilcoxon Signed-Rank Test, dass die gepaarten Werte der TLG-Erhebung (Läsionen aller Volumina gegenüber Läsionen mit einer Mindestgröße von 1 ml;  $n=1050$ ) statistisch signifikant voneinander abweichen ( $p<0.001$ ).

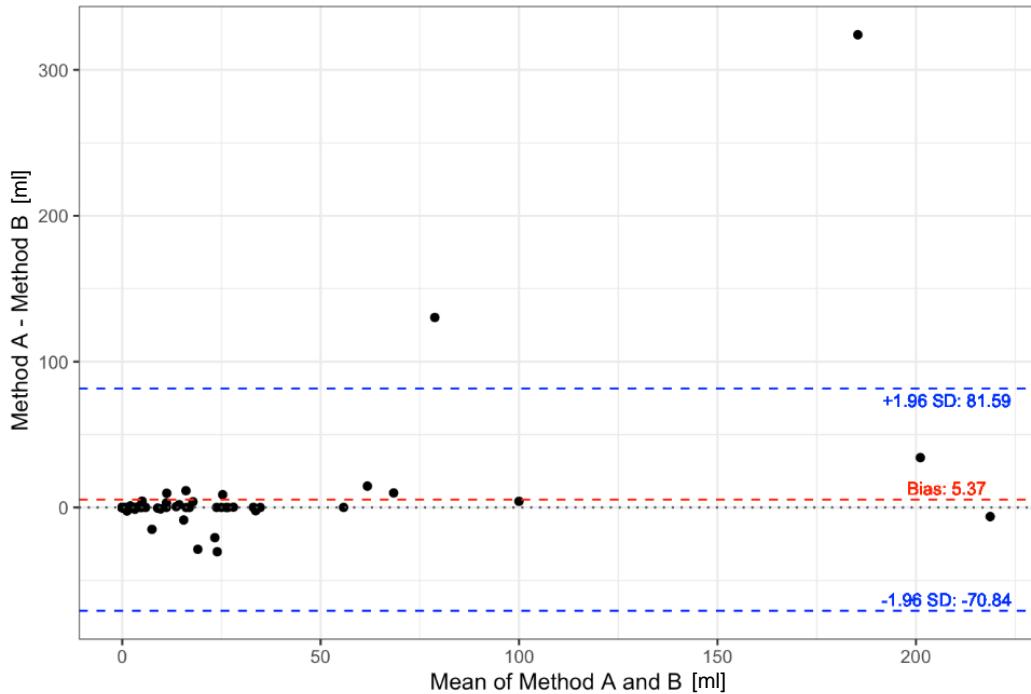
### 3.4 Inter-Observer-Vergleich

Der sich anschließende Abschnitt beinhaltet die Ergebnisse der Analyse nach Bland und Altman und der Untersuchung hinsichtlich der Verteilung und Korrelation der Messwerte beider Beobachter. Zusätzlich werden die Ergebnisse der Interklassenkorrelationsanalyse beschrieben.

#### 3.4.1 Analyse nach Bland und Altman

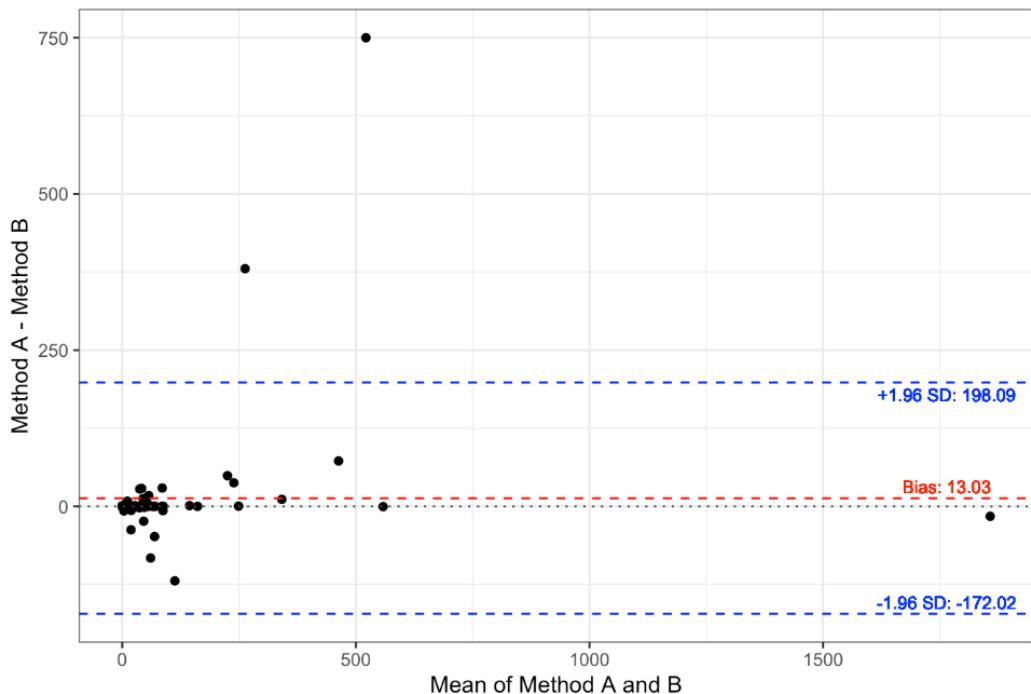
Die nachfolgenden Bland-Altman-Plots (Abbildungen 6 und 7) stellen die Übereinstimmung zwischen den Messergebnissen beider Beobachter für das MTV und die TLG bei Läsionen ohne Volumengrenze ( $n=83$ ) dar. Entlang der x-Achse sind die Mittelwerte der Messwerte der

Observer abgebildet, die y-Achse zeigt jeweils die Differenzen zwischen den beiden Messungen an.



**Abbildung 6:** Bland-Altman-Plot zur Analyse der Übereinstimmung beider Beobachter für die MTV-Werte aller ROI-Größen.  
SD: Standardabweichung

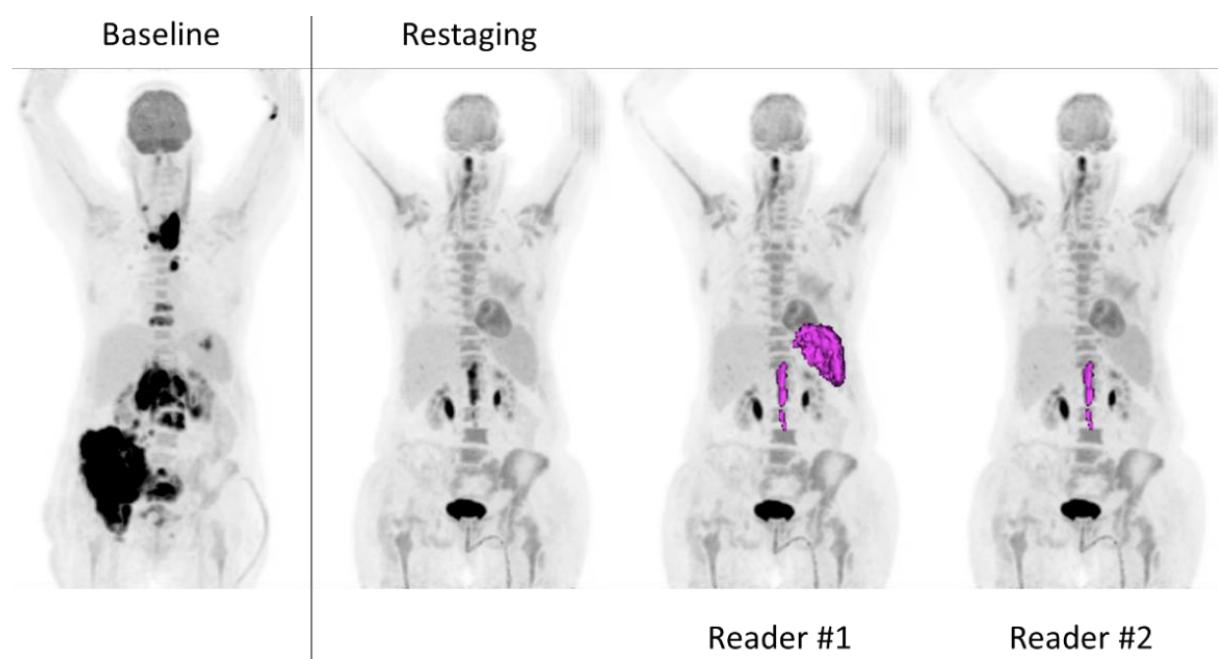
Die oberen und unteren Übereinstimmungsgrenzen, dargestellt durch die blauen, unterbrochenen Linien, lagen bei 81.59 bzw. -70.84 für das MTV und 198.09 bzw. -172.02 für die TLG.



**Abbildung 7:** Bland-Altman-Plot zur Analyse der Übereinstimmung beider Beobachter für die TLG-Werte aller ROI-Größen.  
SD: Standardabweichung

Verglichen mit den breiten Streubereichen, zeigte der Bias mit 5.37 für das MTV und 13.03 für die TLG eine geringe systematische Abweichung zwischen den beiden Beobachtern. Der große Streubereich wird durch die beiden Ausreißer, die Datenpunkte außerhalb der 95 %-Übereinstimmungsgrenzen, beeinflusst und verbreitert. Dabei handelt sich um zwei Patientenfälle, bei denen die Messungen der Beobachter deutlich voneinander abweichen. Diese werden im Folgenden näher beschrieben und in den Abbildungen 8 und 9 graphisch dargestellt.

Bei ersterem Patienten wurde von beiden Beobachtern eine Lymphommanifestation im ZNS gemessen, von nur einem Beobachter zusätzlich auch ein Milzbefall. Dadurch ergaben sich MTV-Ergebnisse von 23.3 ml und 347.4 ml und TLG-Werte von 146.7 und 896.6 durch die beiden Observer.



**Abbildung 8:** Maximum-Intensity-Projections (MIP) der prätherapeutischen FDG-PET und nach Induktionstherapie aus ventraler Sicht ohne und nach Konturierung durch zwei unabhängige Befürger. Nach initial extranodalen Manifestationen mit unter anderem fokalem Milzbefall und nach Induktionstherapie deutlich PET-positiven Resten intraspinal differieren die Bewertungen der diffus erhöhten Milzspeicherung.

Bei dem zweiten Patienten wurden beide Male paraaortale Lymphknoten als vitales Lymphomgewebe festgestellt, ein Beobachter hat zusätzlich infiltriertes Knochenmark im Becken ausgemessen. Die Messwerte für das MTV lauteten 13.62 ml und 143.90 ml, für die TLG 72.79 und 453.10.

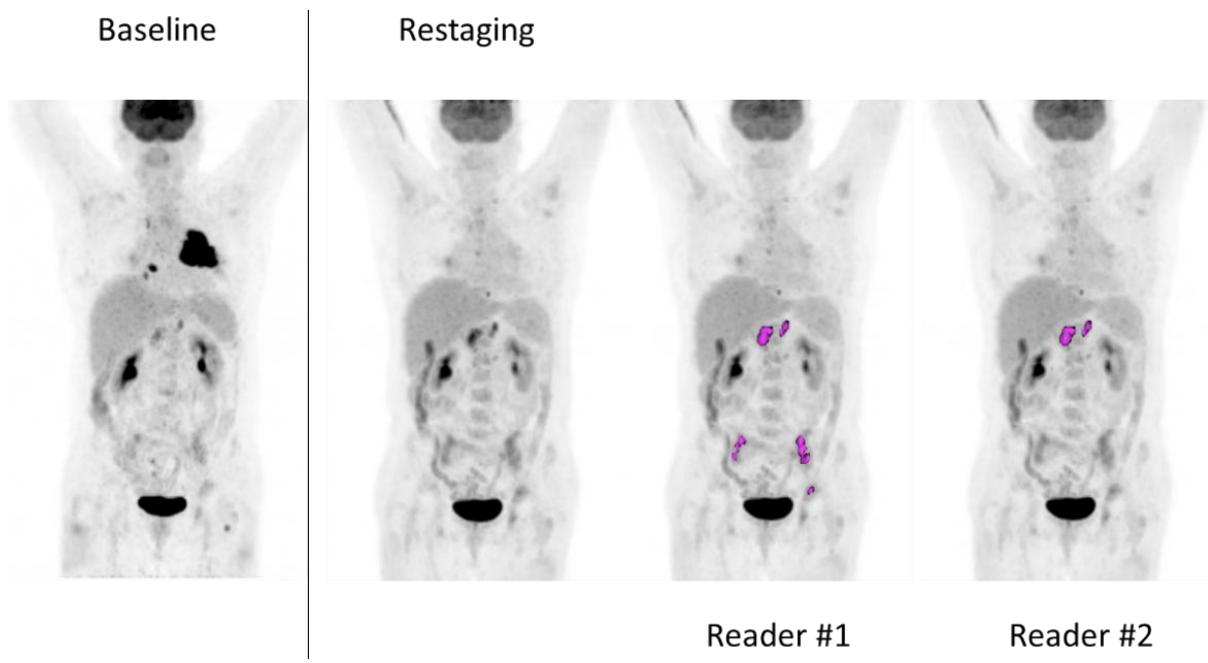


Abbildung 9: Maximum-Intensity-Projections (MIP) der prätherapeutischen FDG-PET und nach Induktionstherapie aus ventraler Sicht ohne und nach Konturierung durch zwei unabhängige Befunder. Nach initialem Lungenbefall und nach Induktionstherapie persistierenden nodalen Manifestationen differieren die Bewertungen der diffus erhöhten Knochenmarkspeicherung im Becken.

Abbildung 10 stellt die Übereinstimmung der  $SUV_{max}$  der Lymphomherde beider Beobachter für Patienten mit PET-Positivität bei allen Läsionsgrößen ( $n=50$ ) dar. Mit einem Bias von -0.02 und einem schmalen Streubereich präsentieren sich die Messwerte beider Beobachter größtenteils sehr konstant mit nur minimalen Abweichungen.

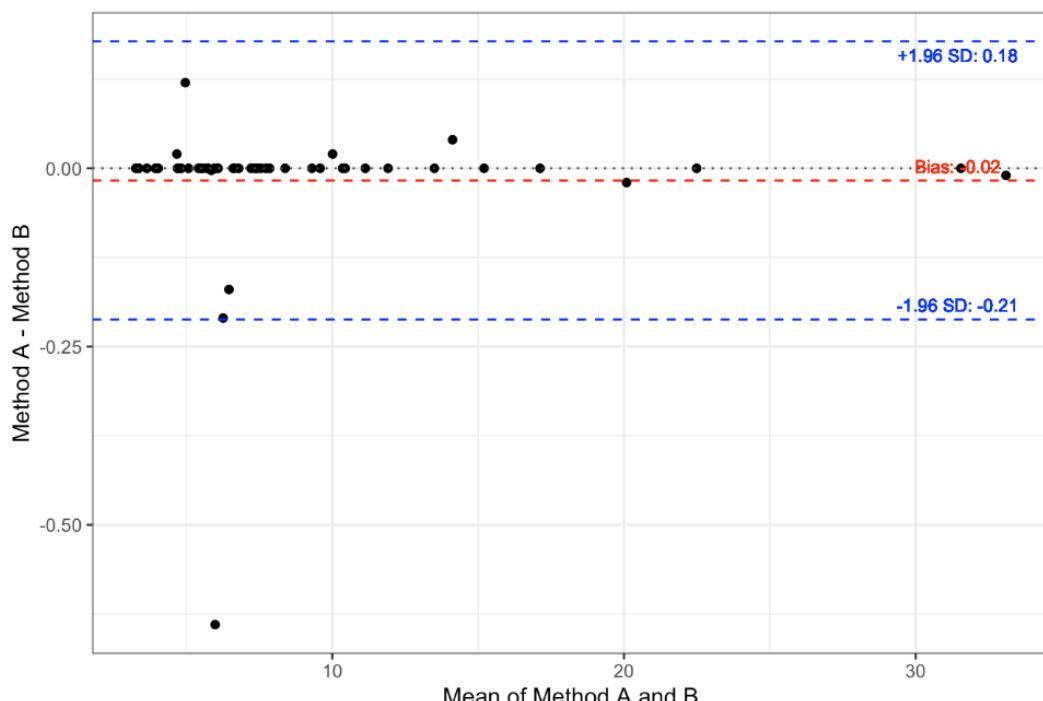
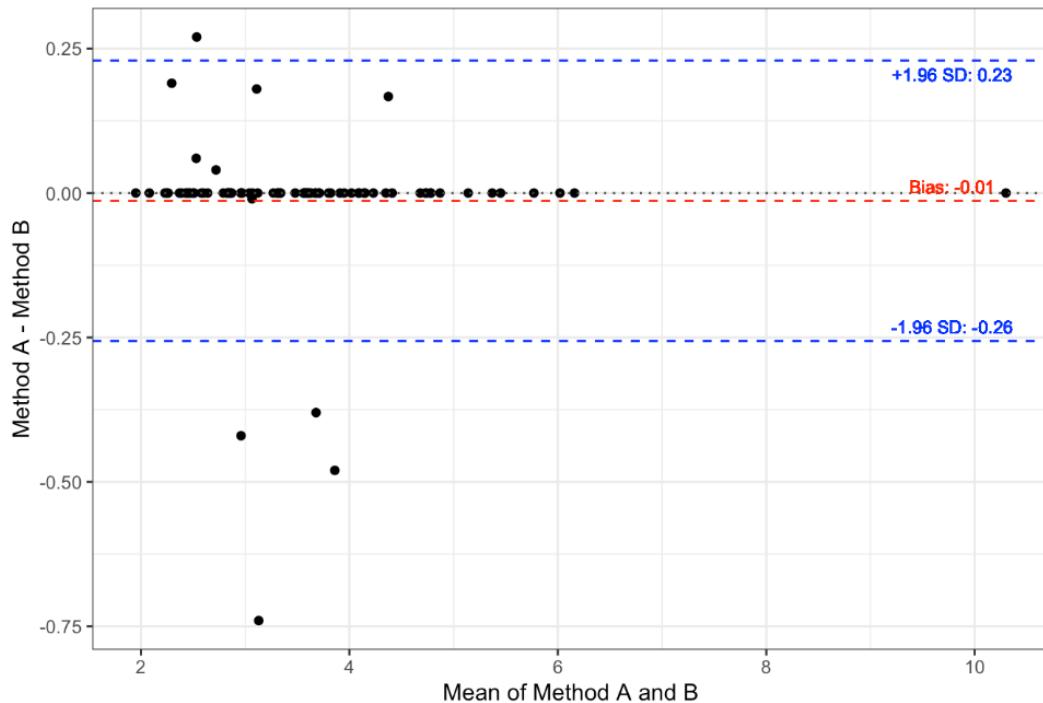


Abbildung 10: Bland-Altman-Plot zur Analyse der Übereinstimmung beider Beobachter für die  $SUV_{max}$  der Lymphomherde der Patienten mit PET-Positivität für alle ROI-Größen. SD: Standardabweichung

Der Bland-Altman-Plot (Abbildung 11) zur Darstellung der Übereinstimmung der  $SUV_{max}$  der Leber beider Beobachter ( $n=83$ ) veranschaulicht mit einem Bias von -0.01 und einem kleinen Streubereich eine geringe durchschnittliche Differenz zwischen den Messwerten der beiden Beobachter.



*Abbildung 11: Bland-Altman-Plot zur Analyse der Übereinstimmung beider Beobachter für die  $SUV_{max}$  der Leber. SD: Standardabweichung*

Gleiches gilt für die  $SUV_{max}$  der Aorta der beiden Beobachter, deren Übereinstimmung in Abbildung 12 präsentiert wird. Mit einem Bias von -0.01 und einem schmalen Streubereich zeigten sich auch hier konstante Messwerte mit geringen Unterschieden zwischen den beiden Beobachtern.

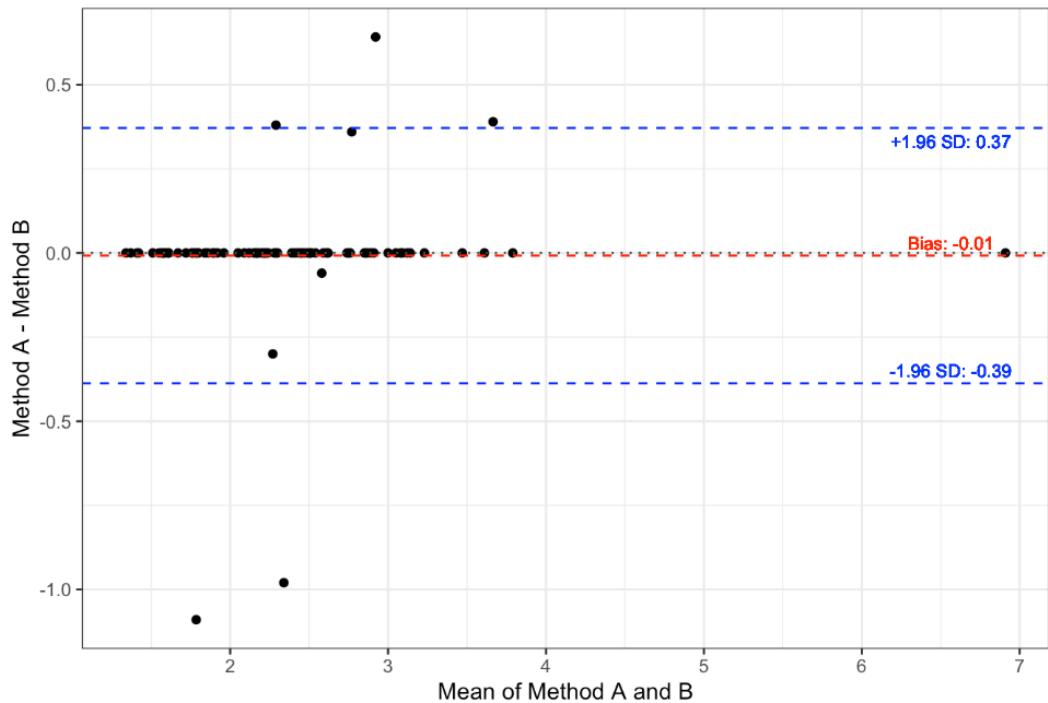


Abbildung 12: Bland-Altman-Plot zur Analyse der Übereinstimmung beider Beobachter für die  $SUV_{max}$  der Aorta. SD: Standardabweichung

Auch die Mittelwerte der Referenzregion Leber der beiden Beobachter, dargestellt in Abbildung 13, weisen mit einem Bias von -0.03 und einem kleinen Streubereich nur geringe Abweichungen auf.

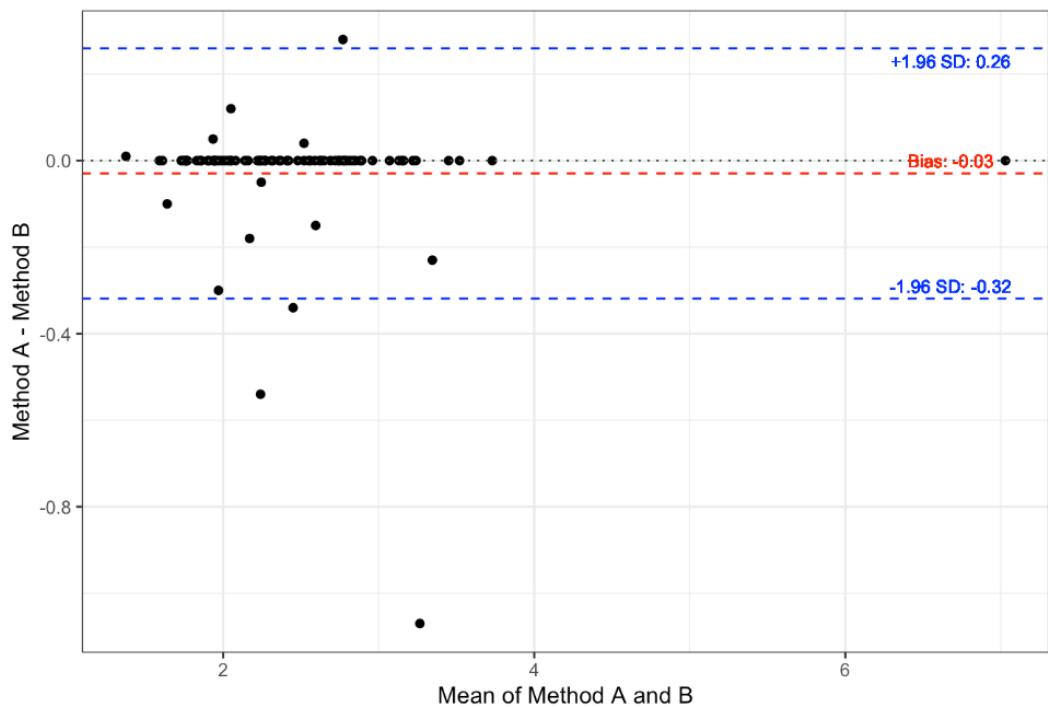
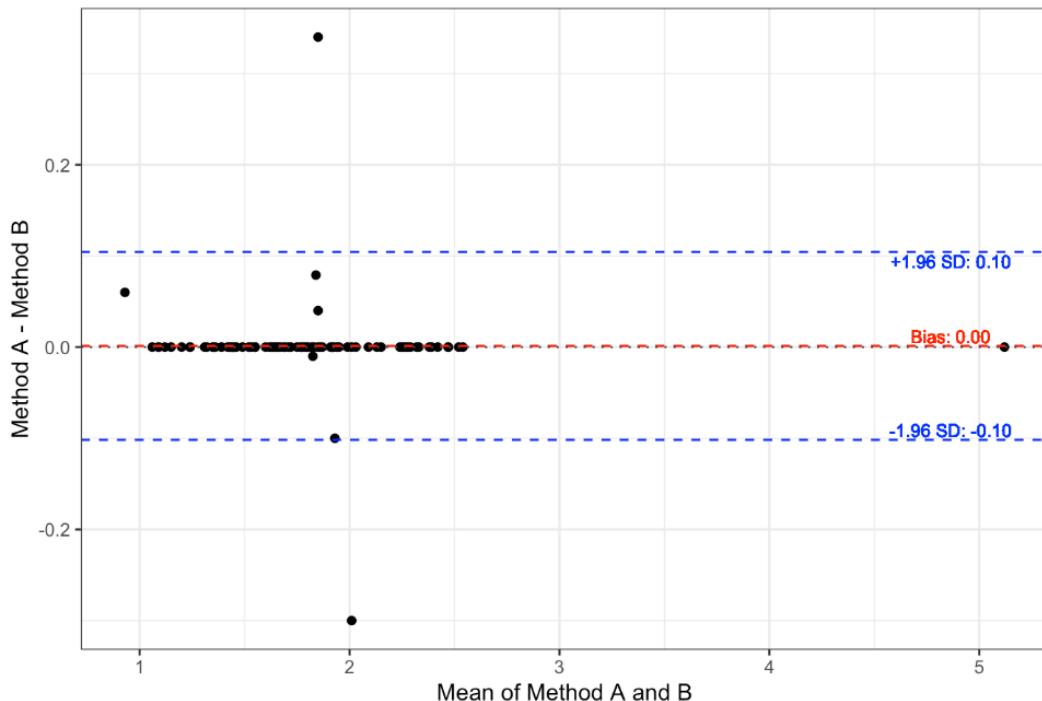


Abbildung 13: Bland-Altman-Plot zur Analyse der Übereinstimmung beider Beobachter für die  $SUV_{mean}$  der Leber. SD: Standardabweichung

Die Übereinstimmung der  $SUV_{mean}$  der Aorta beider Beobachter wird in Abbildung 14 dargestellt. Der Bias von 0.00 und der schmale Streubereich erklären die konstanten Messwerte und die marginalen Abweichungen zwischen den Beobachtern.



**Abbildung 14:** Bland-Altman-Plot zur Analyse der Übereinstimmung beider Beobachter für die  $SUV_{mean}$  der Aorta. SD: Standardabweichung

### 3.4.2 Analyse der Verteilung und Korrelation mithilfe von Scatterplots, Histogrammen und Kendall's Tau Korrelationskoeffizienten

Die sich anschließenden Pairsplots visualisieren die Übereinstimmung zwischen den Messungen der beiden Beobachter. Abbildung 15 stellt die Messergebnisse für das MTV unter Einschluss von Läsionen ohne Volumengrenze dar.

Die Histogramme visualisieren die ähnliche Verteilung der Messwerte der beiden Beobachter. Die Berechnung des Kendall's Tau Korrelationskoeffizienten für die MTV-Ergebnisse der beiden Beobachter ( $n=83$ ) ergab eine statistisch signifikante ( $p<0.001$ ) positive Korrelation mit einem Wert von 0.864. Der Scatterplot unterstreicht die positive Korrelation: Außer den in 3.4.1 beschriebenen Ausreißern orientieren sich alle Datenpunkte entlang der Trendlinie.

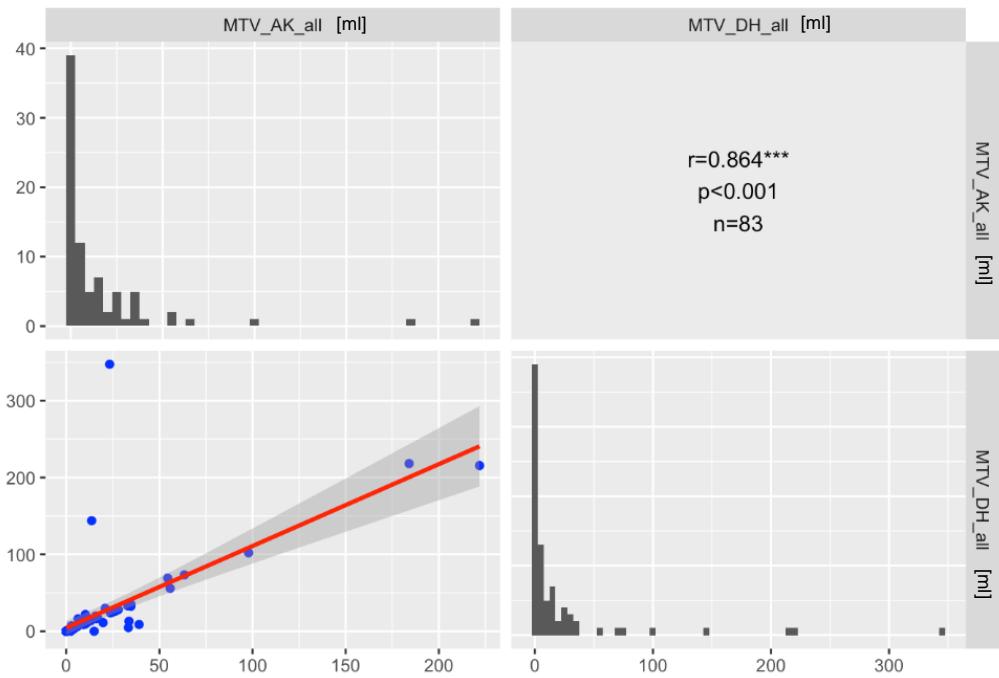


Abbildung 15: Pairsplot zur Darstellung der Beziehung und Verteilung der MTV-Ergebnisse unter Einschluss von Läsionen aller ROI-Größen beider Beobachter

Die sich anschließende Abbildung 16 zeigt die statistisch signifikante ( $p<0.001$ ) positive Korrelation  $r=0.771$  zwischen dem MTV und der Anzahl der ausgewählten Regionen eines Beobachters: Mit zunehmendem MTV steigt auch die Anzahl der ROIs.

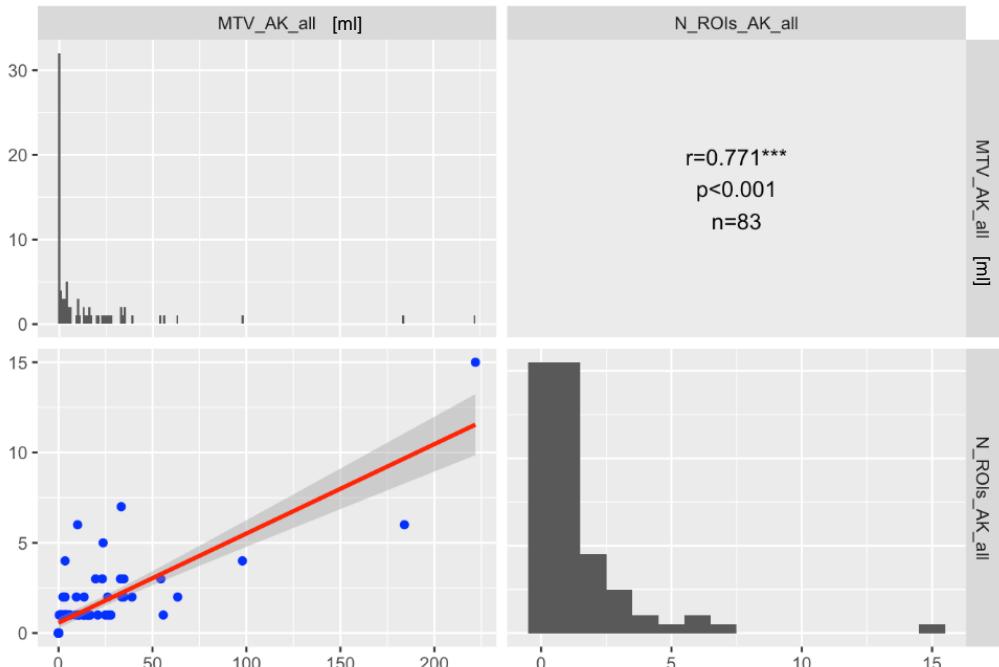
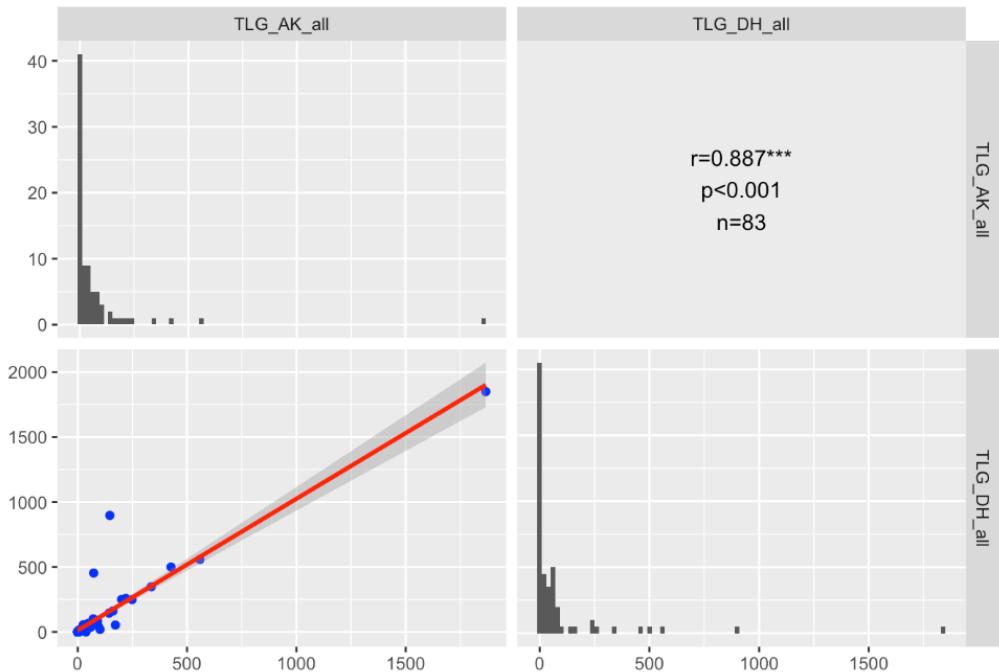


Abbildung 16: Pairsplot zur Darstellung der Beziehung der Anzahl der ROIs und der MTV-Ergebnisse unter Einschluss von Läsionen ohne Volumengrenze eines Beobachters

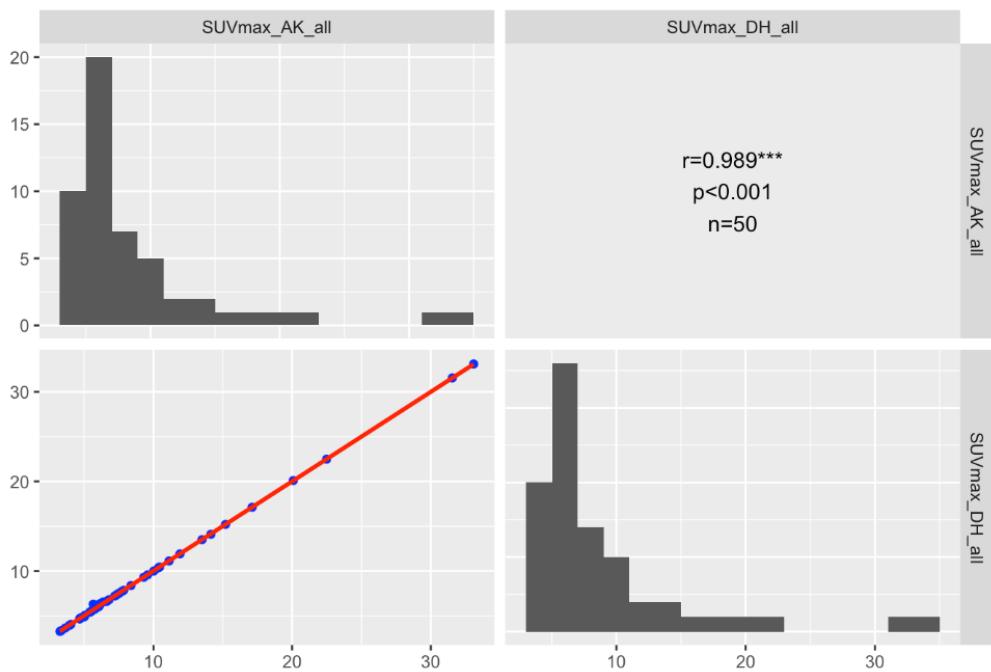
Die nachfolgende Abbildung 17 stellt die Beziehung und Verteilung der TLG-Werte der beiden Beobachter unter Berücksichtigung von Läsionen ohne Volumengrenze ( $n=83$ ) graphisch dar.

Die Histogramme zeigen die ähnliche Verteilung der Ergebnisse der beiden Beobachter. Die statistisch signifikante ( $p<0.001$ ) positive Korrelation ( $r=0.887$ ) wird durch den Scatterplot verdeutlicht: Abgesehen von den in 3.4.1 beschriebenen Ausreißern ordnen sich alle Datenpunkte entlang der Trendlinie an.



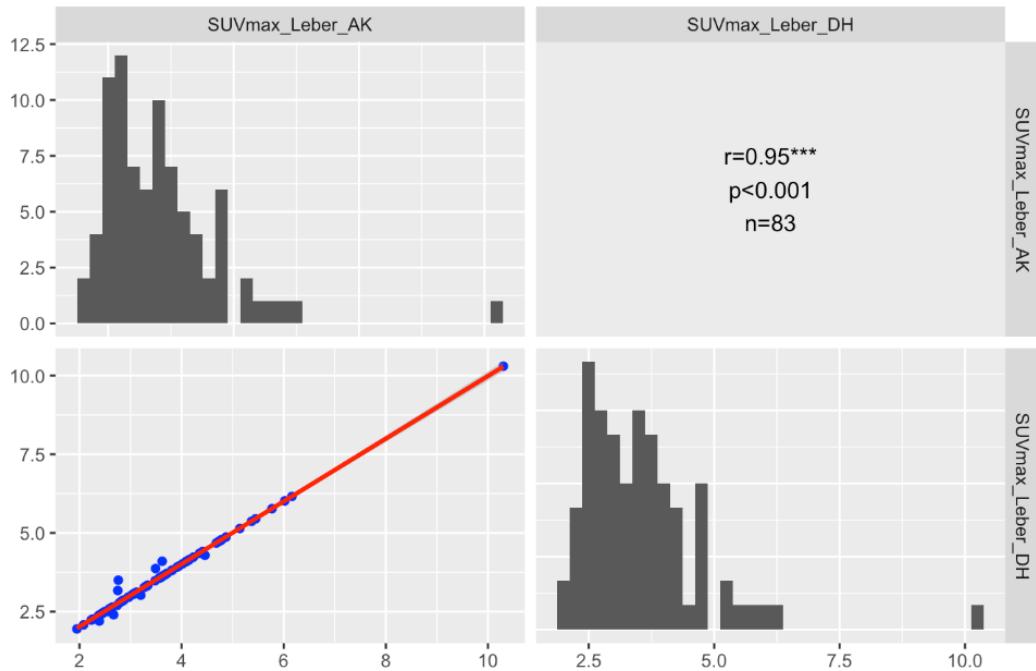
**Abbildung 17:** Pairsplot zur Darstellung der Beziehung und der Verteilung der TLG-Werte beider Beobachter unter Einschluss von Läsionen ohne Volumengrenze

Die Beziehung und Verteilung der  $SUV_{max}$  der Lymphomherde für Läsionen ohne Volumengrenze beider Beobachter ( $n=50$ ) ist in Abbildung 18 dargestellt. Die Histogramme zeigen die übereinstimmende Verteilung der Messwerte der beiden Beobachter, in den Scatterplots präsentieren sich die Datenpunkte ohne Ausreißer stringent entlang einer Geraden. Die beinahe vollständige Übereinstimmung wird auch durch den statistisch signifikanten ( $p<0.001$ ) Kendall's Tau Korrelationskoeffizienten ( $r=0.989$ ) unterstrichen.



**Abbildung 18:** Pairsplot zur Darstellung der Beziehung und Verteilung der  $SUV_{max}$  der Lymphomherde beider Beobachter unter Einschluss von Läsionen ohne Volumengrenze

Die Übereinstimmung zwischen den erhobenen  $SUV_{max}$  für die Referenzregionen Leber und Aorta ( $n=83$ ) ist ersichtlich in den Abbildungen 19 und 20. Die Histogramme stellen die übereinstimmende Verteilung der Messwerte der beiden Beobachter dar, in den Scatterplots orientieren sich die Datenpunkte ohne markante Ausreißer entlang der Trendlinie.



**Abbildung 19:** Pairsplot zur Darstellung der Beziehung und Verteilung der  $SUV_{max}$  der Leber beider Beobachter

Die positive Korrelation für die  $SUV_{max}$  der Referenzregionen wird durch die statistisch signifikanten ( $p<0.001$ ) Kendall's Tau Korrelationskoeffizienten ( $r=0.95$ ; Leber) und ( $r=0.907$ ; Aorta) unterstrichen.

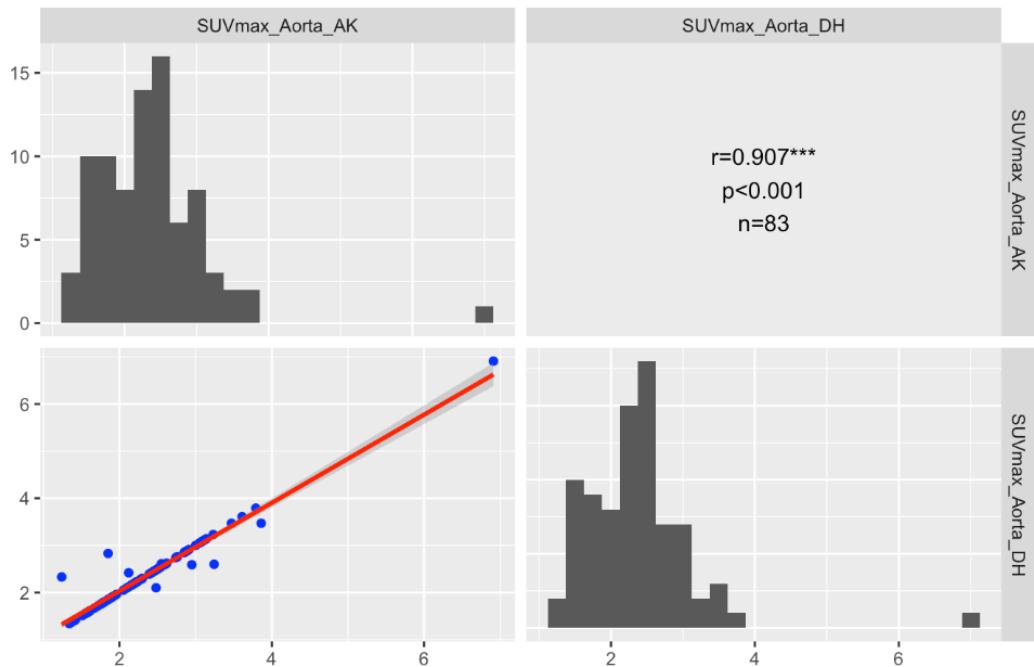


Abbildung 20: Pairsplot zur Darstellung der Beziehung und Verteilung der  $SUV_{max}$  der Aorta beider Beobachter

Gleiches gilt für die Übereinstimmung der  $SUV_{mean}$  der Leber und der Aorta der beiden Beobachter, dargestellt in Abbildung 21 und 22.

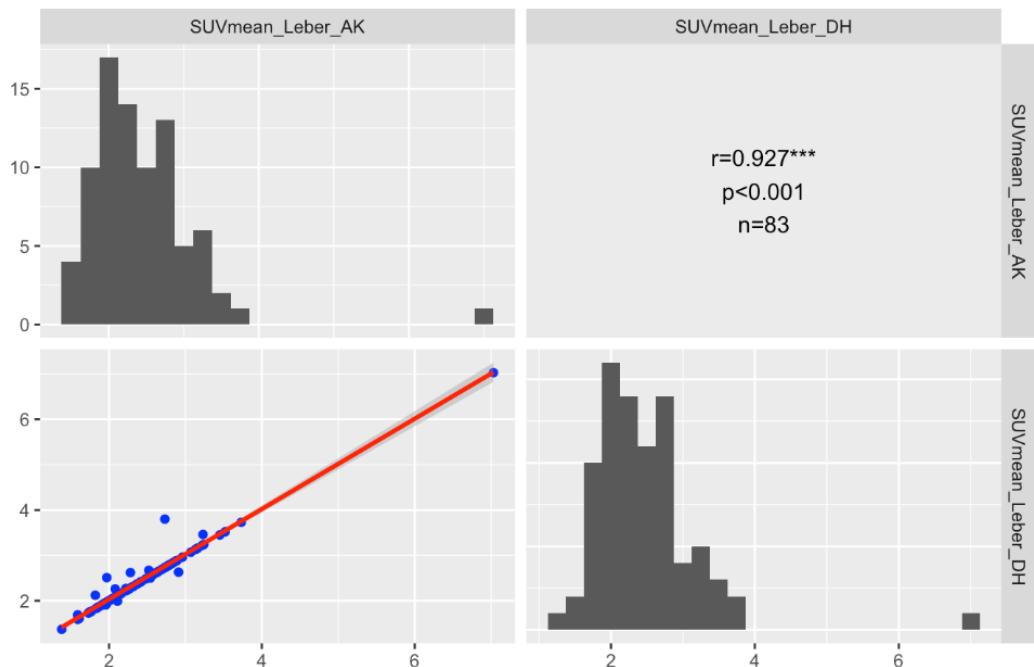
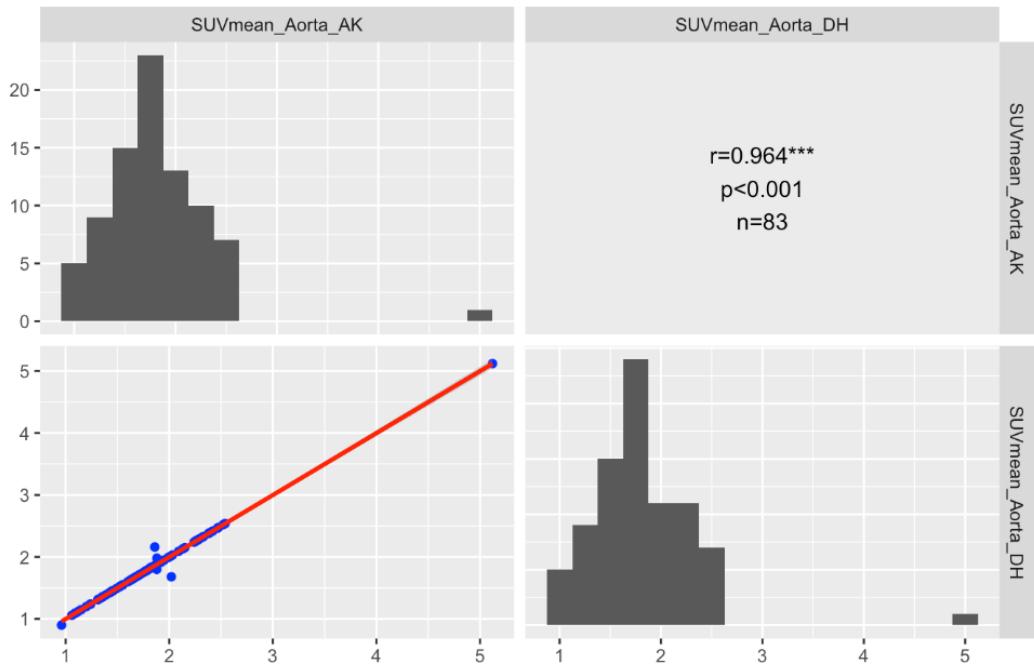


Abbildung 21: Pairsplot zur Darstellung der Beziehung und Verteilung der  $SUV_{mean}$  der Referenzregion Leber beider Beobachter

Auch hier zeigt sich eine ähnliche Verteilung der Messwerte der beiden Observer in den Histogrammen und eine positive Korrelation in den Scatterplots, welche durch die statistisch signifikanten ( $p<0.001$ ) Korrelationskoeffizienten ( $r=0.927$ ; Leber) und ( $r=0.964$ ; Aorta) verdeutlicht wird.



*Abbildung 22: Pairsplot zur Darstellung der Beziehung und Verteilung der  $SUV_{mean}$  der Referenzregion Aorta beider Beobachter*

Die nachfolgende Abbildung 23 zeigt anhand der  $SUV_{max}$  der Leber und der Aorta eines Beobachters die positive Korrelation ( $r=0.619$ ;  $p<0.001$ ) zwischen den Messwerten der Referenzregionen: Mit Zunahme der Traceranreicherung in der Leber steigt auch die Radionuklidspeicherung in der Aorta.

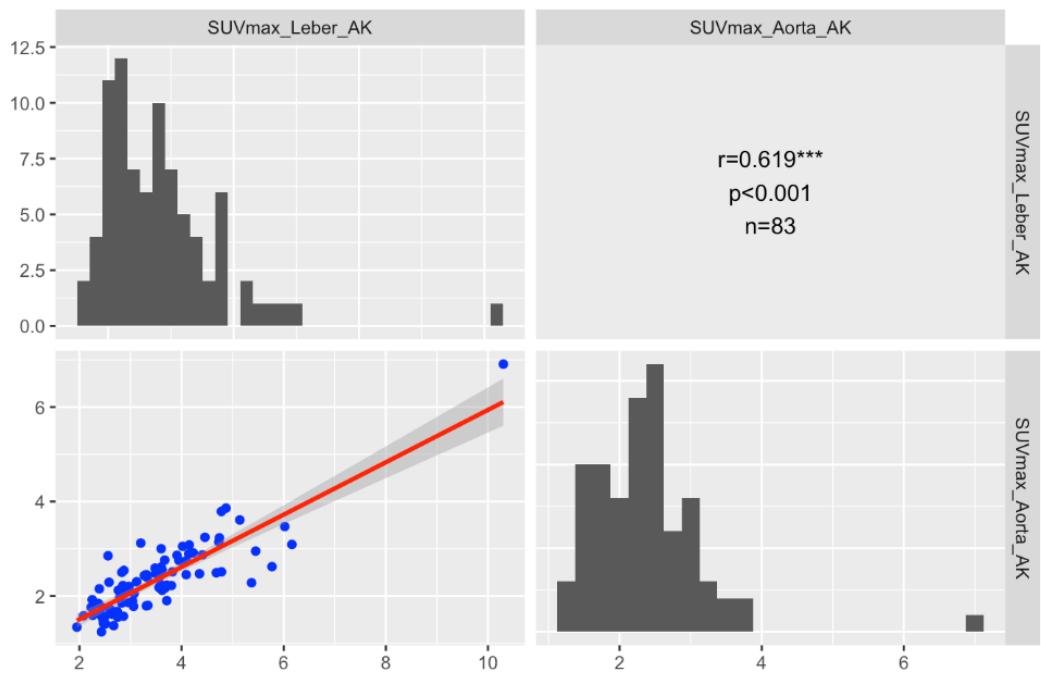


Abbildung 23: Pairsplot zur Darstellung der Beziehung der  $SUV_{max}$  der Referenzregion Leber und Aorta eines Beobachters

Die Übereinstimmung der gemessenen PERCIST-Schwellen ( $n=83$ ) der beiden Beobachter wird in Abbildung 24 graphisch dargestellt. Die ähnliche Verteilung der Messwerte ergibt sich aus den Histogrammen. Der Scatterplot zeigt die positive Korrelation der PERCIST-Werte der beiden Observer mit einem statistisch signifikanten ( $p<0.001$ ) Korrelationskoeffizienten ( $r=0.948$ ).

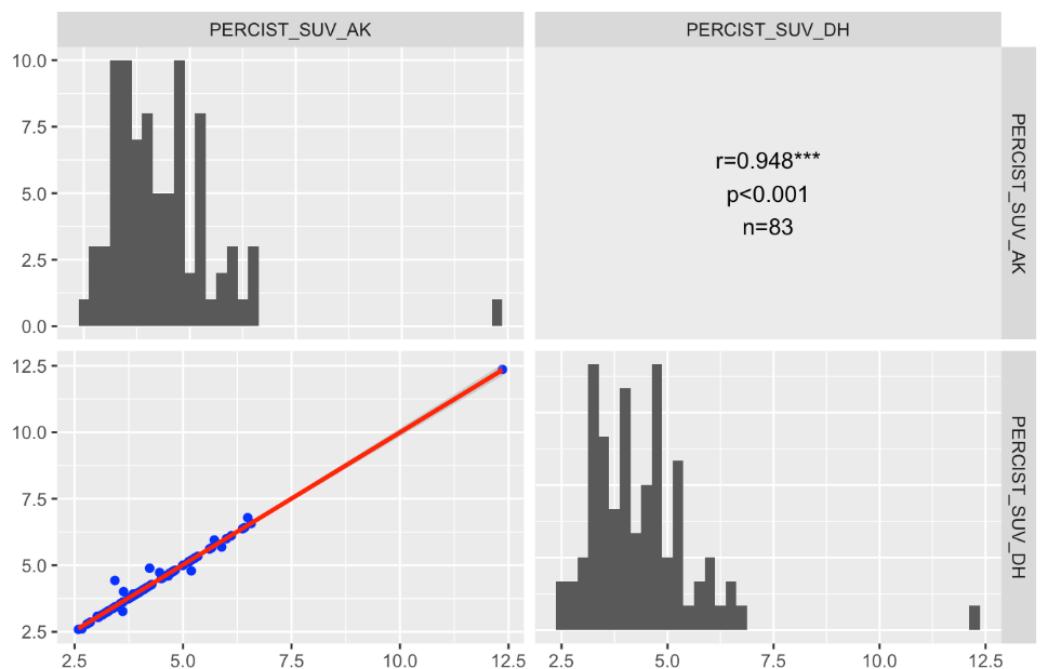


Abbildung 24: Pairsplot zur Darstellung der Verteilung der PERCIST-Schwellen beider Beobachter

Die Gegenüberstellung der Konturierungszeiten der beiden Beobachter (Abbildung 25) zeigt, dass ein Beobachter maximal 20 Minuten zur Befundung benötigt, wohingegen der andere davon abweichende Zeiten bis knapp über 40 Minuten verzeichnet.

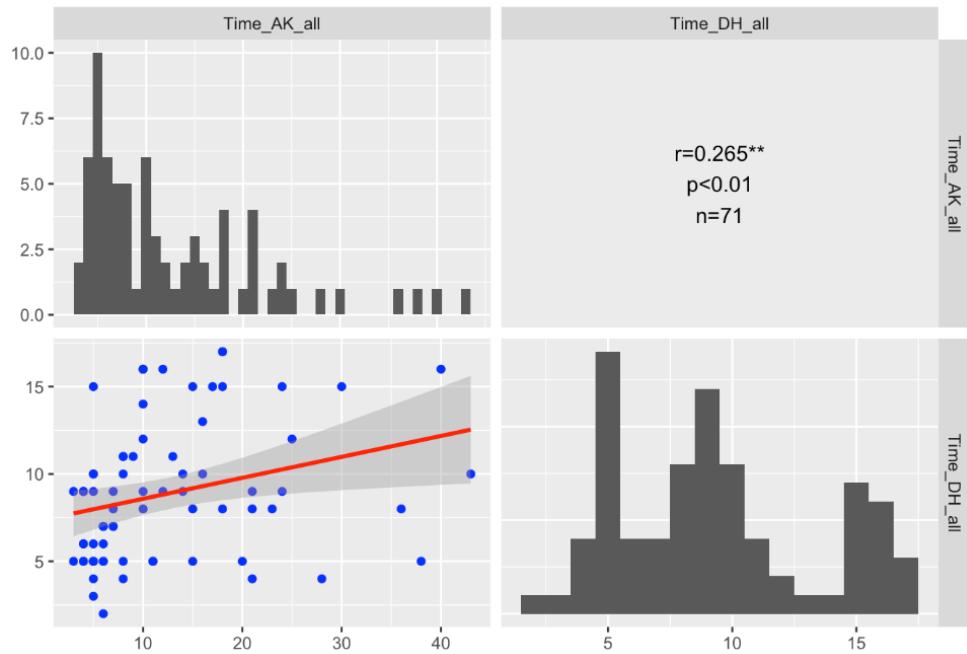


Abbildung 25: Pairsplot zum Vergleich der Konturierungszeiten der beiden Beobachter

Abbildung 26 verdeutlicht den positiven linearen Zusammenhang zwischen der Konturierungszeit und der Anzahl der ausgewählten Regionen eines Beobachters ( $r=0.42$ ;  $p<0.001$ ): Mit Zunahme der ROIs steigt auch die benötigte Zeit zum Konturieren der Läsionen.

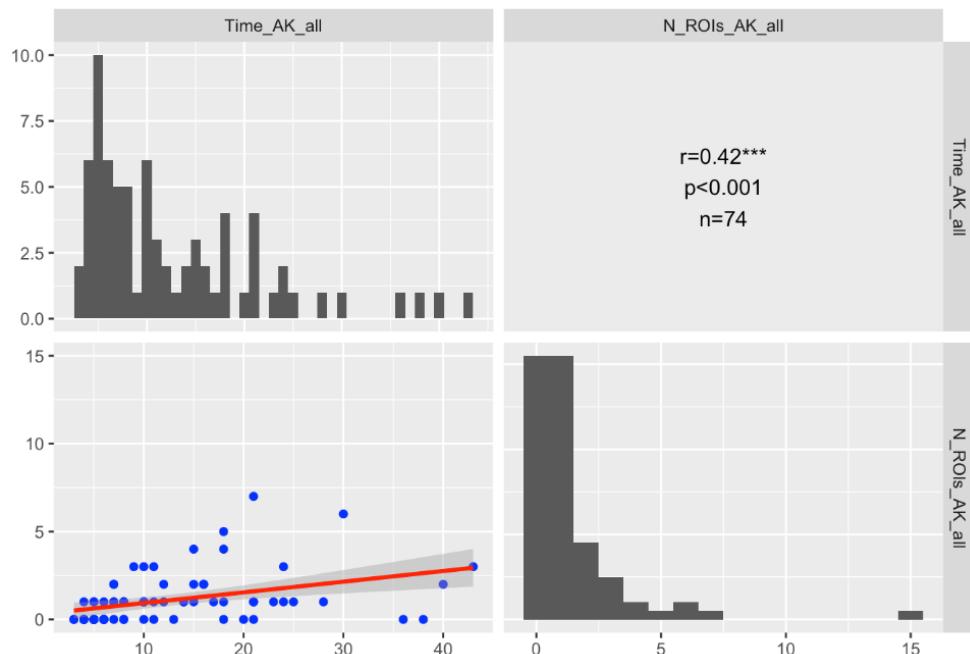


Abbildung 26: Pairsplot zur Darstellung der Beziehung der Konturierungszeit und der Anzahl der ausgewählten Regionen eines Beobachters

### 3.4.3 Interklassenkorrelationsanalyse

Die Interklassenkorrelationsanalyse der MTV-Ergebnisse unter Einschluss von Läsionen ohne Volumengrenze (n=83) ergab einen Interklassenkorrelationskoeffizienten (ICC(3,1)) von 0.768 mit einem 95 %-Konfidenzintervall (95 % CI) von 0.642 bis 0.850. Das bedeutet, dass 23.2 % der Gesamtvarianz der Messungen auf Unterschiede in den Messungen der beiden Observer zurückzuführen sind.

Für die TLG-Werte für Läsionen ohne Volumengrenze der beiden Beobachter (n=83) zeigte sich ein ICC(3,1) von 0.956 (95 % CI: 0.933 - 0.972).

Die  $SUV_{max}$  der Lymphomherde für Läsionen ohne Volumengrenze (n=50) beider Beobachter betrachtet, ergab sich ein ICC(3,1) von 1.00 (95 % CI: 1.00 - 1.00).

## 3.5 Tumor-Sink-Effekt

Nach Ausschluss der Normalverteilung der  $SUV_{max}$  der Leber und der MTV-Messwerte für Läsionen aller Volumina größer null (n=230) durch den Shapiro-Wilk-Test ( $p<0.001$ ) wurde der statistisch signifikante ( $p=0.038$ ) Kendall's Tau Korrelationskoeffizient mit -0.092 berechnet. Ein Patient konnte nicht in die nachfolgende Analyse eingeschlossen werden, da aufgrund separater Scans von Lymphomherd und Leber kein  $SUV_{max}$  zur Verfügung stand.

Die Annahmen für die Durchführung der linearen Regressionsanalyse wurden durch die nachfolgenden Tests bestätigt: Durch den Durbin-Watson-Test mit einem Wert von 1.427 konnte die Unabhängigkeit der Residuen bestätigt werden. Der Breusch-Pagan-Test ( $p=0.247$ ) schloss eine Heteroskedastizität der Residuen aus. Die maximale Cook's Distance von 0.132 bewies, dass das Modell nicht durch einzelne Beobachtungen übermäßig beeinflusst wird. Die Normalverteilung der Residuen bestätigte sich durch das nachstehende Histogramm (Abbildung 27).

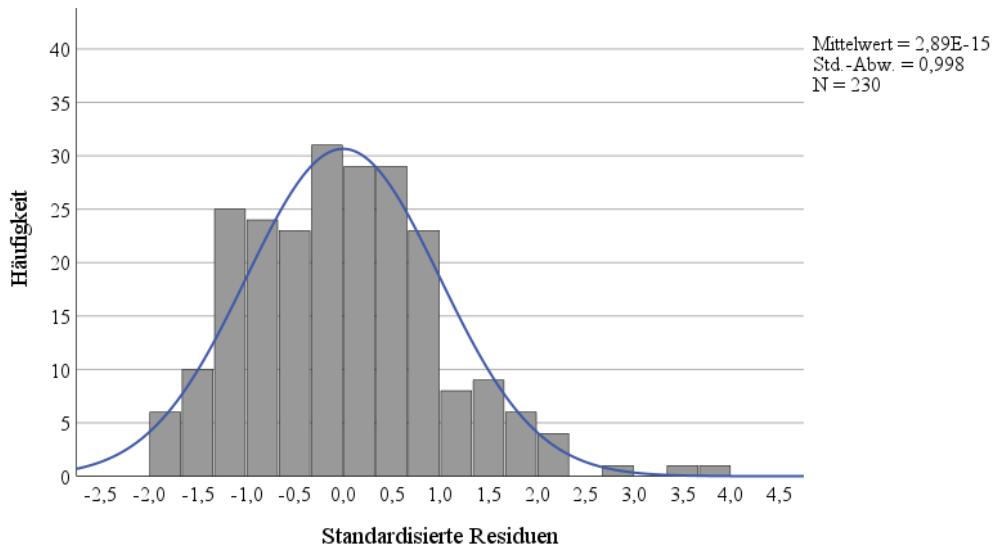


Abbildung 27: Histogramm zur Testung auf Normalverteilung der Residuen

In der sich anschließenden linearen Regressionsanalyse ergab sich ein  $R^2$ -Wert von 0.021.

Die Abbildung 28 veranschaulicht den negativ-linearen, statistisch signifikanten ( $p=0.027$ ) Zusammenhang des MTV für Läsionen ohne Volumengrenze größer null und der  $SUV_{max}$  der Leber. Die Regressionsgleichung lautete:  $SUV_{max} (\text{Leber}) = -0.002 \times MTV + 3.370$

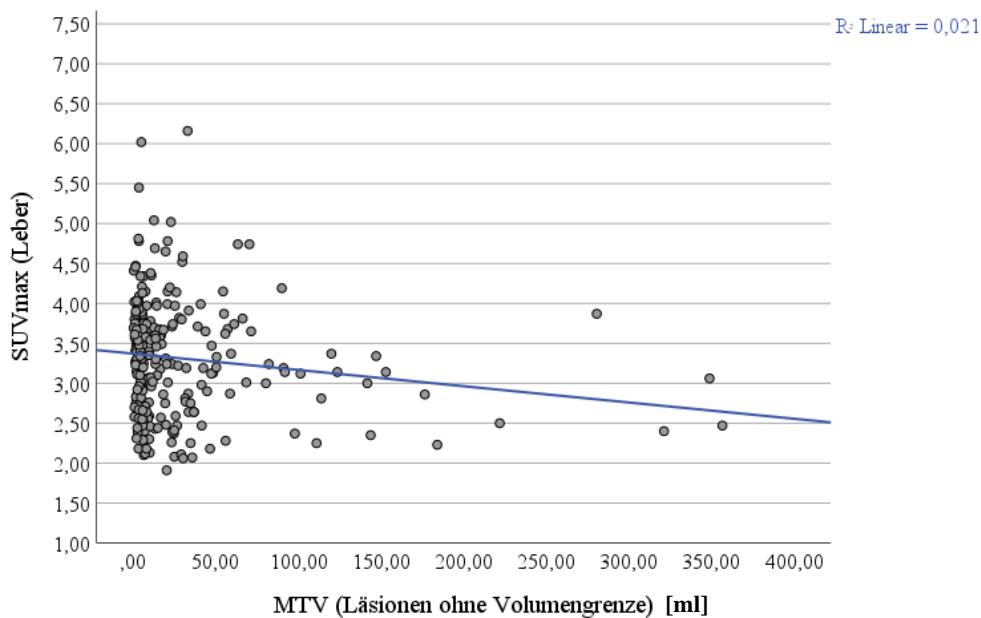


Abbildung 28: Scatterplot zur graphischen Darstellung des Zusammenhangs von MTV und  $SUV_{max}$  der Leber

Die Standardfehler für das MTV lag bei 0.001. Mit den T-Werten (-2.223 für das MTV und 60.119 für die Konstante) wurde die Abweichung der Koeffizienten von null bestätigt.

Auch nicht-lineare Modellierungen, einschließlich quadratischer, kubischer, logarithmischer und exponentieller Regression, sowie eine Log-Transformation der Variablen wurde

durchgeführt, um den Zusammenhang zwischen MTV und  $SUV_{max}$  zu modellieren. Keine der eben genannten Methoden konnte jedoch eine signifikante Besserung des  $R^2$ -Werts erzielen.

Zusätzlich zur Korrelation von MTV und der FDG-Anreicherung der Leber wurde auch der Zusammenhang zwischen MTV und der Traceranreicherung in der Aorta geprüft. Nach Ausschluss der Normalverteilung der  $SUV_{max}$  der Aorta und der MTV-Messwerte für Läsionen aller Volumina größer null ( $n=230$ ) durch den Shapiro-Wilk-Test ( $p<0.001$ ) wurde der statistisch signifikante ( $p=0.032$ ) Kendall's Tau Korrelationskoeffizient mit -0.095 berechnet. Ein Patient konnte nicht in die nachfolgende Analyse eingeschlossen werden, da aufgrund separater Scans von Läsion und Aorta kein  $SUV_{max}$  bestimmt werden konnte.

Die Annahmen für die Durchführung der linearen Regressionsanalyse wurden durch die nachfolgenden Tests bestätigt: Durch den Durbin-Watson-Test mit einem Wert von 1.457 konnte die Unabhängigkeit der Residuen gezeigt werden. Der Breusch-Pagan-Test ( $p=0.537$ ) widerlegte eine Heteroskedastizität der Residuen. Die maximale Cook's Distance von 0.313 bestätigte, dass das Modell nicht durch einzelne Beobachtungen übermäßig beeinflusst wird. Die Normalverteilung der Residuen wurde durch das nachstehende Histogramm (Abbildung 29) dargestellt.

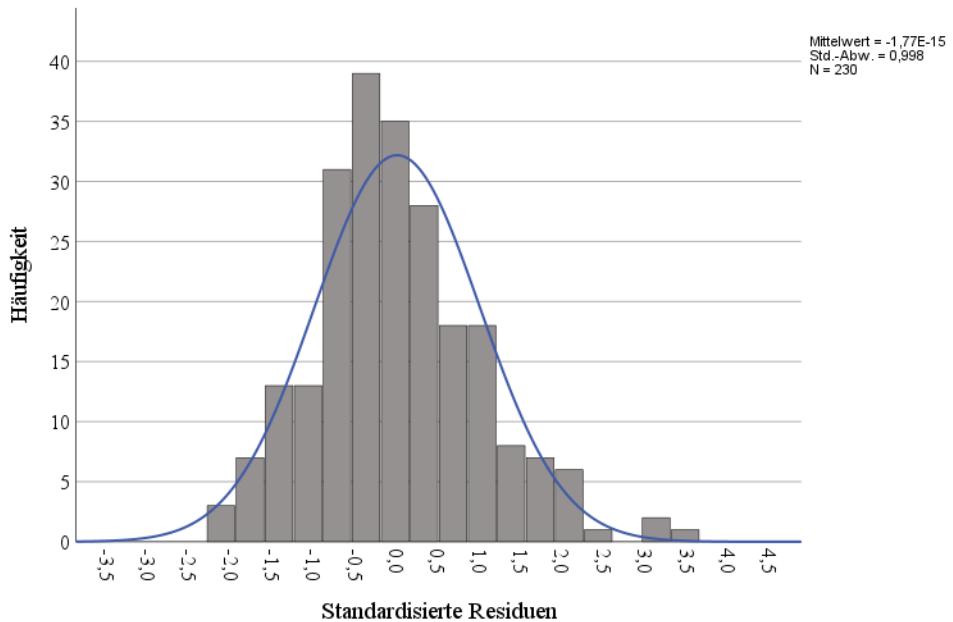
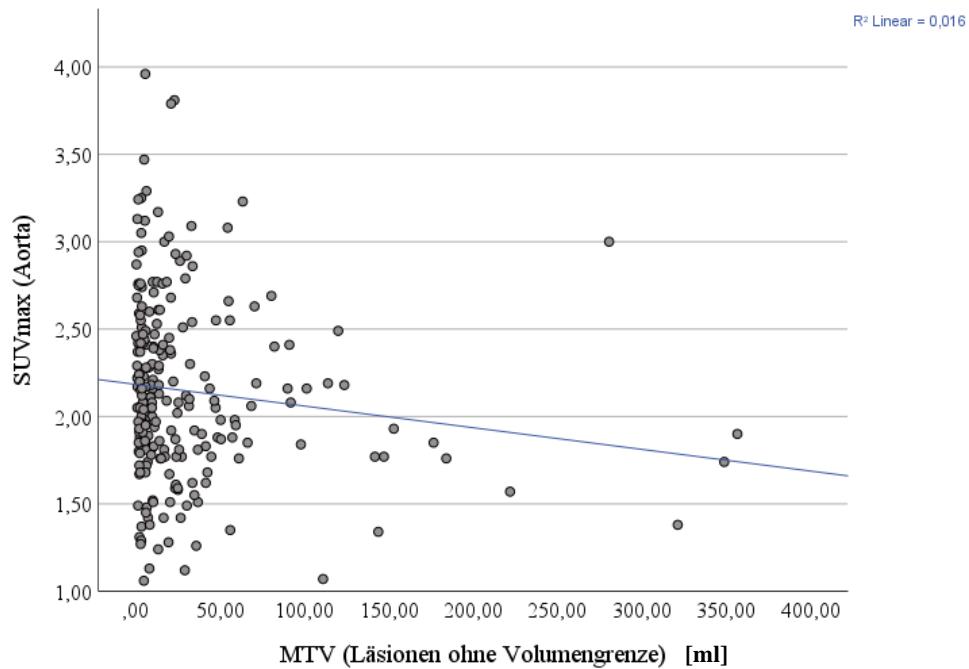


Abbildung 29: Histogramm zur Testung auf Normalverteilung der Residuen

In der sich anschließenden linearen Regressionsanalyse ergab sich ein  $R^2$ -Wert von 0.016.

Die Abbildung 30 veranschaulicht den negativ-linearen, statistisch nicht signifikanten ( $p=0,057$ ) Zusammenhang des MTV für Läsionen ohne Volumengrenze größer null und der  $SUV_{max}$  der Aorta.



*Abbildung 30: Scatterplot zur graphischen Darstellung des Zusammenhangs von MTV und  $SUV_{max}$  der Aorta*

Die Regressionsgleichung lautete:  $SUV_{max} (\text{Aorta}) = -0,001 \times MTV + 2,184$

Die Standardfehler für das MTV lagen bei 0.001. Mit den T-Werten (-1.914 für das MTV und 55.082 für die Konstante) wurde die Abweichung der Koeffizienten von null bestätigt.

Zusammenfassend ist ein geringer, aber statistisch signifikanter Tumor-Sink-Effekt auf die Speicherintensität in der Leber beobachtbar, allerdings nicht für die FDG-Anreicherung im Blutpool.

## 4. Diskussion

Der folgende Abschnitt liefert einen kurzen Überblick über die wichtigsten Ergebnisse der Arbeit und vergleicht das untersuchte Patientenkollektiv, die verwendeten Methoden und die Ergebnisse der Dissertation mit aktuellen wissenschaftlichen Studien. Am Ende des Kapitels folgen eine Erläuterung der Limitationen der durchgeführten Analyse und ein Ausblick.

### 4.1 Überblick über die Kernaussagen der Arbeit

1. Beim Vergleich der beiden Befundungsverfahren an den quantifizierbaren 1049 EoT-PETs der OPTIMAL>60-Studie war die Positivitätsrate unter Anwendung der visuellen Analysemethode beinahe doppelt so hoch wie jene unter Einsatz des quantitativen SUV-basierten Deauville-Scores (siehe Tabelle 3).
2. Die Analyse der 1050 EoT-PETs der OPTIMAL>60-Studie zeigte, dass es einen statistisch signifikanten Unterschied in den MTV- und TLG-Ergebnissen gibt, wenn nur Läsionen mit einem Mindestvolumen von 1 ml gezählt werden bzw. alle Läsionsgrößen berücksichtigt werden (siehe Kapitel 3.2 und 3.3).
3. Im Inter-Observer-Vergleich ergab sich eine sehr gute Übereinstimmung der Messwerte zwischen den beiden Beobachtern, welche sich sowohl graphisch als auch in verschiedenen statistischen Tests zeigte (siehe Kapitel 3.4).
4. Bei der Untersuchung hinsichtlich eines Tumor-Sink-Effekts präsentierte sich eine schwache, aber statistisch signifikante Korrelation zwischen den  $SUV_{max}$  der Leber und dem MTV. Der Zusammenhang der  $SUV_{max}$  der Aorta mit dem MTV war statistisch nicht signifikant (siehe Kapitel 3.5).

### 4.2 Vergleich mit Patientenkollektiven anderer Studien

Für die vorliegende Arbeit wurde das Patientenkollektiv der OPTIMAL>60-Studie analysiert, welche 1152 Patienten zwischen 61 und 80 Jahren mit noch unbehandelten, neu-diagnostizierten, histologisch gesicherten, CD20-positiven aggressiven Non-Hodgkin-Lymphomen umfasst. Als vergleichbare Studie lässt sich die HOVON84 Studie (EudraCT 2006-005174-42), eine internationale, multizentrische, randomisierte kontrollierte Studie nennen. Im Gegensatz zur OPTIMAL>60-Studie, welche beispielsweise auch Patienten mit follikulärem Lymphom Grad IIIb, plasmoblastischem Lymphom und/oder primärem Effusionslymphom einschließt, umfasst diese nur Patienten mit der Lymphomentität des neu-diagnostizierten, histologisch gesicherten, CD20-positiven diffusen großzelligen B-Zell-Lymphoms (50). Es gibt aktuell nur wenige PET-gestützte Studien zu aggressiven Non-Hodgkin-Lymphomen, welche verschiedene Lymphomentitäten miteinschließen. Ein Großteil

der Untersuchungen stützt sich allein auf die diffusen großzelligen B-Zell-Lymphome (34, 51, 52). Jene 575 Patienten der oben erwähnten HOVON84 Studie waren zwischen 18 und 80 Jahre alt und wurden hinsichtlich eines standardisierten und eines intensivierten R-CHOP-Schemas mit zusätzlicher Rituximab-Gabe randomisiert. Nach vier Zyklen wurde eine Interim-PET und nach sechs bis acht Zyklen eine EoT-PET durchgeführt, welche mithilfe des Deauville-Scores hinsichtlich der PET-Positivität von zwei unabhängigen Beobachtern bewertet wurden (50). Im Rahmen der OPTIMAL>60-Studie wird unter anderem auch liposomales Vincristin im Rahmen des R-CHLIP-Schemas eingesetzt und mit dem konventionellen Vincristin (R-CHOP-Schema) hinsichtlich der Neurotoxizität und des progressionsfreien Überlebens verglichen. Bisher ist liposomales Vincristin vor allem aus Phase-II-Studien zur Behandlung von Patienten mit rezidivierendem und refraktären diffusen großzelligen B-Zell-Lymphom bekannt (53, 54). In der OPTIMAL>60-Studie wird liposomales Vincristin erstmals als First-Line-Therapie mit konventionellem Vincristin direkt verglichen.

Das untersuchte Subkollektiv mit PET-positiven Lymphomherden in der EoT-PET der OPTIMAL>60-Studie zeigt hinsichtlich seiner demographischen Merkmale (siehe Kapitel 2.1.4) Ähnlichkeiten zu Patientenpopulationen aus anderen Studien. Ein vergleichbares Patientenkollektiv untersuchten Allioux et al. in ihrem Paper zur Risikostratifizierung von Patienten mit diffusem großzelligen B-Zell-Lymphom in der EoT-PET. Von den 108 analysierten Patienten wiesen mehr als die Hälfte, ähnlich wie das oben genannte Subkollektiv der OPTIMAL>60-Studie, eine über den Normwert erhöhte LDH, ein Lymphomausbreitungsmuster nach Ann-Arbor-Stadium IV sowie eine ähnliche Prävalenz von Bulkformationen und extralymphatischen Manifestationen auf. Anders als in der OPTIMAL>60-Studie war die Altersspanne der Patienten aber deutlich größer und reichte von 19 bis 84 Jahre (Mittelwert  $60\pm13$ ) (55). PET-gestützte Forschung zu Non-Hodgkin-Lymphomen, die sich speziell auf ältere Patienten konzentriert, wie die OPTIMAL>60-Studie, ist bislang nur begrenzt vorhanden. Betrachtet man das durchschnittliche Erkrankungsalter von 67 Jahren bei Non-Hodgkin-Lymphomen (4), wird deutlich, dass mehr Studien mit Fokus auf ältere Patienten benötigt werden.

#### 4.3 Vergleich mit Methodiken anderer Studien

Die OPTIMAL>60-Studie zeichnet sich als prospektive Studie mit sehr hoher Fallzahl aus. Es gibt nur wenige aktuelle Studien dieser Größenordnung, die die PET-Diagnostik, vor allem auch die EoT-PET, bei aggressiven Non-Hodgkin-Lymphomen untersuchen. Als vergleichbare Arbeit lässt sich die GOYA-Studie nennen. In dieser Phase-III-Studie wurde das

progressionsfreie Überleben sowie das Gesamtüberleben von 1418 Patienten, welche in die beiden Therapiearme CHOP mit Rituximab und alternativ CHOP mit Obinutuzumab randomisiert wurden, anhand des PET-Status untersucht (56). Viele andere Studien hingegen fokussieren sich auf die initiale und/oder Interim-PET (57, 58) oder umfassen eine deutlich kleinere Studienpopulation (59, 60).

Die PET-Bewertung im Rahmen der OPTIMAL>60-Studie erfolgte auf Basis eines visuellen Vergleichs der FDG-speichernden Lymphomareale mit dem mediastinalen Blutpool als Referenzorgan. Läsionen, die visuell eine höhere Radionuklidspeicherung als die Aorta verzeichneten, wurden als PET-positiv gewertet (siehe Kapitel 2.1.2). Zu Beginn der OPTIMAL>60-Studie existierte die erst später eingeführte Lugano-Klassifikation mit dem daraus hervorgehenden Deauville-Score noch nicht. Seit Einführung jenes ist der Deauville-Score eines der wichtigsten PET-Bewertungsschemata. Er wurde sowohl in der vorliegenden Dissertation als auch in zahlreichen aktuellen Studien angewandt (50, 52, 56). Weitere Bewertungsverfahren stehen im Fokus wissenschaftlicher Arbeiten. Dazu zählt unter anderem die Bewertung anhand  $\Delta SUV_{max}$ , wobei die  $SUV_{max}$  der Lymphomherde in der prätherapeutischen initialen PET und in der EoT-PET miteinander verglichen werden, um einen Therapieerfolg zu bestimmen (55):

$$\Delta SUV_{max} = \frac{Baseline\ SUV_{max}\ (Läsion) - EoT\ SUV_{max}\ (Läsion)}{Baseline\ SUV_{max}\ (Läsion)}$$

Auch die *Lesion-to-Liver-SUV<sub>max</sub>-Ratio (LLR)* wird als mögliche Alternative zum Deauville-Score diskutiert. Diese ergibt sich aus dem Quotienten von  $SUV_{max}$  der Läsion geteilt durch den  $SUV_{max}$  der Leber: Das Ansprechen der Lymphomläsionen wird anhand der Tracer-Aufnahme des Tumors im Vergleich zur Radionuklidspeicherung der Leber gemessen (61). Inwiefern die beiden genannten Verfahren den Deauville-Score in der klinischen Praxis ergänzen bzw. ersetzen könnten, bedarf weiterer Untersuchung.

Die Konturierung der Läsionen in der vorliegenden Dissertation erfolgte gemäß Empfehlung der Europäischen Gesellschaft für Nuklearmedizin (EANM) anhand eines Schwellenwertes von 41 % des läsionalen  $SUV_{max}$  (36). Mehrere Studien (34, 62) zeigten, dass die prognostische Aussagekraft des MTV unabhängig von der angewandten Segmentierungsmethode ähnlich war. Obwohl die prognostische Aussagekraft vergleichbar war, ergaben sich bei Anwendung verschiedener Verfahren stark abweichende MTV-Messwerte. Es war daher bisher nicht möglich, einen Schwellenwert festzulegen, der zwischen Hochrisiko- und Niedrigrisiko-Patienten unterscheidet (63). Eine Standardisierung der MTV-Messmethoden ist daher

notwendig (64). Boellaard et al. erarbeiteten einen internationalen Benchmark-Datensatz, um MTV-Messungen bei Lymphomen zu standardisieren und als reproduzierbaren Biomarker zugänglich zu machen. Sie fokussierten sich unter anderem auf die SUV4.0-Methode, ein Schwellenwert-basiertes Verfahren (65). Jenes stellte sich in anderen Studien als erfolgreichste Methode zur MTV-Messung bei Patienten mit diffusem großzelligen B-Zell-Lymphom heraus (51) und zeigte verglichen mit anderen am wenigsten Beeinflussung von PET-Rekonstruktionsmechanismen (66). Boellaard et al. betonten jedoch, dass es weiterer Untersuchung der SUV4.0-Methode in klinischen Studien bedarf (65).

Für den vorliegenden Inter-Observer-Vergleich wurden 83 EoT-PETs der OPTIMAL>60-Studie zufällig nach festgelegten Kriterien (siehe Kapitel 2.3) ausgewählt. Analysiert wurde die Inter-Observer-Reliabilität von MTV, TLG, der benötigten Konturierungszeit und der Anzahl der ROIs sowie die Reproduzierbarkeit der SUVs der Lymphomherde und der Referenzregionen. Die Konturierung erfolgte anhand eines Schwellenwertes von 41 % des läsionalen  $SUV_{max}$ . Ein Vergleich lässt sich ziehen mit der Arbeit von Zwezerijnen et al. Diese untersucht die Interim-PETs von 45 zufällig ausgewählten Patienten mit DS 4- und DS 5-Läsionen der HOVON84-Studie hinsichtlich der MTV-Ermittlung mittels sechs verschiedener Segmentierungsverfahren, unter anderem auch jener Methode, welche einen Schwellenwert von 41 % des läsionalen  $SUV_{max}$  nutzt (67). Im Gegensatz zur Vergleichsstudie von Zwezerijnen et al., welche sich auf Übereinstimmung von quantitativen, volumetrischen Daten bezieht, beschäftigen sich andere Inter-Observer-Vergleiche mit der Untersuchung von qualitativen Beurteilungen, PET-Bewertung (positiv oder negativ), und der Übereinstimmung der Auswahl segmentierter nodaler und extranodaler Regionen (50, 68). Der im Rahmen der Dissertation durchgeführte Inter-Observer-Vergleich behandelt vor allem die Reliabilität der volumetrischen Messungen, erörtert aber auch Unterschiede in der Auswahl von segmentierten Regionen.

#### 4.4 Vergleich mit Ergebnissen anderer Studien

##### 4.4.1 PET-Bewertungsverfahren

Verschiedene Verfahren zur Bewertung des Therapieansprechens lagen im Fokus der Arbeit. Dabei zeigte sich, dass die Positivitätsrate unter Anwendung des visuellen Bewertungsverfahrens (40.6 %) beinahe doppelt so hoch ist wie die unter Einsatz des quantitativen SUV-basierten Deauville-Scores (22.0 %). Das liegt unter anderem daran, dass das visuelle Befundungsverfahren gemäß der Protokollvorgabe in der OPTIMAL>60-Studie alle Läsionen als PET-positiv zählt, die visuell eine höhere Traceranreicherung als die Intensität

in der Aorta zeigen. Unter Anwendung der allgemein üblichen Bewertungskriterien für den Deauville-Score (12) werden nur jene Lymphomherde als PET-positiv gezählt, deren Radionuklidspeicherung quantitativ die der Leber überschreitet (DS 4 und 5), siehe Kapitel 1.1.3. Die Positivitätsrate der Patienten der OPTIMAL>60-Studie liegt mit 22.0 bzw. 40.6 % nahe an anderen Studienergebnissen, die die PET-Positivität von Patienten mit aggressiven Non-Hodgkin-Lymphomen nach R-CHOP-Standardtherapie untersuchten. Coiffier et al. zufolge können 30 bis 40 % aller mit R-CHOP behandelten Patienten durch die Therapie nicht geheilt werden: Etwa die Hälfte jener Patienten spreche kaum bis gar nicht auf die Therapie an, während fünf Prozent eine partielle Remission zeigen (69).

Um die prognostische Aussagekraft der beiden unterschiedlichen PET-Bewertungsverfahren vergleichen zu können, bedarf es einer Korrelation mit Überlebensdaten, welche aufgrund des gültigen Daten-Embargos noch nicht zur Verfügung stehen. Die aktuellen Studien zum Zusammenhang der verschiedenen PET-Bewertungsverfahren mit dem progressionsfreien Überleben und Gesamtüberleben liefern differierende Ergebnisse. Albano et al. zeigten in ihrer Untersuchung an EoT-PETs von Patienten mit Burkitt-Lymphom, sowohl für die PET-Bewertung mithilfe des Deauville-Scores als auch für jene anhand der visuellen IHP-Kriterien (siehe Kapitel 2.1.2) eine signifikante Korrelation mit dem progressionsfreien Überleben (PFS) und Gesamtüberleben (Overall Survival, OS) (70). In der Studie von Han et al. an EoT-PETs von Patienten mit diffusem großzelligem B-Zell-Lymphom zeigten sich hingegen weder für die Befundung nach DS noch nach IHP statistisch signifikante Korrelationen mit dem PFS und OS (71).

Fallanca et al. führten in ihrem Vergleich der PET-Bewertungsverfahren, IHP und Deauville-Score, ein weiteres an: einen angepassten Deauville-Score, der auch PETs mit der Bewertung DS 3 als positiv zählt. Die drei Befundungsverfahren miteinander verglichen, präsentierte der Deauville-Score mit positiver Bewertung bei DS 4 und 5 die besten Ergebnisse hinsichtlich der Überlebensvorhersage (72). Der Großteil der aktuellen Studien nutzt den Deauville-Score mit PET-Positivität bei DS 4 und 5 (56, 57, 67, 73). Eine Untersuchung von Lee et. al. zeigte auch, dass sich die Prognose von Patienten mit DS 3 Lymphomherden in der EoT-PET nicht signifikant von Patienten mit DS 1 und DS 2 Läsionen unterschied. Im Vergleich zu Patienten mit DS 4 und DS 5 Lymphomherden verzeichneten Patienten mit DS 1, 2 und/oder 3 Lymphomanifestationen eine gute Prognose (74).

Obwohl vor allem der Deauville-Score mit PET-Positivität bei DS 4 und 5 die Kriterien des IHP in der klinischen Forschung zunehmend verdrängt hat, existiert aktuell kein Goldstandard zur Bewertung von EoT-PETs bei Patienten mit aggressiven Non-Hodgkin-Lymphomen. Mit aufstrebenden PET-Befundungsschemata in der Erprobung (Lesion-to-Liver-SUV<sub>max</sub>-Ratio,  $\Delta$ SUV<sub>max</sub>, siehe Kapitel 4.3) bleibt das ideale Verfahren zur PET-Bewertung weiterhin unklar.

#### 4.4.2 MTV-Messwerte

Zum Vergleich der erhobenen MTV-Messwerte eignet sich eine Gegenüberstellung mit den MTV-Ergebnissen, bei welchen ein Schwellenwert von 41 % des läsionalen SUV<sub>max</sub> zur Konturierung der Lymphomherde diente, aus der oben beschriebenen Untersuchung von Zwezerijnen et al. Dabei wurden bei 45 PET-Scans 60 DS 4- und DS 5-Läsionen mit einem Median von 27.6 ml (Interquartilbereich 3.5-214.6 ml) detektiert (67). Die PET-Aufnahmen mit PET-Bewertung *positiv* der OPTIMAL>60-Studie (n=231) verzeichneten 638 DS 4- und DS 5-Lymphomherde mit einem Median von 11.8 ml (Interquartilbereich 4.6-32.6 ml). Beim Vergleich der beiden MTV-Ergebnisse fallen der deutlich kleinere Median und Interquartilabstand der OPTIMAL>60-Studie auf, welche sich höchstwahrscheinlich auf die deutlich größere Stichprobe zurückführen lassen.

Ebenso wie die log-transformierten MTV-Werte der Untersuchung von Zwezerijnen et al. (67) zeigten auch die MTV-Werte der OPTIMAL>60-Studie eine Normalverteilung nach Logarithmierung (siehe Kapitel 3.2).

#### 4.4.3 ROI-Größen

Ein zentraler Forschungspunkt der vorliegenden Arbeit handelt von der Untersuchung der ROI-Größe bei der Konturierung von Lymphomherden. Die dazu erfolgten vergleichende Analysen zeigten, dass sich die MTV- und die TLG-Ergebnisreihen mit den beiden unterschiedlichen ROI-Größen jeweils signifikant voneinander unterscheiden.

Von den 231 PET-positiven Patienten hatten sieben Patienten nur Läsionen, deren gemessenes Volumen kleiner als 1 ml war. Aufschlussreich wäre nun eine Überlebensanalyse jener Patienten, welche jedoch aufgrund der noch unter Embargo stehenden Überlebensdaten nicht möglich ist.

Zur Fragestellung der Relevanz von kleinen Lymphomläsionen hinsichtlich des Verlaufs bzw. des Fortschreitens der Erkrankung fehlen aktuelle wissenschaftliche Daten. Ein möglicher Grund dafür könnte die bisher fehlende klare Empfehlung zur Integration von Volumenschwellen in der Lugano-Klassifikation sein (30). Möglicherweise spielen auch die in

Kapitel 1.2.5 beschriebenen technischen Schwierigkeiten, die mit der Konturierung von kleinen Läsionen einhergehen, eine Rolle. Der Einbezug der kleinen Läsionen führt zu einem statistisch signifikanten Unterschied der MTV- und der TLG-Ergebnisse. Inwiefern die Berücksichtigung von kleinen Läsionen das Überleben beeinflusst, ist jedoch noch unklar.

#### 4.4.4 Inter-Observer-Vergleich

Die sehr gute Übereinstimmung und damit die Reproduzierbarkeit der Ergebnisse zwischen beiden Beobachtern konnte sowohl graphisch durch Visualisierung der Messwerte in Bland-Altman-Plots und Scatterplots gezeigt werden, als auch quantitativ durch Berechnung der Kendall's Tau Korrelationskoeffizienten und der Interklassenkorrelationsanalyse. Wie in Kapitel 4.3 bereits beschrieben, bedarf es einer Standardisierung der MTV-Messung zum Einsatz des MTV als prognostischen Biomarker im klinischen Alltag. Dabei spielt auch die Reliabilität der Segmentierungsmethode eine große Rolle. Karimjee et al. haben dazu einen Inter-Observer- und Inter-Instrument-Vergleich durchgeführt. Dabei wurden vollautomatische Segmentierungsverfahren mittels Künstlicher Intelligenz (KI) mit der auch in dieser Arbeit verwendeten semi-automatischen Segmentierungsmethode (siehe Kapitel 2.2) anhand von Bland-Altman-Plots und ICC verglichen. Unabhängig von der Segmentierungsmethode wurde eine sehr gute Reliabilität der MTV- und TLG-Messungen gezeigt. Die ICC des Inter-Observer-Vergleichs für das semi-automatische Segmentierungsverfahren von Karimjee et al. waren vergleichbar mit den im Rahmen der Dissertation berechneten Ergebnissen, lediglich die ICC für die MTV-Werte waren bei Karimjee et al. etwas höher (75). Ein möglicher Grund hierfür könnten Verzerrungen durch die beiden in Kapitel 3.4.1 beschriebenen differierenden PET-Bewertungen sein. Der ICC-Wert zur Beschreibung der Übereinstimmung der TLG-Ergebnisse wird durch die beiden abweichenden Beurteilungen weniger beeinflusst. Das liegt möglicherweise daran, dass die exzellente Übereinstimmung der SUVs der Lymphomherde die differierenden MTV-Werte im Produkt ausgleicht ( $TLG = SUV_{mean} \times MTV$ ).

Neben der sehr guten Reproduzierbarkeit der Segmentierungsmethode spielt auch die Konturierungszeit in der klinischen Praxis eine bedeutende Rolle. Die benötigte Dauer zur Analyse der Patienten der OPTIMAL>60-Studie betrug beim erfahrenen Beobachter bis zu maximal 20 Minuten (siehe Kapitel 3.4.2). Ilyas et al. verzeichneten mit bis zu 21 Minuten vergleichbare Zeiten (34), ähnlich Burggraaf et al. mit durchschnittlich 22 Minuten bei Notwendigkeit von manuellen Korrekturen (76). Karimjee et al. zeigten unter Anwendung der Künstlichen Intelligenz deutlich geringere Konturierungszeiten von nur bis zu zehn Minuten,

mit einer mittleren Zeitersparnis von zwei Minuten und 30 Sekunden pro Messung, verglichen mit der semi-automatischen Segmentierungsmethode (75).

Die Herausforderungen in der Segmentierung von Lymphomherden liegen für KI-basierte Verfahren vor allem in der stark variablen Lage und Größe der Lymphomläsionen, sowie der Notwendigkeit Organe mit physiologischer FDG-Speicherung auszuschließen. Dies führt zu geringerer Leistung bei der Segmentierung von Lymphommanifestationen, verglichen mit Primärtumoren anderer Krebsarten (77). Auch Unterschiede in der Bildbeschaffenheit können die Leistung KI-basierter Segmentierungsverfahren beeinflussen. Ein entsprechend vielfältiger Trainingsdatensatz kann zwar die Variationen in der Bildqualität abdecken, jedoch können die dafür benötigten Ressourcen ein limitierender Faktor sein (78).

Entgegen den eben beschriebenen Herausforderungen haben Yousefirizi et al. ein KI-basiertes Segmentierungsnetzwerk zur MTV-Ermittlung entwickelt, welches anhand von PET/CT-Scans von Melanom-, Bronchialkarzinom- und Lymphompatienten entwickelt und an multizentrischen PET/CT-Scans von Patienten mit unterschiedlichen Lymphomsubtypen und Krankheitsstadien getestet wurde. Es zeigte gute Ergebnisse für die MTV-Ermittlung von Patienten mit verschiedenen Lymphomsubtypen. Zukünftig soll auch eine implementierte benutzerfreundliche Oberfläche den Ärzten ermöglichen, aktiv in den Lernprozess des Segmentierungsmodells einzutreten und ihn zu verbessern (79).

Vollautomatisierte Segmentierungsverfahren stehen im Fokus aktueller Forschung und können in allen Stadien des klinischen Managements (Staging, Therapieplanung und Evaluierung des Behandlungserfolgs) wertvolle Unterstützung leisten (80). Zukünftig könnten sie beispielsweise auch mit einer Zweitmeinung die befundenen Ärzte entlasten (81).

#### 4.4.5 Tumor-Sink-Effekt

Der negative Kendall's Tau Korrelationskoeffizient von -0.092 und die negative Steigung  $\beta_1$  von -0.002 der linearen Regressionsanalyse konnten einen statistisch signifikanten, wenn auch nur schwachen, negativ-linearen Zusammenhang zwischen dem  $SUV_{max}$  der Leber und dem MTV nachweisen: Mit steigendem MTV sinkt die FDG-Speicherung der Leber. Analoge Beobachtungen machten auch Wu et al. (40) und Oliveira et al. (41): Mit gesunkenem MTV nach Chemotherapie stieg die FDG-Aufnahme der Leber. Neben der Radionuklidspeicherung der Leber untersuchten Wu et al. in ihrem Paper auch den Zusammenhang zwischen dem MTV und der Tracer-Speicherung der Aorta. Letztere blieb anders als die SUVs der Leber über den Verlauf der Chemotherapie stabil (40). Die Untersuchung des Subkollektivs der

OPTIMAL>60-Studie brachte ähnliche Ergebnisse: Die Korrelation zwischen MTV und  $SUV_{max}$  der Aorta zeigte in der linearen Regressionsanalyse einen schwächeren, negativ-linearen, nicht signifikanten Zusammenhang ( $\beta_1 = -0.001$ ) verglichen mit der Untersuchung der SUVs der Leber. Nicht auszuschließen ist daher auch eine Erhöhung der Radionuklidspeicherung durch reversible Änderungen im Lebermetabolismus, beispielsweise Steatosis hepatis (82), im Rahmen der Chemotherapie (83).

Kenntnis über die FDG-Dynamik der Leber und der Aorta im Rahmen der Chemotherapie ist vor allem aufgrund ihrer Rolle als Referenzorgane bedeutend. Der Deauville-Score nutzt zur Einschätzung der Tracer-Aufnahme von Läsionen die Radionuklidspeicherung der Leber und der Aorta als Vergleich. Während manche Studien daher andere PET-Bewertungsverfahren bei Verfügbarkeit einer initialen PET, wie die Methode des  $\Delta SUV_{max}$ , befürworten (55, 84), zeigten andere Untersuchungen, dass das  $\Delta SUV_{max}$ -Verfahren dem Deauville-Score nicht überlegen war (85).

Zur Bewertung von EoT-PETs werden häufig die Maxima der SUVs herangezogen (61). Wu et al. fanden in ihrer Untersuchung jedoch, dass die SUV-Mittelwerte der Leber besser mit dem SUV-Maximum des Tumors, dem MTV und der TLG korrelierten (40). Zusätzlich zeigten Keramida et al., dass die Chemotherapie-induzierte Steatosis die  $SUV_{max}$  der Leber erhöht, nicht aber die  $SUV_{mean}$  (86). Die  $SUV_{max}$  würden auch durch Rauschen und Schwankungen der Bildqualität stärker beeinflusst als die  $SUV_{mean}$  oder  $SUV_{peak}$  (87, 88). Zwezerijnen et al. empfehlen daher das SUV-Maximum der Leber nicht zur Quantifizierung der FDG-Speicherung von Leber und Tumorherden heranzuziehen. Wurde der  $SUV_{max}$  bisher aufgrund seiner sehr guten Inter-Observer-Reliabilität verwendet, sollte nun laut Zwezerijnen et al. bei Verfügbarkeit zur Berechnung der *Lesion-to-Liver-SUV<sub>max</sub>-Ratio* der  $SUV_{peak}$  des Tumors geteilt durch den  $SUV_{mean}$  der Leber verwendet werden. Aktuell gibt es jedoch keine einheitlichen Vorgaben, wie die SUVs der Leber eingesetzt werden sollten (39).

#### 4.5 Limitationen

Die vorliegende Arbeit weist einige Einschränkungen auf. Dazu zählen zum einen die fehlenden Überlebensdaten der Patienten der OPTIMAL>60-Studie. Zum Zeitpunkt der Analyse stehen diese noch unter Embargo, was die Korrelation der erhobenen Ergebnisse mit Überlebensdaten unmöglich macht. Prognostische Aussagen über PET-Bewertungsverfahren und ROI-Größen sind daher aktuell nicht möglich.

Zum anderen stellt auch die Unvollständigkeit der Datengrundlage eine Limitation dar. Es fehlen Informationen über die 97 ursprünglich eingeschlossenen Patienten, bei welchen weder eine EoT-PET dokumentiert noch entsprechende Bilddaten verfügbar waren. Ohne Kenntnis der Daten sind die potenziellen Einflüsse auf den Gesamtdatensatz nicht abschätzbar. Die mangelnde Verfügbarkeit der Restaging-PET spielt auch in der klinischen Praxis eine bedeutende Rolle. Die EoT-PET bei Patienten mit aggressiven Non-Hodgkin-Lymphomen wird im Gegensatz zur initialen PET von der gesetzlichen Krankenversicherung in Deutschland noch nicht bezahlt. Die Übernahme der Kosten für die Baseline-PET trat mit dem G-BA-Beschluss (Gemeinsamer Bundesausschuss) im März 2020 in Kraft (6).

#### 4.6 Ausblick

Die Chancen der vorliegenden Studie liegen in der Eignung der erhobenen Ergebnisse für weiterführende Analysen, vor allem der Korrelation mit den nach Wegfall des Daten-Embargos zur Verfügung stehenden Überlebensdaten. Sie kann wichtige Informationen über die EoT-PET bei Patienten mit aggressiven Non-Hodgkin-Lymphomen liefern, vor allem auch hinsichtlich des Einsatzes von Volumengrenzen und verschiedener PET-Bewertungsverfahren. In Zukunft werden vermutlich auch die in Kapitel 4.4.4 beschriebenen vollautomatischen Segmentierungsverfahren eine zunehmend wichtige Rolle im klinischen Alltag spielen.

## 5. Zusammenfassung

Die vorliegende Arbeit analysiert die PET-Positivität und das Metabolic Tumor Volume (MTV) in der End-of-Treatment-PET (EoT-PET) bei Patienten mit aggressiven Non-Hodgkin-Lymphomen anhand des Patientenkollektivs der OPTIMAL>60-Studie (n=1152). Ziele der Forschung waren die Analyse der MTV-Messung mit verschiedenen Volumenschwellen, der Vergleich zweier PET-Befundungsverfahren (quantitativ SUV-basierter Deauville-Score versus visuelle Bewertung), die Überprüfung der Reproduzierbarkeit quantitativer PET-basierter Biomarker sowie die Analyse hinsichtlich eines Tumor-Sink-Effekts.

Dazu wurden die 1050 EoT-PETs der Studie mithilfe des Software-Paktes *syngo.via* von Siemens unter Anwendung des Deauville-Scores ausgewertet (negativ DS 1, 2 und 3, positiv: DS 4 und 5). Lymphomherde mit  $SUV_{max}$  über der PERCIST-Schwelle ( $1.5 \times SUV_{mean}^{Leber} + 2 \times SUV_{SD}^{Leber}$ ) wurden anhand einer Schwelle von 41 % des läsionalen  $SUV_{max}$  konturiert.

Die visuelle Analysemethode der OPTIMAL>60-Studie (positiv bereits bei läsionaler Anreicherung über mediastinalem Blutpool) ergab eine PET-Positivität von 40.6 % gegenüber 22.0 % bei Verwendung der modernen Lugano-Kriterien für PET-Positivität mit DS 4 und DS 5 ( $p<0.001$  im McNemar-Test). Von den 1050 analysierten EoT-PETs waren 231 Patienten nach Deauville-Kriterien PET-positiv mit einem (ohne läsionales Mindestvolumen gemessenen) MTV von  $30.6 \pm 53.0$  ml. Bei läsionalem Mindestvolumen von 1 ml war das MTV bei den verbleibenden 224 PET-positiven Patienten mit  $31.4 \pm 53.5$  ml signifikant höher ( $p<0.001$ ). Der Inter-Observer-Vergleich zweier unabhängiger Beobachter an einem Subkollektiv von 83 Patienten ergab eine sehr gute Reproduzierbarkeit der MTV-Ergebnisse mit  $ICC(3,1)=0.768$ , der TLG-Werte mit  $ICC(3,1)=0.956$  und der läsionalen  $SUV_{max}$  mit  $ICC(3,1)=1.00$ . Ein schwacher, aber signifikanter Tumor-Sink-Effekt zwischen  $SUV_{max}^{Leber}$  und MTV war in einer linearen Regressionsanalyse mit negativer Steigung von -0.002 bei  $p<0.027$  zu beobachten. Für  $SUV_{max}^{Aorta}$  konnte keine solche Korrelation mit dem MTV gefunden werden.

Die beschriebenen Ergebnisse verdeutlichen, dass die visuellen Kriterien der OPTIMAL>60-Studie zu einer signifikant höheren Positivitätsrate als die moderne Lugano-Klassifikation führen und dass in der EoT-PET die Messung des MTV von Herden, die nach Lugano-Kriterien PET-positiv sind, mit hoher Reproduzierbarkeit möglich ist. Die Beurteilung der prognostischen Aussagekraft der EoT-PET kann mit den hier erzeugten Daten anhand der Überlebensdaten der OPTIMAL>60-Studie nach Wegfall des Daten-Embargos im Jahr 2025 erfolgen.

## 6. Anhang

### a. Tabellenverzeichnis

<b>Tabelle 1:</b> Deauville-Score zur Beurteilung des Therapieansprechens der Non-Hodgkin-Lymphome .....	9
<b>Tabelle 2:</b> Klinische Charakteristika des Subkollektivs mit PET-positivem Lymphomgewebe in der End-of-Treatment-PET .....	18
<b>Tabelle 3:</b> Vergleich der PET-Bewertungen, gewonnen nach visuellem Analyseverfahren und mithilfe des Deauville-Scores (DS). PET-Bewertung Deauville, basierend auf Vergleich läsionales SUV-Maximum gegenüber Referenzorganen Blutpool bzw. Leber: positiv bei DS 4 und 5, negativ bei DS 1, 2 und 3; PET-Bewertung visuell: Einschätzung im Vergleich zum mediastinalen Blutpool. Positiv entspricht läsionaler Speicherintensität über, negativ unter mediastinalem Blutpool.....	25
<b>Tabelle 4:</b> Deskriptive Statistik für MTV, TLG, $SUV_{max}$ und die Anzahl der ROIs unter Berücksichtigung der Läsionen mit einem Mindestvolumen von 1 ml. SD: Standardabweichung .....	26
<b>Tabelle 5:</b> Deskriptive Statistik für MTV, TLG, $SUV_{max}$ und die Anzahl der ROIs unter Berücksichtigung von Läsionen aller Volumina. SD: Standardabweichung .....	27

## b. Abbildungsverzeichnis

<b>Abbildung 1:</b> STARD-Flowchart zur Darstellung des analysierten Patientenkollektivs. RE2-PET: Restaging-PET, STARD: Standards for Reporting of Diagnostic Accuracy.....	17
<b>Abbildung 2:</b> Häufigkeitsverteilung der MTV-Werte bei Läsionen aller Volumina.....	27
<b>Abbildung 3:</b> Häufigkeitsverteilung der logarithmierten MTV-Werte bei Läsionen aller Volumina.....	28
<b>Abbildung 4:</b> Häufigkeitsverteilung der TLG-Werte bei Läsionen aller Volumina.....	28
<b>Abbildung 5:</b> Häufigkeitsverteilung der logarithmierten TLG-Werte bei Läsionen aller Volumina.....	29
<b>Abbildung 6:</b> Bland-Altman-Plot zur Analyse der Übereinstimmung beider Beobachter für die MTV-Werte aller ROI-Größen. SD: Standardabweichung .....	30
<b>Abbildung 7:</b> Bland-Altman-Plot zur Analyse der Übereinstimmung beider Beobachter für die TLG-Werte aller ROI-Größen. SD: Standardabweichung .....	30
<b>Abbildung 8:</b> Maximum-Intensity-Projections (MIP) der prätherapeutischen FDG-PET und nach Induktionstherapie aus ventraler Sicht ohne und nach Konturierung durch zwei unabhängige Befunder. Nach initial extranodalen Manifestationen mit unter anderem fokalem Milzbefall und nach Induktionstherapie deutlich PET-positiven Resten intraspinal differieren die Bewertungen der diffus erhöhten Milzspeicherung. ....	31
<b>Abbildung 9:</b> Maximum-Intensity-Projections (MIP) der prätherapeutischen FDG-PET und nach Induktionstherapie aus ventraler Sicht ohne und nach Konturierung durch zwei unabhängige Befunder. Nach initialem Lungenbefall und nach Induktionstherapie persistierenden nodalen Manifestationen differieren die Bewertungen der diffus erhöhten Knochenmarkspeicherung im Becken.....	32
<b>Abbildung 10:</b> Bland-Altman-Plot zur Analyse der Übereinstimmung beider Beobachter für die SUV <sub>max</sub> der Lymphomherde der Patienten mit PET-Positivität für alle ROI-Größen. SD: Standardabweichung .....	32
<b>Abbildung 11:</b> Bland-Altman-Plot zur Analyse der Übereinstimmung beider Beobachter für die SUV <sub>max</sub> der Leber. SD: Standardabweichung .....	33
<b>Abbildung 12:</b> Bland-Altman-Plot zur Analyse der Übereinstimmung beider Beobachter für die SUV <sub>max</sub> der Aorta. SD: Standardabweichung .....	34
<b>Abbildung 13:</b> Bland-Altman-Plot zur Analyse der Übereinstimmung beider Beobachter für die SUV <sub>mean</sub> der Leber. SD: Standardabweichung .....	34
<b>Abbildung 14:</b> Bland-Altman-Plot zur Analyse der Übereinstimmung beider Beobachter für die SUV <sub>mean</sub> der Aorta. SD: Standardabweichung .....	35
<b>Abbildung 15:</b> Pairsplot zur Darstellung der Beziehung und Verteilung der MTV-Ergebnisse unter Einschluss von Läsionen aller ROI-Größen beider Beobachter .....	36
<b>Abbildung 16:</b> Pairsplot zur Darstellung der Beziehung der Anzahl der ROIs und der MTV-Ergebnisse unter Einschluss von Läsionen ohne Volumengrenze eines Beobachters .....	36
<b>Abbildung 17:</b> Pairsplot zur Darstellung der Beziehung und der Verteilung der TLG-Werte beider Beobachter unter Einschluss von Läsionen ohne Volumengrenze .....	37
<b>Abbildung 18:</b> Pairsplot zur Darstellung der Beziehung und Verteilung der SUV <sub>max</sub> der Lymphomherde beider Beobachter unter Einschluss von Läsionen ohne Volumengrenze .....	38
<b>Abbildung 19:</b> Pairsplot zur Darstellung der Beziehung und Verteilung der SUV <sub>max</sub> der Leber beider Beobachter.....	38
<b>Abbildung 20:</b> Pairsplot zur Darstellung der Beziehung und Verteilung der SUV <sub>max</sub> der Aorta beider Beobachter.....	39

<b>Abbildung 21:</b> Pairsplot zur Darstellung der Beziehung und Verteilung der $SUV_{mean}$ der Referenzregion Leber beider Beobachter.....	39
<b>Abbildung 22:</b> Pairsplot zur Darstellung der Beziehung und Verteilung der $SUV_{mean}$ der Referenzregion Aorta beider Beobachter .....	40
<b>Abbildung 23:</b> Pairsplot zur Darstellung der Beziehung der $SUV_{max}$ der Referenzregion Leber und Aorta eines Beobachters .....	41
<b>Abbildung 24:</b> Pairsplot zur Darstellung der Verteilung der PERCIST-Schwellen beider Beobachter.....	41
<b>Abbildung 25:</b> Pairsplot zum Vergleich der Konturierungszeiten der beiden Beobachter....	42
<b>Abbildung 26:</b> Pairsplot zur Darstellung der Beziehung der Konturierungszeit und der Anzahl der ausgewählten Regionen eines Beobachters .....	42
<b>Abbildung 27:</b> Histogramm zur Testung auf Normalverteilung der Residuen.....	44
<b>Abbildung 28:</b> Scatterplot zur graphischen Darstellung des Zusammenhangs von MTV und $SUV_{max}$ der Leber .....	44
<b>Abbildung 29:</b> Histogramm zur Testung auf Normalverteilung der Residuen.....	45
<b>Abbildung 30:</b> Scatterplot zur graphischen Darstellung des Zusammenhangs von MTV und $SUV_{max}$ der Aorta .....	46

### c. Abkürzungsverzeichnis

BiTEs	Bispezifische Antikörper
CAR-T-Zell-Therapie	Chimärer Antigenrezeptor-T-Zell-Therapie
CI	Konfidenzintervall
CT	Computertomographie
DLBCL	Diffuses großzelliges B-Zell-Lymphom
DS	Deauville-Score
EANM	European Association of Nuclear Medicine
EBV	Epstein-Barr-Virus
ECOG	Eastern Cooperative Oncology Group
EoT	End-of-Treatment
FDG	F-18-2-Fluoro-2-desoxy-D-glucose
G-BA	Gemeinsamer Bundesausschuss
G-CSF	granulocyte-colony stimulating factor
ICC	Interklassenkorrelationskoeffizient
IHP	International Harmonization Project
IPI	International Prognostic Index
KI	Künstliche Intelligenz
LDH	Laktatdehydrogenase
LOR	Line Of Response
Max	Maximum
Min	Minimum
MIP	Maximum Intensity Projection
MTV	Metabolic Tumor Volume

NHL	Non-Hodgkin-Lymphome
OS	Gesamtüberleben
PERCIST	Positron Emission Response Criteria in Solid Tumors
PET	Positronenemissionstomographie
PFS	Progressionsfreies Überleben
R-CHLIP	Rituximab, Cyclophosphamid, Doxorubicin, liposomales Vincristin und Prednison
R-CHOP	Rituximab, Cyclophosphamid, Doxorubicin, konventionelles Vincristin und Prednison
r-IPI	revised-International Prognostic Index
RKI	Robert-Koch-Institut
ROI	region of interest
SD	Standardabweichung
STARD	Standards for Reporting of Diagnostic Accuracy
SUV	Standardized Uptake Value
TLG	Tumor Lesion Glycolysis
WHO	World Health Organization

## 7. Literaturverzeichnis

1. Bowzyk Al-Naeeb A, Ajithkumar T, Behan S, Hodson DJ. Non-Hodgkin lymphoma. *BMJ* 2018; 362:1-7.
2. Ansell SM. Non-Hodgkin Lymphoma: Diagnosis and Treatment. *Mayo Clin Proc* 2015; 90(8):1152–63. doi: 10.1016/j.mayocp.2015.04.025.
3. Swerdlow SH, Campo E, Pileri SA, Harris NL, Stein H, Siebert R et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood* 2016; 127(20):2375–90. doi: 10.1182/blood-2016-01-643569.
4. Thandra KC, Barsouk A, Saginala K, Padala SA, Barsouk A, Rawla P. Epidemiology of Non-Hodgkin's Lymphoma. *Med Sci (Basel)* 2021; 9(1) [Stand: 15.11.2023]. Verfügbar unter: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7930980/pdf/medsci-09-00005.pdf>.
5. Alaggio R, Amador C, Anagnostopoulos I, Attygalle AD, Araujo IBdO, Berti E et al. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Lymphoid Neoplasms. *Leukemia* 2022; 36(7):1720–48 [Stand: 08.07.2024]. Verfügbar unter: [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9214472/pdf/41375\\_2022\\_Article\\_1620.pdf](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9214472/pdf/41375_2022_Article_1620.pdf).
6. Leitlinienprogramm Onkologie (Deutsche Krebsgesellschaft, Deutsche Krebshilfe, AWMF). Diagnostik, Therapie und Nachsorge für erwachsene Patient\*innen mit einem diffusen großzelligen B-Zell-Lymphom und verwandten Entitäten: Langversion 1.0, AWMF-Registernummer: 018/038OL 2022 [Stand: 26.09.2024]. Verfügbar unter: [https://www.leitlinienprogramm-onkologie.de/fileadmin/user\\_upload/Downloads/Leitlinien/DLBCL/Version\\_1/LL\\_DLBC\\_L\\_Langversion\\_1.0.pdf](https://www.leitlinienprogramm-onkologie.de/fileadmin/user_upload/Downloads/Leitlinien/DLBCL/Version_1/LL_DLBC_L_Langversion_1.0.pdf).
7. Ernst M, Dührsen U, Hellwig D, Lenz G, Skoetz N, Borchmann P. Diffuse Large B-Cell Lymphoma and Related Entities. *Dtsch Arztebl Int* 2023; 120(17):289–96.
8. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2024; 74(3):229–63.
9. Robert Koch-Institut, Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V., Hrsg. Krebs in Deutschland für 2019/2020. Berlin; 2023.
10. Shaffer AL, Young RM, Staudt LM. Pathogenesis of human B cell lymphomas. *Annu Rev Immunol* 2012; 30:565–610.
11. Vockerodt M, Yap L-F, Shannon-Lowe C, Curley H, Wei W, Vrzalikova K et al. The Epstein-Barr virus and the pathogenesis of lymphoma. *J Pathol* 2015; 235(2):312–22. doi: 10.1002/path.4459.
12. Cheson BD, Fisher RI, Barrington SF, Cavalli F, Schwartz LH, Zucca E et al. Recommendations for initial evaluation, staging, and response assessment of Hodgkin and non-Hodgkin lymphoma: the Lugano classification. *J Clin Oncol* 2014; 32(27):3059–68. doi: 10.1200/JCO.2013.54.8800.

13. Barrington SF, Mikhael NG, Kostakoglu L, Meignan M, Hutchings M, Müeller SP et al. Role of imaging in the staging and response assessment of lymphoma: consensus of the International Conference on Malignant Lymphomas Imaging Working Group. *J Clin Oncol* 2014; 32(27):3048–58. doi: 10.1200/JCO.2013.53.5229.
14. Juweid ME, Stroobants S, Hoekstra OS, Mottaghy FM, Dietlein M, Guermazi A et al. Use of positron emission tomography for response assessment of lymphoma: consensus of the Imaging Subcommittee of International Harmonization Project in Lymphoma. *JCO* 2007; 25(5):571–8. doi: 10.1200/JCO.2006.08.2305.
15. Martelli M, Ceriani L, Zucca E, Zinzani PL, Ferreri AJM, Vitolo U et al. <sup>18</sup>Ffluorodeoxyglucose positron emission tomography predicts survival after chemoimmunotherapy for primary mediastinal large B-cell lymphoma: results of the International Extranodal Lymphoma Study Group IELSG-26 Study. *J Clin Oncol* 2014; 32(17):1769–75. doi: 10.1200/JCO.2013.51.7524.
16. Vodicka P, Klener P, Trneny M. Diffuse Large B-Cell Lymphoma (DLBCL): Early Patient Management and Emerging Treatment Options. *Onco Targets Ther* 2022; 15:1481–501 [Stand: 25.06.2024]. Verfügbar unter: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9739046/pdf/ott-15-1481.pdf>.
17. Oertel M, Berdel C, Held G, Herfarth K, Schmidberger H, Ernst M et al. The new German evidence-based guideline on diffuse large B-cell lymphoma-key aspects for radiation oncologists. *Strahlenther Onkol* 2023; 199(2):115–20. Verfügbar unter: <https://link.springer.com/article/10.1007/s00066-022-02035-9>.
18. International Non-Hodgkin's Lymphoma Prognostic Factors Project. A predictive model for aggressive non-Hodgkin's lymphoma. *N Engl J Med* 1993; 329(14):987–94. doi: 10.1056/NEJM199309303291402.
19. Sehn LH, Berry B, Chhanabhai M, Fitzgerald C, Gill K, Hoskins P et al. The revised International Prognostic Index (R-IPI) is a better predictor of outcome than the standard IPI for patients with diffuse large B-cell lymphoma treated with R-CHOP. *Blood* 2007; 109(5):1857–61. Verfügbar unter: <https://pubmed.ncbi.nlm.nih.gov/17105812/>.
20. Hellwig D, Dührsen U. Non-Hodgkin-Lymphome – PET in Diagnostik und Therapiesteuerung. *Onkologe* 2019; 25(10):867–79.
21. Yu S. Review of F-FDG Synthesis and Quality Control. *Biomed Imaging Interv J* 2006; 2(4):1-11. doi: 10.2349/bij.2.4.e57.
22. Warburg O, Wind F, Negelein E. THE METABOLISM OF TUMORS IN THE BODY. *J Gen Physiol* 1927; 8(6):519–30. doi: 10.1085/jgp.8.6.519.
23. Luna A, Vilanova JC, Da Hygino Cruz LC, Rossi SE. Functional Imaging in Oncology. Berlin, Heidelberg: Springer Berlin Heidelberg; 2014.
24. Ruhlmann J. PET in der Onkologie: Grundlagen und Klinische Anwendung. Berlin, Heidelberg: Springer Berlin / Heidelberg; 1998.
25. Dietlein M, Kopka K, Schmidt M, Hrsg. Nuklearmedizin: Basiswissen und klinische Anwendung. 9. aktualisierte und erweiterte Auflage. Stuttgart: Thieme; 2023.

26. Reader AJ, Pan B. AI for PET image reconstruction. *Br J Radiol* 2023; 96(1150):20230292. Verfügbar unter: <https://academic.oup.com/bjr/article/96/1150/20230292/7499053>.
27. Xie M, Wu K, Liu Y, Jiang Q, Xie Y. Predictive value of F-18 FDG PET/CT quantization parameters in diffuse large B cell lymphoma: a meta-analysis with 702 participants. *Med Oncol* 2015; 32(1):446 [Stand: 26.06.2024]. Verfügbar unter: <https://link.springer.com/content/pdf/10.1007/s12032-014-0446-1.pdf>.
28. Guo B, Tan X, Ke Q, Cen H. Prognostic value of baseline metabolic tumor volume and total lesion glycolysis in patients with lymphoma: A meta-analysis. *PLoS One* 2019; 14(1):e0210224.
29. Adams MC, Turkington TG, Wilson JM, Wong TZ. A systematic review of the factors affecting accuracy of SUV measurements. *AJR. American journal of roentgenology* 2010; 195(2):310–20. Verfügbar unter: <https://pubmed.ncbi.nlm.nih.gov/20651185/>.
30. Ricard F, Barrington S, Korn R, Brueggenwerth G, Trotman J, Cheson B et al. Application of the Lugano Classification for Initial Evaluation, Staging, and Response Assessment of Hodgkin and Non-Hodgkin Lymphoma: The PRoLoG Consensus Initiative (Part 2-Technical). *J Nucl Med* 2023; 64(2):239–43. Verfügbar unter: <https://jnm.snmjournals.org/content/64/2/239>.
31. Zhao W, Wu X, Huang S, Wang H, Fu H. Evaluation of therapeutic effect and prognostic value of 18F-FDG PET/CT in different treatment nodes of DLBCL patients. *EJNMMI Res* 2024; 14(1):20. doi: 10.1186/s13550-024-01074-w.
32. Boellaard R, Krak NC, Hoekstra OS, Lammertsma AA. Effects of noise, image resolution, and ROI definition on the accuracy of standard uptake values: a simulation study. *J Nucl Med* 2004; 45(9):1519–27.
33. Akamatsu G, Shimada N, Matsumoto K, Daisaki H, Suzuki K, Watabe H et al. New standards for phantom image quality and SUV harmonization range for multicenter oncology PET studies. *Ann Nucl Med* 2022; 36(2):144–61. Verfügbar unter: <https://link.springer.com/article/10.1007/s12149-021-01709-1>.
34. Ilyas H, Mikhaeel NG, Dunn JT, Rahman F, Møller H, Smith D et al. Defining the optimal method for measuring baseline metabolic tumour volume in diffuse large B cell lymphoma. *Eur J Nucl Med Mol Imaging* 2018; 45(7):1142–54. Verfügbar unter: <https://link.springer.com/article/10.1007/s00259-018-3953-z>.
35. Im H-J, Bradshaw T, Solaiyappan M, Cho SY. Current Methods to Define Metabolic Tumor Volume in Positron Emission Tomography: Which One is Better? *Nucl Med Mol Imaging* 2018; 52(1):5–15. doi: 10.1007/s13139-017-0493-6.
36. Boellaard R, Delgado-Bolton R, Oyen WJG, Giammarile F, Tatsch K, Eschner W et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging* 2015; 42(2):328–54. doi: 10.1007/s00259-014-2961-x.
37. van der Vos CS, Koopman D, Rijnsdorp S, Arends AJ, Boellaard R, van Dalen JA et al. Quantification, improvement, and harmonization of small lesion detection with state-of-the-art PET. *Eur J Nucl Med Mol Imaging* 2017; 44(Suppl 1):4–16. Verfügbar unter: <https://link.springer.com/article/10.1007/s00259-017-3727-z>.

38. Soret M, Bacharach SL, Buvat I. Partial-volume effect in PET tumor imaging. *J Nucl Med* 2007; 48(6):932–45. Verfügbar unter: <https://jnm.snmjournals.org/content/48/6/932>.
39. Zwezerijnen GJC, Eertink JJ, Ferrández MC, Wiegers SE, Burggraaff CN, Lugtenburg PJ et al. Reproducibility of 18FFDG PET/CT liver SUV as reference or normalisation factor. *Eur J Nucl Med Mol Imaging* 2023; 50(2):486–93. doi: 10.1007/s00259-022-05977-5.
40. Wu X, Bhattacharai A, Korkola P, Pertovaara H, Eskola H, Kellokumpu-Lehtinen P-L. The Association Between Liver and Tumor 18FFDG Uptake in Patients with Diffuse Large B Cell Lymphoma During Chemotherapy. *Mol Imaging Biol* 2017; 19(5):787–94. doi: 10.1007/s11307-017-1044-3.
41. Oliveira M, Lasnon C, Nganoa C, Gac A-C, Damaj G, Aide N. Comprehensive analysis of the influence of G-CSF on the biodistribution of 18F-FDG in lymphoma patients: insights for PET/CT scheduling. *EJNMMI Res* 2019; 9(1):79. doi: 10.1186/s13550-019-0546-1.
42. Gallamini A, Hutchings M, Rigacci L, Specht L, Merli F, Hansen M et al. Early interim 2-18Ffluoro-2-deoxy-D-glucose positron emission tomography is prognostically superior to international prognostic score in advanced-stage Hodgkin's lymphoma: a report from a joint Italian-Danish study. *JCO* 2007; 25(24):3746–52.
43. syngo.via – Siemens Healthineers Deutschland; 2024 [Stand: 06.06.2024]. Verfügbar unter: <https://www.siemens-healthineers.com/de/digital-health-solutions/syngovia>.
44. 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(Suppl 1):122S-50S. doi: 10.2967/jnumed.108.057307.
45. Leonhart R. Lehrbuch Statistik: Einstieg und Vertiefung. 5., überarbeitete Auflage. Bern: Hogrefe; 2022 [Stand: 01.07.2024]. Verfügbar unter: <https://elibrary.hogrefe.com/content/pdf/10.1024/86258-000.pdf>.
46. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999; 8(2):135–60. doi: 10.1177/096228029900800204.
47. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of chiropractic medicine* 2016; 15(2):155–63. Verfügbar unter: <https://pubmed.ncbi.nlm.nih.gov/27330520/>.
48. Dreger C. Ökonometrie: Grundlagen – Methoden – Beispiele. 5., überarb. u. akt. Aufl. 2014. Wiesbaden: Springer Gabler; 2014. (SpringerLink Bücher).
49. Bühner M, Ziegler M. Statistik für Psychologen und Sozialwissenschaftler. 2., aktualisierte und erweiterte Auflage. Hallbergmoos: Pearson Deutschland GmbH; 2017. (Studium Psychologie). Verfügbar unter: <https://elibrary.pearson.de/book/99.150005/9783863268091>.
50. Burggraaff CN, Cornelisse AC, Hoekstra OS, Lugtenburg PJ, Keizer B de, Arens AIJ et al. Interobserver Agreement of Interim and End-of-Treatment 18F-FDG PET/CT in Diffuse Large B-Cell Lymphoma: Impact on Clinical Practice and Trials. *J Nucl Med* 2018; 59(12):1831–6. doi: 10.2967/jnumed.118.210807.

51. Barrington SF, Zwezerijnen BGJC, Vet HCW de, Heymans MW, Mikhaeel NG, Burggraaff CN et al. Automated Segmentation of Baseline Metabolic Total Tumor Burden in Diffuse Large B-Cell Lymphoma: Which Method Is Most Successful? A Study on Behalf of the PETRA Consortium. *J Nucl Med* 2021; 62(3):332–7. doi: 10.2967/jnumed.119.238923.
52. Kanemasa Y, Shimoyama T, Sasaki Y, Tamura M, Sawada T, Omuro Y et al. Analysis of prognostic value of complete response by PET-CT and further stratification by clinical and biological markers in DLBCL patients. *Med Oncol* 2017; 34(2):29. doi: 10.1007/s12032-017-0885-6.
53. Kaplan LD, Deitcher SR, Silverman JA, Morgan G. Phase II study of vincristine sulfate liposome injection (Marqibo) and rituximab for patients with relapsed and refractory diffuse large B-Cell lymphoma or mantle cell lymphoma in need of palliative therapy. *Clin Lymphoma Myeloma Leuk* 2014; 14(1):37–42. doi: 10.1016/j.clml.2013.09.009.
54. Hagemeister F, Rodriguez MA, Deitcher SR, Younes A, Fayad L, Goy A et al. Long term results of a phase 2 study of vincristine sulfate liposome injection (Marqibo(®)) substituted for non-liposomal vincristine in cyclophosphamide, doxorubicin, vincristine, prednisone with or without rituximab for patients with untreated aggressive non-Hodgkin lymphomas. *Br J Haematol* 2013; 162(5):631–8. doi: 10.1111/bjh.12446.
55. Allioux F, Gandhi D, Vilque J-P, Nganoa C, Gac A-C, Aide N et al. End-of-treatment 18F-FDG PET/CT in diffuse large B cell lymphoma patients: ΔSUV outperforms Deauville score. *Leuk Lymphoma* 2021; 62(12):2890–8.
56. Kostakoglu L, Martelli M, Sehn LH, Belada D, Carella A-M, Chua N et al. End-of-treatment PET/CT predicts PFS and OS in DLBCL after first-line treatment: results from GOYA. *Blood Adv* 2021; 5(5):1283–90. doi: 10.1182/bloodadvances.2020002690.
57. Zucca E, Cascione L, Ruberto T, Facchinelli D, Schär S, Hayoz S et al. Prognostic models integrating quantitative parameters from baseline and interim positron emission computed tomography in patients with diffuse large B-cell lymphoma: post-hoc analysis from the SAKK38/07 clinical trial. *Hematol Oncol* 2020; 38(5):715–25. doi: 10.1002/hon.2805.
58. Dührsen U, Müller S, Hertenstein B, Thomssen H, Kotzerke J, Mesters R et al. Positron Emission Tomography-Guided Therapy of Aggressive Non-Hodgkin Lymphomas (PETAL): A Multicenter, Randomized Phase III Trial. *JCO* 2018; 36(20):2024–34. doi: 10.1200/JCO.2017.76.8093.
59. Michaud L, Bantilan K, Mauguen A, Moskowitz CH, Zelenetz AD, Schöder H. Prognostic Value of 18F-FDG PET/CT in Diffuse Large B-Cell Lymphoma Treated with a Risk-Adapted Immunotherapy Regimen. *J Nucl Med* 2023; 64(4):536–41. doi: 10.2967/jnumed.122.264740.
60. Freitas AC, Carvalho IP, Esteves S, Salgado L, Da Gomes Silva M. End of treatment FDG-PET in primary mediastinal B-cell lymphoma treated with R-chemotherapy: Prognostic indicator and implications for consolidation radiotherapy. *Eur J Haematol* 2022; 108(2):118–24. doi: 10.1111/ejh.13715.
61. Li Y-H, Zhao Y-M, Jiang Y-L, Tang S, Chen M-T, Xiao Z-Z et al. The prognostic value of end-of-treatment FDG-PET/CT in diffuse large B cell lymphoma: comparison of visual

- Deauville criteria and a lesion-to-liver SUV<sub>max</sub> ratio-based evaluation system. *Eur J Nucl Med Mol Imaging* 2022; 49(4):1311–21. doi: 10.1007/s00259-021-05581-z.
- 62. Cottreau A-S, Hapdey S, Chartier L, Modzelewski R, Casasnovas O, Itti E et al. Baseline Total Metabolic Tumor Volume Measured with Fixed or Different Adaptive Thresholding Methods Equally Predicts Outcome in Peripheral T Cell Lymphoma. *J Nucl Med* 2017; 58(2):276–81. doi: 10.2967/jnumed.116.180406.
  - 63. El-Galaly TC, Villa D, Cheah CY, Gormsen LC. Pre-treatment total metabolic tumour volumes in lymphoma: Does quantity matter? *Br J Haematol* 2022; 197(2):139–55. doi: 10.1111/bjh.18016.
  - 64. Barrington SF, Meignan M. Time to Prepare for Risk Adaptation in Lymphoma by Standardizing Measurement of Metabolic Tumor Burden. *J Nucl Med* 2019; 60(8):1096–102. doi: 10.2967/jnumed.119.227249.
  - 65. Boellaard R, Buvat I, Nioche C, Ceriani L, Cottreau A-S, Guerra L et al. International Benchmark for Total Metabolic Tumor Volume Measurement in Baseline 18F-FDG PET/CT of Lymphoma Patients: A Milestone Toward Clinical Implementation. *J Nucl Med* 2024; 65(9):1343–8. doi: 10.2967/jnumed.124.267789.
  - 66. Ferrández MC, Eertink JJ, Golla SSV, Wiegers SE, Zwezerijnen GJC, Pieplenbosch S et al. Combatting the effect of image reconstruction settings on lymphoma 18FFDG PET metabolic tumor volume assessment using various segmentation methods. *EJNMMI Res* 2022; 12(1):44. doi: 10.1186/s13550-022-00916-9.
  - 67. Zwezerijnen GJC, Eertink JJ, Burggraaff CN, Wiegers SE, Shaban EAIN, Pieplenbosch S et al. Interobserver Agreement on Automated Metabolic Tumor Volume Measurements of Deauville Score 4 and 5 Lesions at Interim 18F-FDG PET in Diffuse Large B-Cell Lymphoma. *J Nucl Med* 2021; 62(11):1531–6. Verfügbar unter: <https://pubmed.ncbi.nlm.nih.gov/33674403/>.
  - 68. Arimoto MK, Nakamoto Y, Higashi T, Ishimori T, Ishibashi M, Togashi K. Intra- and inter-observer agreement in the visual interpretation of interim 18F-FDG PET/CT in malignant lymphoma: influence of clinical information. *Acta Radiol* 2018; 59(10):1218–24. doi: 10.1177/0284185117751279.
  - 69. Coiffier B, Sarkozy C. Diffuse large B-cell lymphoma: R-CHOP failure-what to do? *Hematology. American Society of Hematology. Education Program* 2016; 2016(1):366–78. Verfügbar unter: <https://pubmed.ncbi.nlm.nih.gov/27913503/>.
  - 70. Albano D, Bosio G, Re A, Pagani C, Giubbini R, Bertagna F. Metabolic behavior and prognostic value of early and end of treatment 18F-FDG PET/CT in adult Burkitt's lymphoma: the role of Deauville and IHP criteria. *Leuk Lymphoma* 2019; 60(2):326–33.
  - 71. Han EJ, O JH, Yoon H, Jung SE, Park G, Choi BO et al. FDG PET/CT response in diffuse large B-cell lymphoma: Reader variability and association with clinical outcome. *Medicine (Baltimore)* 2016; 95(39):e4983. doi: 10.1097/MD.0000000000004983.
  - 72. Fallanca F, Alongi P, Incerti E, Gianolli L, Picchio M, Kayani I et al. Diagnostic accuracy of FDG PET/CT for clinical evaluation at the end of treatment of HL and NHL: a comparison of the Deauville Criteria (DC) and the International Harmonization Project Criteria (IHPC). *Eur J Nucl Med Mol Imaging* 2016; 43(10):1837–48.

73. Kumar R, Krishnaraju VS, Mittal BR, Shankar K, Singh H, Ramachandran A et al. Pathologic Validation of Deauville Score-Based Disease on F-18 FDG PET/CT after First-Line Treatment in Patients with Lymphoma. *Nucl Med Mol Imaging* 2024;1–9.
74. Lee JW, Oh D, Eom K-Y, Kim JH, Kim WC, Chung MJ et al. The prognostic value of PET/CT evaluation with Deauville score on the recurrence and survival in diffuse large B-cell lymphoma: a multi-institutional study of KROG 17-02. *Clin Exp Metastasis* 2020; 37(1):125–31. Verfügbar unter: <https://link.springer.com/article/10.1007/s10585-019-09992-z>.
75. Karimjee M, Delaby G, Huglo D, Baillet C, Willaume A, Dujardin S et al. Evaluation of a convolution neural network for baseline total tumor metabolic volume on 18FFDG PET in diffuse large B cell lymphoma. *Eur Radiol* 2023; 33(5):3386–95. doi: 10.1007/s00330-022-09375-1.
76. Burggraaff CN, Rahman F, Kaßner I, Pieplenbosch S, Barrington SF, Jauw YWS et al. Optimizing Workflows for Fast and Reliable Metabolic Tumor Volume Measurements in Diffuse Large B Cell Lymphoma. *Mol Imaging Biol* 2020; 22(4):1102–10. doi: 10.1007/s11307-020-01474-z.
77. Hasani N, Paravastu SS, Farhadi F, Yousefirizi F, Morris MA, Rahmim A et al. Artificial Intelligence in Lymphoma PET Imaging:: A Scoping Review (Current Trends and Future Directions). *PET Clin* 2022; 17(1):145–74.
78. Hadjiiski L, Cha K, Chan H-P, Drukker K, Morra L, Näppi JJ et al. AAPM task group report 273: Recommendations on best practices for AI and machine learning for computer-aided diagnosis in medical imaging. *Med Phys* 2023; 50(2):e1-e24. doi: 10.1002/mp.16188.
79. Yousefirizi F, Klyuzhin IS, O JH, Harsini S, Tie X, Shiri I et al. TMTV-Net: fully automated total metabolic tumor volume segmentation in lymphoma PET/CT images - a multi-center generalizability analysis. *Eur J Nucl Med Mol Imaging* 2024; 51(7):1937–54.
80. Veziroglu EM, Farhadi F, Hasani N, Nikpanah M, Roschewski M, Summers RM et al. Role of Artificial Intelligence in PET/CT Imaging for Management of Lymphoma. *Semin Nucl Med* 2023; 53(3):426–48.
81. Häggström I, Leithner D, Alvén J, Campanella G, Abusamra M, Zhang H et al. Deep learning for 18Ffluorodeoxyglucose-PET-CT classification in patients with lymphoma: a dual-centre retrospective analysis. *Lancet Digit Health* 2024; 6(2):e114-e125. doi: 10.1016/S2589-7500(23)00203-0.
82. Salomon T, Nganoa C, Gac A-C, Fruchart C, Damaj G, Aide N et al. Assessment of alteration in liver 18F-FDG uptake due to steatosis in lymphoma patients and its impact on the Deauville score. *Eur J Nucl Med Mol Imaging* 2018; 45(6):941–50. doi: 10.1007/s00259-017-3914-y.
83. Ceriani L, Suriano S, Ruberto T, Zucca E, Giovanella L. 18F-FDG uptake changes in liver and mediastinum during chemotherapy in patients with diffuse large B-cell lymphoma. *Clin Nucl Med* 2012; 37(10):949–52.

84. Itti E, Meignan M, Berriolo-Riedinger A, Biggi A, Cashen AF, Véra P et al. An international confirmatory study of the prognostic value of early PET/CT in diffuse large B-cell lymphoma: comparison between Deauville criteria and  $\Delta$ SUVmax. *Eur J Nucl Med Mol Imaging* 2013; 40(9):1312–20. doi: 10.1007/s00259-013-2435-6.
85. Ng DZ, Lee CY, Lam WW, Tong AK, Tan SH, Khoo LP et al. Prognostication of diffuse large B-cell lymphoma patients with Deauville score of 3 or 4 at end-of-treatment PET evaluation: a comparison of the Deauville 5-point scale and the  $\Delta$ SUVmax method. *Leuk Lymphoma* 2022; 63(1):256–9. doi: 10.1080/10428194.2021.1992624.
86. Keramida G, Potts J, Bush J, Dizdarevic S, Peters AM. Hepatic steatosis is associated with increased hepatic FDG uptake. *Eur J Radiol* 2014; 83(5):751–5.
87. Sher A, Lacoeuille F, Fosse P, Vervueren L, Cahouet-Vannier A, Dabli D et al. For avid glucose tumors, the SUV peak is the most reliable parameter for (18)FFDG-PET/CT quantification, regardless of acquisition time. *EJNMMI Res* 2016; 6(1):21. doi: 10.1186/s13550-016-0177-8.
88. Kaalep A, Sera T, Rijnsdorp S, Yaqub M, Talsma A, Lodge MA et al. Feasibility of state of the art PET/CT systems performance harmonisation. *Eur J Nucl Med Mol Imaging* 2018; 45(8):1344–61. doi: 10.1007/s00259-018-3977-4.

## 8. Danksagung

Zahlreiche Personen haben in unterschiedlichster Weise zum erfolgreichen Gelingen dieser Arbeit beigetragen. Bei all jenen möchte ich mich an dieser Stelle herzlich bedanken.

Mein großer Dank gilt meinem Doktorvater und Betreuer Herrn Prof. Dr. Dirk Hellwig für die Bereitstellung des Themas und das damit gesetzte Vertrauen in mich. Sein unermüdliches Engagement, seine herausragende Expertise und seine freundliche und wertschätzende Unterstützung haben mich stets durch die Verfassung der vorliegenden Forschungsarbeit geleitet. Die inspirierenden Gespräche, die gelehrten Kenntnisse sowie die übertragene Faszination für die bearbeitete Thematik werden mich definitiv auch auf meinem weiteren Weg begleiten – vielen Dank!

Weiters möchte ich mich bei allen Mitarbeitern der Abteilung für Nuklearmedizin des Universitätsklinikums Regensburg bedanken, welche mir während der vorausgehenden Famulatur durch ihre freundliche Aufnahme und die gewonnenen fachlichen Erfahrungen einen spannenden, praktischen Einstieg in das Fachgebiet der Nuklearmedizin ermöglicht haben.

An dieser Stelle möchte ich auch Frau Beate Zalesky aus dem Sekretariat der Abteilung für Nuklearmedizin für die stets freundliche und unkomplizierte Möglichkeit zur Terminvereinbarung danken.

Besonderer Dank gilt auch meinen Eltern und meinem Bruder Tobias für ihre liebevolle Unterstützung auf fachlicher, technischer sowie emotionaler Ebene, für ihre ermutigenden Worte und für ihr beständiges Vertrauen in mich.

Auch meinen Mitdoktoranden Michael und Lukas möchte ich für die gute Zusammenarbeit meinen Dank aussprechen.

Abschließend möchte ich mich bei meinen Freunden für den motivierenden Austausch und das unterstützende Miteinander während der schönen gemeinsamen Zeit bedanken.