

Philipps



Universität
Marburg

Aus der Klinik für Strahlendiagnostik

Geschäftsführender Direktor: Prof. Dr. med. Klaus J. Klose
des Fachbereichs Medizin der Philipps-Universität Marburg

in Zusammenarbeit mit dem Universitätsklinikum Gießen und Marburg GmbH,
Standort Marburg

Softwareunterstützte semiautomatische Tumorvolumetrie von Lungentumoren

Inaugural-Dissertation zur Erlangung des Doktorgrades der Humanmedizin

dem Fachbereich Medizin der Philipps-Universität Marburg

vorgelegt von

Katharina Haag

aus Frankfurt am Main

Marburg, 2011

Angenommen vom Fachbereich Medizin der Philipps-Universität Marburg am:
12.12.2011

Gedruckt mit Genehmigung des Fachbereichs.

Dekan: Prof. Dr. Matthias Rothmund

Referent: Prof. Dr. Dr. Johannes Heverhagen

Korreferent: PD Dr. Gerd Strassmann

Gewidmet meinen Eltern

Annette Haag, geb. Baumann

und

Dr. Detlev Haag

Inhaltsverzeichnis

| | | |
|----------|---|-----------|
| 1 | Einleitung..... | 1 |
| 1.1 | Quantifizierung von Lungentumoren | 1 |
| 1.2 | WHO - Kriterien..... | 2 |
| 1.3 | RECIST - Kriterien 1.0 | 4 |
| 1.4 | RECIST - Kriterien 1.1 | 8 |
| 1.5 | Grundlagen der Computertomographie..... | 11 |
| 1.6 | Volumetrie..... | 15 |
| 1.7 | Fragestellung | 18 |
| 2 | Material und Methoden..... | 19 |
| 2.1 | Patientenkollektiv | 19 |
| 2.2 | Computertomographie | 20 |
| 2.3 | Methoden zur Bestimmung der Tumorlast..... | 21 |
| 2.3.1 | Manuelle Messung | 21 |
| 2.3.2 | Semiautomatische Volumetrie | 21 |
| 2.4 | Statistische Auswertung | 23 |
| 3 | Ergebnisse | 25 |
| 3.1 | Durchmesser und Volumen der Lungenläsionen | 25 |
| 3.2 | Vergleich der Einteilung nach RECIST-Schwellenwerten (30/20%) und an RECIST angepassten Schwellenwerten (65/73%) zur Existenzabschätzung von Schwellenwerten/Referenzbereichen und zur Beurteilung des Unterschieds zwischen den Klassifikationen..... | 26 |
| 3.3 | Bewertung der Übereinstimmung der Messwerte bei gleicher Technik und Methode aber wechselnden Beobachtern (Reproduzierbarkeit) | 31 |
| 3.3.1 | Vergleich der manuellen Messung des größten Durchmessers von zwei Untersuchern | 32 |
| 3.3.2 | Vergleich der semiautomatischen Messung des größten Durchmessers von zwei Untersuchern..... | 35 |

| | | |
|----------|--|-----------|
| 3.3.3 | Tabellarische Übersicht der Ergebnisse | 39 |
| 3.4 | Bewertung der Übereinstimmung der Messwerte bei unterschiedlicher Technik und Methode und gleichen Beobachtern..... | 40 |
| 3.4.1 | Vergleich der manuellen mit der semiautomatischen Messung von Untersucher 1 | 40 |
| 3.4.2 | Vergleich der manuellen mit der semiautomatischen Messung von Untersucher 2 | 43 |
| 3.4.3 | Tabellarische Übersicht der Ergebnisse | 47 |
| 3.5 | Interaktionszeitanalyse mit Hinblick auf die Praktikabilität | 48 |
| 4 | Diskussion | 49 |
| 5 | Zusammenfassung..... | 55 |
| 5.1 | Zusammenfassung | 55 |
| 5.2 | Summary | 57 |
| 6 | Literaturverzeichnis..... | 59 |
| 7 | Abbildungsverzeichnis/Tabellenverzeichnis | 64 |
| 8 | Verzeichnis der akademischen Lehrer | 68 |
| 9 | Danksagung | 69 |

Abkürzungsverzeichnis

| | |
|-----------------|--|
| bspw. | Beispielsweise |
| bzw. | Beziehungsweise |
| ca. | Circa |
| CMEA | Council for Mutual Economic Assistance |
| CR | Complete Remission |
| CT | Computertomographie |
| d.h. | das heißt |
| EORTC | European Organization for Research and Treatment of Cancer |
| <i>et al.</i> | und andere |
| ggf. | gegebenenfalls |
| HE | Houndsfieldeinheiten |
| <i>in vitro</i> | Latein „im Glas“ |
| <i>in vivo</i> | Latein: „am lebenden Objekt“ |
| KI | Konfidenzintervall |
| KV | Kilovolt |
| mAs | Milliampersekunden |
| max. | Maximal |
| min | Minimal |
| ml | Milliliter |
| mm | Millimeter |
| MRT | Magnetresonanztomographie |
| n | Fallzahl |
| NC | No Change |
| NCI | National Cancer Institute |
| NSCLC | Non Small Cell Carcinoma |
| p | Statistische Signifikanz |
| PACS | Picture Archiving System |
| PD | Progressive Disease |
| PR | Partial Response |
| R | Korrelationskoeffizient |
| RECIST | Response Criteria In Solid Tumors |
| SCLC | Small Cell Carcinoma |

| | |
|------|---------------------------------------|
| s | Sekunden |
| SD | Standardabweichung |
| SD | Stable Disease |
| sog. | sogenannt |
| u.a. | unter anderem |
| UICC | Union Internationale Contre le Cancer |
| VOI | Volume of Interest |
| vs. | Versus, Latein: "gegenüber" |
| WHO | World Health Organization |
| z.B. | zum Beispiel |
| z.T. | zum Teil |
| % | Prozent |

1 Einleitung

1.1 Quantifizierung von Lungentumoren

Die geschätzte Inzidenz von Tumorneuerkrankungen lag im Jahr 2008 weltweit bei ca. 12.4 Millionen, wovon ca. 7.6 Millionen Menschen ihrer Tumorerkrankung erlagen. Europaweit belief sich die Zahl der Tumorinzidenzen auf ca. 3.4 Millionen, die der Todesfälle durch maligne Tumore auf ca. 1.8 Millionen. Sowohl europaweit, als auch weltweit, war die häufigste Krebsart bei Männern der Lungenkrebs und bei Frauen der Brustkrebs (WHO 2008).

Um eine genaue Vorstellung über die Lokalisation und die Ausbreitung eines Tumors zu bekommen, werden in der Onkologie multiple bildgebende Verfahren genutzt. Am häufigsten ist die computertomographische Untersuchung, die sowohl zum Tumorstaging als auch zur Therapiekontrolle einen großen Beitrag leistet (Hensing 2005; Collins, Haines et al. 2007; Erasmus and Sabloff 2008; Klein and Braff 2008).

Hierbei ist die Quantifizierung von Tumorgrößen eine zentrale klinische Fragestellung, die bei Diagnose, Therapieplanung und Therapiemonitoring sowohl in der Routine als auch insbesondere in klinischen Studien gestellt wird. Wie diese Quantifizierung am besten und am effizientesten erfolgen soll, wird bis heute intensiv und kontrovers diskutiert (James, Eisenhauer et al. 1999; Fraioli, Bertoletti et al. 2006; Gottlieb, Litwin et al. 2008; Levy and Rubin 2008).

Um die Tumorlast eines Patienten quantitativ bestimmen zu können, wurden verschiedene Kriterien entwickelt, die nun näher vorgestellt werden.

1.2 WHO - Kriterien

Aufgrund expandierender Forschung in der Onkologie und einer steigenden Anzahl an Studien in diesem Bereich, gründete die World Health Organization (WHO) eine Initiative zur Vereinheitlichung von Mess- und Bewertungskriterien um vergleichbare Resultate bei der Verlaufsbeurteilung von Tumoren zu ermöglichen.

Unter der Schirmherrschaft der WHO trafen sich Vertreter verschiedener Organisationen, darunter die European Organization for Research and Treatment of Cancer (EORTC), das National Cancer Institute (NCI), die Union Internationale Contre le Cancer (UICC) und der Council for Mutual Economic Assistance (CMEA). Sie veröffentlichten nach mehrjähriger Zusammenarbeit 1979 das „WHO Handbook For Reporting Results On Cancer Treatment“ (WHO 1979). Diese Richtlinien sollten als internationaler Standard für Studien und die klinische Routine dienen.

Das Messverfahren der WHO-Kriterien basiert auf der Multiplikation des größten Längsdurchmessers des Tumors mit dem darauf senkrecht stehenden größten Durchmesser (siehe Abbildung 1). Bei multiplen Läsionen wird die Summe aller Einzelwerte berechnet.

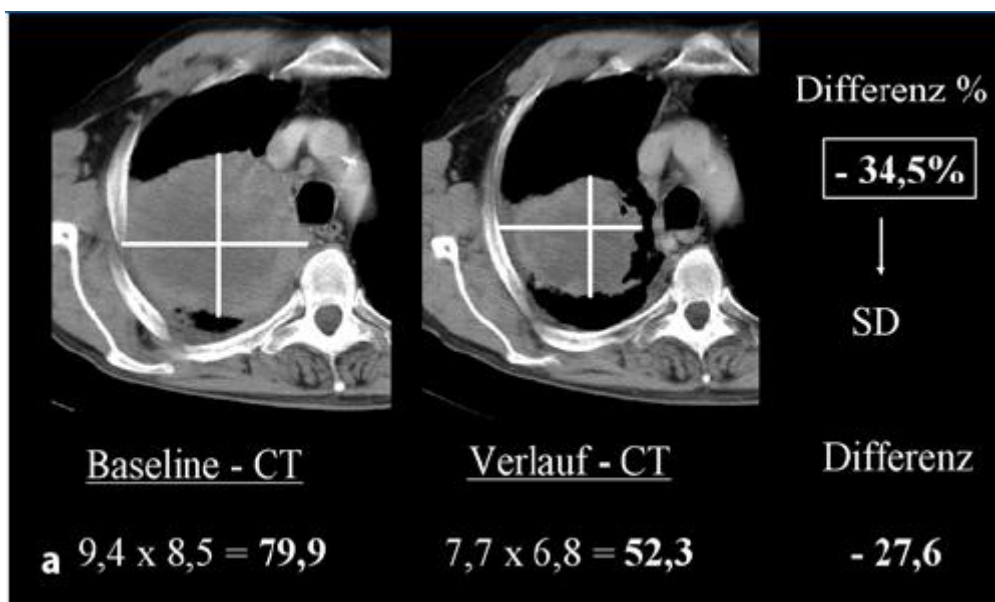


Abbildung 1: Bidimensionale Messung eines Bronchialkarzinoms im Verlauf nach WHO-Kriterien. Der Vergleich der bidimensionalen Messergebnisse der Ausgangs-CT (1. Spalte) und der Verlaufs-CT (2. Spalte) ergab mit einer Reduktion der Tumorgröße von 34,5% nach WHO-Kriterien SD (stable disease) (Hauth, Stattaus et al. 2007).

Die erste Messung wird vor Therapiebeginn vorgenommen um einen Ausgangswert (sog. baseline) definieren zu können. Zum Vergleich der Messergebnisse vor Therapiebeginn mit den darauf folgenden gibt es verschiedene Kategorien des Therapieansprechens (sog. Responsekategorien). Als eine komplette Remission (CR) gilt das vollständige Verschwinden aller Tumormassen, bestätigt durch zweimalige Kontrollen in einem Abstand von mindestens vier Wochen. Die partielle Remission (PR) wird durch die Reduktion aller Tumormassen um mindestens 50% und die Abwesenheit neu aufgetretener Läsionen definiert, ebenfalls kontrolliert durch zwei Untersuchungen in einem vierwöchigen Abstand. „No change“ (NC) oder „stable disease“ (SD) werden verwendet, wenn es weder eine 50 prozentige Reduktion noch eine Vergrößerung der Tumormassen um 25% zu verzeichnen gibt. Die Progression der Tumorerkrankung, „progressive disease“ (PD), entspricht einer Zunahme der Größe einer oder mehrerer Läsionen um mindestens 25% und bzw. oder das Auftreten neuer Läsionen (siehe Tabelle 1).

Tabelle 1: Responsekategorien nach WHO (1979).

| | | Ziel - Läsionen |
|----------------------------|----|---|
| Complete Response | CR | Komplettes Verschwinden aller detektierten Läsionen |
| Partial Response | PR | ≥ 50% Abnahme des Produktes in einer oder mehrerer der evaluierten Läsionen |
| Stable Disease | SD | Weder Remission noch Progression |
| Progressive Disease | PD | ≥ 25% Zunahme des Produktes in einer oder mehreren evaluierten Läsionen oder Auftreten neuer Läsionen |

1.3 RECIST - Kriterien 1.0

Die oben beschriebenen WHO-Kriterien waren lange Zeit weltweit die am meisten genutzten einheitlichen Richtlinien zur Verlaufskontrolle von Tumorgrößen. Im Laufe der Zeit traten jedoch Probleme bei der Nutzung auf, da zum Beispiel die minimale Größe einer Läsion oder auch die gemessene Anzahl von Tumorknoten unter den Forschungsgruppen stark variierte. Weiterhin wurde die Definition einer Progression der Krankheit von einigen an einer einzigen Läsion festgemacht, andere wiederum nutzten die Summe aller gemessenen Läsionen. Zudem führte das Fortschreiten der CT- und MR-Technik zu der Fragestellung, ob dreidimensionale Werte in die Responsekriterien eingeführt werden sollten.

Aufgrund der unterschiedlichen Interpretationen und der Unsicherheiten bezüglich der WHO-Kriterien waren diese nicht länger als international vergleichbarer Standard anzusehen.

1994 bildete die EORTC, das NCI und das National Cancer Institute of Canada erneut eine Arbeitsgruppe um die existierenden Kriterien zu reevaluieren. Nach mehrjähriger Zusammenarbeit, multiplen Versionen und einigen Konsensustreffen wurde die endgültige Version im Juni 1999 beim Journal of the National Cancer Institute als „Response Evaluation Criteria In Solid Tumors“ (RECIST) eingereicht und im Februar 2000 veröffentlicht. Diese Kriterien basierten auf retrospektiven Daten von über 4000 Patienten, die in gemeinsamen Studien von Europa und den USA zusammengetragen wurden.

Ziel der Arbeitsgruppe war es nicht, komplett neue Kriterien zu erstellen, sondern nur eine Revision der WHO-Kriterien zu schaffen, die die konfliktträchtigen Areale adressierte und die Unsicherheiten klärte. Deshalb wurden bei den neuen RECIST-Kriterien auch weiterhin die vier Responsekategorien „complete response“, „partial response“, „stable disease“, und „progressive disease“ beibehalten. Um zukünftige Forschungsergebnisse mit den momentanen Standards vergleichen zu können, wurde die Definition der partiellen Remission ebenfalls nicht verändert. Auch wenn zu diesem Zeitpunkt die Technologie weltweit vorangeschritten war, so dass zum Beispiel Änderungen in Tumolvolumen oder Tumormetabolismus messbar gewesen wären, war es nicht im Sinne der Forschungsgruppe die Messbarkeit von Läsionen zu präzisieren, sondern nur die Standardisierung und Simplifizierung der Methodik.

Das Vorgehen nach den RECIST-Leitlinien stellte sich wie folgt dar:

Zunächst musste zwischen messbaren und nicht-messbaren Läsionen unterschieden werden. Messbare Läsionen wurden mit einer Größe ≥ 20 mm in der konventionellen Röntgentechnik, oder einer Größe ≥ 10 mm in der Computertomographie definiert. Nicht-messbare Läsionen waren alle, die nicht in die beiden oben genannten Kategorien fielen oder Läsionen, die wirklich nicht quantifizierbar waren, wie zum Beispiel Knochenläsionen, Meningeosis und Lymphangiosis carcinomatosa, Aszites, Pleura- oder Perikardergüsse, inflammatorischer Brustkrebs und zystische Läsionen. Als beste diagnostische Mittel wurden die Computertomographie und die Magnetresonanztomographie (MRT) empfohlen, wobei die Schichtdicke 10 mm oder weniger, und beim Spiral-CT 5 mm betragen sollte. Generell sollte die Schichtdicke mindestens doppelt so groß sein wie der Durchmesser der kleinsten evaluierten Läsion und die ausgewählten Läsionen sollten sowohl repräsentativ für den Krankheitsverlauf als auch gut darstellbar sein. Die Gabe eines intravenösen Kontrastmittels wurde für die Untersuchung mittels Computertomographie empfohlen um vaskuläre Strukturen von angrenzendem Gewebe besser unterscheiden zu können. Besonders bei der computertomographischen Untersuchung des Thorax sollten die Läsionen in jeder der nachfolgenden Untersuchungen in dem jeweils gleichen Fenster (Lungen- oder Weichteilfenster) und Rekonstruktionsalgorithmus gemessen werden.

Zu Beginn der Dokumentation (Baseline) sollten maximal zehn Läsionen, wovon sich maximal fünf in demselben Organ befinden sollten, als Target-Läsionen ausgewählt werden. Eine exakte Anzahl an Läsionen wurde in der Leitlinie nicht gewählt. Die Summe des längsten Durchmessers aller Target-Läsionen wurde als Baseline-Durchmesser bestimmt. Alle anderen Läsionen wurden als Nontarget-Läsionen festgesetzt und ihr Vorhandensein ebenfalls dokumentiert, wobei die genaue Messung dieser Läsionen nicht als notwendig erachtet wurde. Die Baseline-Dokumentation sollte möglichst zeitnah zu Beginn der Therapie, spätestens aber vier Wochen nach Therapiebeginn, erfolgen. Außerdem wurde eine Reevaluation durch einen zweiten Beobachter empfohlen.

Die Messung der Tumorlast wurde im Vergleich zu den WHO-Kriterien nun dahingehend vereinfacht, dass nur noch der längste axiale Durchmesser einer Läsion zur Vermessung (unidimensional) herangezogen wurde und als Gesamtwert korrespondierend die Summe aller längsten Durchmesser sämtlicher Target-Läsionen errechnet wurde (siehe Abbildung 2).

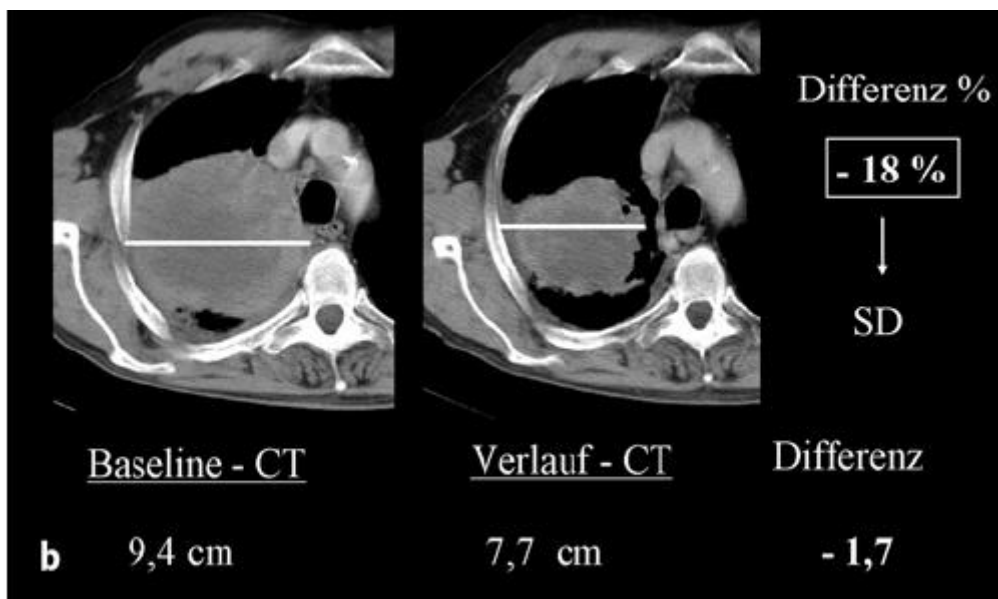


Abbildung 2: Unidimensionale Messung des Bronchialkarzinoms im Verlauf nach RECIST. Der Vergleich des unidimensionalen Messergebnisses der Ausgangs-CT (1. Spalte) und der Verlaufs-CT (2. Spalte) ergab mit einer Reduktion der TumorgroÙe von 18% nach RECIST-Kriterien SD (stable disease) (Hauth, Stattaus et al. 2007).

Die Einteilung der Responsekriterien verlief analog zu den WHO-Kriterien mit der Ausnahme, dass eine partielle Remission durch RECIST mit einer Abnahme von nur noch mindestens 30% der GröÙe definiert wurde. Eine Progression wurde von RECIST nun bei einer Zunahme der GröÙe um 20% erklärt (siehe Tabelle 2).

Tabelle 2: Responsekategorien nach RECIST 1.0 (2000).

| | | Target-Läsionen | Nontarget-Läsionen |
|----------------------------|----|---|--|
| Complete Response | CR | Vollständiges Verschwinden aller Läsionen | Vollständiges Verschwinden aller Läsionen |
| Partial Response | PR | ≥ 30% Reduktion der Summe der Durchmesser | ---- |
| Stable Disease | SD | Weder Progression noch Regression | Verbleib ≥ 1 Läsion(en) |
| Progressive Disease | PD | ≥ 20% Zunahme der Summe der Durchmesser Oder ≥ 1 neue Läsion | ≥ 1 neue Läsion oder deutliche Vergrößerung der Läsionen |

Die Frequenz einer Reevaluation des Krankheitsverlaufs sollte protokollspezifisch bestimmt werden. Von RECIST wurde eine Empfehlung für eine Follow-Up-Untersuchung in einem Zeitrahmen von 6–8 Wochen ausgesprochen. Für die Auswertung des Gesamtansprechens wurden jeweils Target-, Nontarget- und neue Läsionen evaluiert (siehe Tabelle 3).

Tabelle 3: Beurteilung des Gesamtansprechens.

| Target-Läsionen | Non-Target-Läsionen | Neue Läsion | Gesamtansprechen |
|------------------------|----------------------------|--------------------|-------------------------|
| CR | CR | nein | CR |
| CR | nicht CR/nicht PD | nein | PR |
| PR | nicht PD | nein | PR |
| SD | nicht PD | nein | SD |
| PD | egal | egal | PD |
| egal | PD | egal | PD |
| egal | egal | ja | PD |

Wenn im Rahmen der Reevaluation der Status einer Remission, partiell oder komplett, aufgestellt wurde, so sollte dies durch eine erneute Untersuchung höchstens vier Wochen später bestätigt werden (Therasse, Arbuck et al. 2000).

Nachdem die RECIST-Kriterien im Jahre 2000 publiziert und weitläufig in akademischen Gruppen, der Forschung und Industrie akzeptiert wurden, ließ sich in vielen Studien zeigen, dass die unidimensionale Messung nach RECIST mit der bidimensionalen Messung nach WHO weitgehend kongruent war (James, Eisenhauer et al. 1999; Sohaib, Turner et al. 2000; Park, Lee et al. 2003; Prasad, Saini et al. 2003; Therasse, Le Cesne et al. 2005).

Hauth et al. untersuchten diese Fragestellung in einer prospektiven Studie an einem Patientenkollektiv (n = 80) mit Lungen- und Lebermetastasen (Hauth, Stattaus et al. 2007). Es zeigte sich eine hohe Übereinstimmung (κ -Index = 0.94) beider Messverfahren bei der Beurteilung des Therapieansprechens solider Tumoren.

Park et al. belegten in einer vierjährigen Multicenterstudie an 79 Patienten mit verschiedenen Tumorentitäten ebenfalls eine ähnliche Konkordanz von 92% zwischen der Messung nach WHO und RECIST (Park, Lee et al. 2003).

Eine der größten Studien wurde durch James et al. durchgeführt. An 569 Patienten in acht Phase II- und Phase III-Studien angewandte Messungen in beiden Verfahren zeigten einen κ -Koeffizienten von 0.95 und somit die höchste Übereinstimmung (James, Eisenhauer et al. 1999).

Entgegen dieser Aussage kommen Mazumdar et al. bei einer retrospektiven Analyse von 130 Patienten zu der Ansicht, dass beim Vergleich beider Systeme im Mittel zwischen 14 – 20% der Patienten unterschiedlich bewertet werden (Mazumdar, Smith et al. 2004). Neben dem Vergleich von bestehenden Studiendaten stellte die Forschungsgruppe außerdem eine Simulationsstudie ($n = 200$) auf um zu hinterfragen, ob die Form des Tumors eine Auswirkung auf die Messung und somit die Einteilung in die verschiedenen Responsekategorien nach WHO und RECIST hat. Bemerkenswertes Resultat dieser Untersuchung war, dass bis zu 35% der Tumoren, die ein irreguläres Wachstums entlang der Raumebenen aufwiesen, je nach Klassifikation unterschiedlich beurteilt wurden. Die Mehrzahl der Gruppe dieser Patienten wurde nach den WHO-Kriterien als „progressive disease“, nach RECIST jedoch als „stable disease“ eingestuft. Diese Studie weist darauf hin, dass bei der Evaluation irregulär wachsender Tumoren die Anwendung beider Konzepte unter Umständen zu einer unterschiedlichen Bewertung des Krankheitsverlaufes führen kann.

1.4 RECIST - Kriterien 1.1

Mittlerweile sind die RECIST-Kriterien die am häufigsten angewandten Responsekategorien in klinischen Studien, die sich mit neuen Methoden für die Behandlung solider Tumore befassen.

Die RECIST-Forschungsgruppe um Therasse et al. begann 2005 eine neue Evaluation der aktuellen Studiendaten, indem sie 60 Studien zwischen den Jahren 2000 und 2005 auswertete und in einem Review veröffentlichte (Therasse, Eisenhauer et al. 2006). Die Studien befassten sich entweder mit der generellen Implementierung von RECIST, des Vergleichs von WHO vs. RECIST oder der Anwendung der Kriterien bei bestimmten Tumorentitäten. Die Mehrheit der untersuchten Studien unterstützte die Nutzung von RECIST innerhalb gewöhnlicher Situationen.

In einer weiteren retrospektiven Studie im Jahr 2008 beschäftigte sich die Gruppe um Boegarts et al. spezifisch mit der Frage, wie hoch die maximale Anzahl an Läsionen sein sollte, die bei jeder Evaluation des Krankheitsverlaufs gemessen werden sollte (Bogaerts, Ford et al. 2009). Daten von 16 klinischen Studien mit Patienten (n = 6512) mit metastasiertem Tumorleiden flossen in diese Untersuchung ein. Bei der Begutachtung des Therapieansprechens ergab sich jeweils für die Messung von zehn, fünf oder drei Läsionen kein signifikanter Unterschied.

Basierend auf diesen Ergebnissen wurden die RECIST-Kriterien revidiert und im Januar 2009 als RECIST 1.1 veröffentlicht (Eisenhauer, Therasse et al. 2009). Die wichtigsten Änderungen bestehen in folgenden Punkten:

- Die minimale Größe einer Läsion muss nun 10 mm betragen, gemessen mittels Computertomographie, oder klinisch mittels Schieblehre, sofern möglich. Auch für Lymphknoten, die bei RECIST 1.0 nicht berücksichtigt wurden, gibt es definierte Größenangaben (siehe
- Tabelle 4). Knochen- und zystische Läsionen, die bisher als nicht-messbar galten, werden ebenfalls eingeschlossen.
- Durch die Studien von Bogaerts et al. (Bogaerts, Ford et al. 2009) konnte die zu messende Anzahl an Tumorknoten auf insgesamt fünf pro Patient und maximal zwei pro Organ reduziert werden.
- Die Definition einer Progression wurde zusätzlich zu der Vergrößerung um 20% um die absolute Zunahme der Summe aller Durchmesser um mindestens 5 mm erweitert.
- Non-Target-Läsionen können gleichfalls für die Bewertung herangezogen werden, sofern eine eindeutige Progression dieser Läsionen zu verzeichnen ist.
- Das Auftreten neuer Läsionen wird weiterhin als Progression betrachtet. Auch wenn Metastasen in einem Organ gefunden werden, welches in der Baseline-Untersuchung noch nicht bildgebend dargestellt wurde, werden diese Läsionen als neu und die Krankheit damit als fortschreitend betrachtet. Neue Befunde aus PET-Untersuchungen müssen durch einen CT-Befund bestätigt werden.
- Sollte eine Target-Läsion im Therapieverlauf zu klein werden, um angemessen evaluiert werden zu können, wird ihr ein Standardwert von 5 mm zugeteilt. Beim Verschwinden der Läsion fällt der Wert entsprechend auf 0 mm.

Tabelle 4: Vergleich alter und neuer RECIST-Kriterien.

| RECIST 1.0 | RECIST 1.1 |
|---|--|
| bis zu 10 Target Läsionen max. 5 pro Organ | bis zu 5 Target Läsionen max. 2 pro Organ |
| LK: Nicht angegeben | LK: Querdurchmesser > 15 mm = Target < 10 mm = normal |
| PD: 20% Zunahme | PD: 20% Zunahme und mindestens 5 mm absolut |
| Konfirmation für CR + PR erforderlich (spätestens nach 4 Wochen) | Nur wenn Response = primärer Endpunkt, nicht bei progressionsfreiem Überleben (PFS) als Endpunkt |
| | FDG-PET für neue Läsionen zulässig |

1.5 Grundlagen der Computertomographie

Die Computertomographie ist eine radiologische Schnittbilddarstellung des menschlichen Körpers. In den siebziger Jahren des letzten Jahrhunderts wurde dieses Verfahren in der radiologischen Diagnostik erstmals eingeführt.

Um Querschnittsbilder des Körpers anzufertigen, wird der Patient mittels einer Röntgenröhre aus verschiedenen Richtungen durchstrahlt. Das CT-System besteht aus einer Gantry mit Röntgenröhre und Detektoren, einem Patiententisch, einem Bildrechner und einer Konsole. Es erfolgen Absorptionsmessungen mit Detektoren, die gegenüber der Röntgenröhre und hinter dem Patienten angeordnet sind. Röntgenröhre und Detektorsystem bewegen sich in einer gleichsinnigen Rotationsbewegung, so dass aus einer 360° Drehung die räumliche Anordnung der absorbierenden Strukturen errechnet werden kann (Wegener 1996) (siehe Abbildung 3).

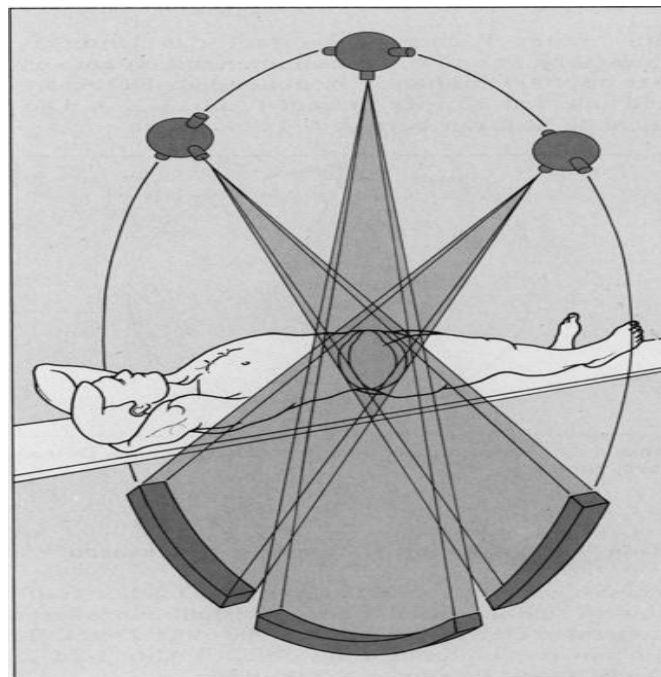


Abbildung 3: Prinzip der Datenerfassung bei der Computertomographie Projektionen einer Schicht werden aus vielen verschiedenen Richtungen senkrecht zur Körperachse durch ein Detektor-System registriert (Wegener 1996).

Die in verschiedenen Graustufen kodierten Schwächungswerte der durchstrahlten Gewebe werden in ein axiales Schnittbild umgesetzt, indem mit Hilfe der Radon-Transformation bzw. der Fourier-Transformation computergestützt das CT-Bild berechnet wird. Es wird zumeist eine Bildauflösung von 512 x 512 Bildpunkten verwendet und die Ergebnisse in Form einer Zahlenmatrix gespeichert. Zahlenwerte zwischen 0 und $2^{12} = 4096$ werden jedem einzelnen Bildpunkt zugeordnet und in einem weiteren Schritt in analoge Grautöne umgesetzt.

Die kleinste Einheit des Tomogramms ist das Pixel. Gleichzeitig repräsentiert jedes Bildelement abhängig von Schichtdicke, Matrixgröße und Scanfelddurchmesser zusätzlich ein Volumenelement (Voxel) (Hofer 2006) (siehe Abbildung 4).

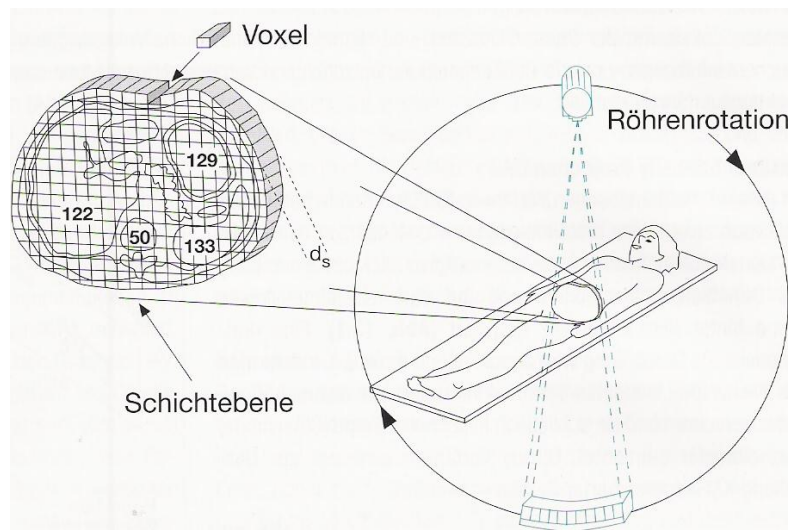


Abbildung 4: Die Schichtebene besteht aus einer Matrix aus gleich großen Quadern (Voxel) mit einer quadratischen Grundfläche und einer definierten Schichtdicke (d_s) (Hofer 2006).

Der spezifische Dichtewert eines Voxels wird nun durch das in ihm enthaltene Gewebe und der entsprechenden Schwächung der Röntgenstrahlung bestimmt. Die Schwächungswerte werden in Hounsfield-Einheiten (HE) angegeben. Der Dichte von Wasser wurde willkürlich der Wert 0 HE und der Dichte von Luft -1000 HE zugeordnet. Die Werte verschiedener Gewebearten werden in Relation zu diesen beiden festgelegten Werten gesetzt. Es sind willkürlich festgesetzte Relativwerte (Wegener 1996) (siehe Abbildung 5).

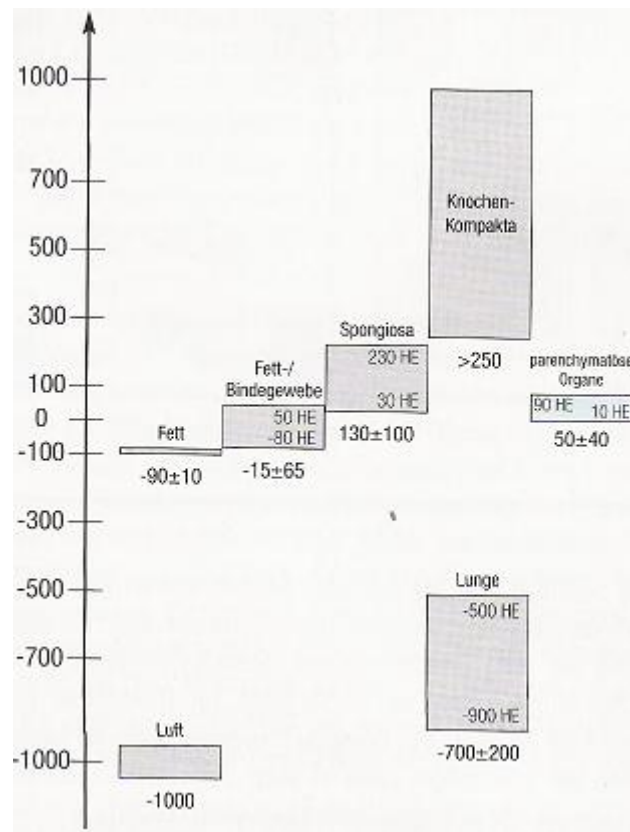


Abbildung 5: Dichtewerte unterschiedlicher Gewebearten angegeben in Hounsfield-Einheiten (Hofer 2006).

Die Gewebedichte wird bei der Beschreibung eines CT-Bilds in Zusammenhang zum umgebenden Gewebe angegeben, also bei gleicher Dichte isodens, bei höherer Dichte hyperdens und bei geringerer Dichte hypodens.

Da das menschliche Auge nur zwischen etwa 20 – 50 Graustufen unterscheiden kann, ist es hilfreich bei der Darstellung einen bestimmten Dichtebereich zu wählen, das sog. Bildfenster. In diesem werden nur die Dichtewerte des gewählten Bereichs angezeigt, Werte oberhalb erscheinen weiß, Werte unterhalb davon schwarz. Das Fenster kann je nach diagnostischer Fragestellung frei gewählt werden oder aus gespeicherten Voreinstellungen als Knochen-, Weichteil- oder Lungenfenster dargestellt werden.

Durch intravenöse oder orale Kontrastmittelgaben lässt sich eine Dichteanhebung der Blutgefäße bzw. des Magen-Darm-Traktes erreichen, um die Abgrenzbarkeit benachbarter Organe, Muskeln oder Knochen zu verbessern.

Seit der Erfindung der konventionellen CT wurde die Technik über die Spiral- und weiter zur Mehrschicht-Spiral-CT (MS-CT) entwickelt. Im Gegensatz zur konventionellen Technik, bei der Schicht für Schicht erarbeitet wird, gibt es bei der Spiral-CT einen kontinuierlichen Tischvorschub, während die Röntgenröhre um den Patienten kreist (Hofer 2006) (siehe Abbildung 6).

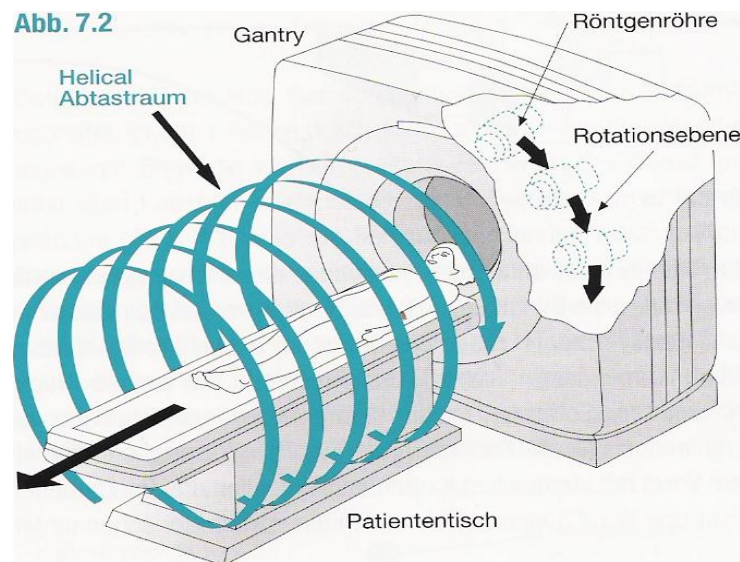


Abbildung 6: Bei der einzeiligen (SSCT = single slice CT) oder mehrzeiligen (MDCT = multi detector CT) Spiral-CT erfolgt die Abtastung kontinuierlich während die Liege gefahren wird (Hofer 2006).

Es entsteht ein einziger Volumen-Datensatz, aus welchem retrospektiv die einzelnen Schichten errechnet werden. Durch die schnelle Messung in einer Atemanhaltephase werden Bewegungsartefakte größtenteils minimiert. Mehrzeilendetektoren sind effektiver in der Nutzung der Röntgenleistung als einzeilige Detektoren. Die simultane Dokumentation mehrerer Schichten verringert die Aufnahmezeit signifikant oder erlaubt die Aufnahme kleinster Details bei kurzer Aufnahmedauer (Uffmann and Prokop 2001).

1.6 Volumetrie

In den letzten Jahren finden softwaregestützte Programme in der Auswertung der Computertomographie des Thorax sowohl zur Detektion als auch zur metrischen Erfassung von Metastasen zunehmend Verwendung. Die dreidimensionale Vermessung, d.h. die Volumetrie, bietet im Vergleich zu den ein- oder zweidimensionalen Messungen den Vorteil, ein irreguläres Tumorwachstum genauer erfassen zu können.

Bereits kurz nach Einführung der Computertomographie wurde die Möglichkeit erkannt Volumina mittels der CT-Technik vermessen zu können. Breiman et al. evaluierten die computergestützte Volumetrie durch Untersuchung von Phantomen sowie Organvolumina von Tieren (Leber, Milz) (Breiman, Beck et al. 1982). Es zeigte sich eine hohe Korrelation ($R = 0.9984$) zwischen dem computertomographisch gemessenen Volumen und dem realen Volumen, welches durch Wasserverdrängung ermittelt wurde.

Besonders in den letzten Jahren konzentrierten sich viele Forschungsgruppen auf die Optimierung der Volumetrie und eine mögliche Integration in den klinischen Alltag.

Verschiedene Rekonstruktions- und Betrachtungsparameter wurden auf ihre Auswirkung in der Volumenquantifizierung untersucht. Vogel et al. beschäftigten sich mit der Genauigkeit der Volumetrie pulmonaler Rundherde bei verschiedenen Schichtdicken (Vogel, Vonthein et al. 2008). Bei 28 Patienten wurde eine Thorax- und Abdomenspirale jeweils in 1, 3 und 5 mm Schichtdicke rekonstruiert. Die volumetrischen Vermessungen zeigten sich bei den Rekonstruktionen mit 1 mm und 3 mm Schichtdicke vergleichbar genau. Disler et al. konnten schon 1994 zeigen, dass mit Abnahme der Schichtdicke eine Abnahme des mittleren Fehlers bei der Volumenquantifizierung folgt (Disler, Marr et al. 1994).

An irregulär konfigurierten Phantomen evaluierten Van Hoe et al. welchen Einfluss die Einstellung von Fensterweite und Fensterzentrum auf eine Volumenuntersuchung hat. Bei verschiedenen Fenstereinstellungen zeigten sich dadurch bedingte Unterschiede in der Volumetrie zwischen 70 und 110% (Van Hoe, Haven et al. 1997).

Um anhand von Volumenmessungen kleiner pulmonaler Herde deren Wachstum und Malignitätsgrad zu bestimmen, untersuchten Yankelevitz et al. sphärische Phantome, verformbare Silikon-Phantome und Patienten ($n = 13$) mit Lungentumoren

verschiedener Größen und Formen (Yankelevitz, Reeves et al. 2000). Die Genauigkeit der Volumetrie der Phantome lag bei $\pm 3\%$, die Tumorverdopplungszeit lag bei den malignen Tumoren bei 177 Tagen, bei den benignen Herden bei 396 Tagen oder darüber.

Wang et al. evaluierten retrospektiv 4225 Lungentumore in 2239 Patienten um den Effekt von Morphologie, Größe und Ort einer Läsion auf die Volumetrie zu bestimmen (Wang, van Klaveren et al. 2008). Besonders juxtavaskuläre und irregulär geformte Läsionen beeinflussten die Variabilität der Volumenmessung, wohingegen intraparenchymal gelegene und glatt berandete Läsionen das beste Messergebnis aufwiesen.

Der Stellenwert der Volumetrie bei einer klinischen Verlaufsbeurteilung solider Tumore wird in der Literatur bisher kontrovers diskutiert (Hopper, Kasales et al. 1996; Dachman, MacEneaney et al. 2001; Werner-Wasik, Xiao et al. 2001; Prasad, Saini et al. 2003). Einige Studien sehen in der Anwendung der Volumetrie im klinischen Alltag keinen Gewinn (Sohaib, Turner et al. 2000; Dachman, MacEneaney et al. 2001), was jedoch unter Berücksichtigung der technischen Weiterentwicklung differenziert betrachtet werden sollte. Marten et al. kamen einige Jahre später zu dem Schluss, dass es im Grunde keine Indikation mehr für die manuelle Vermessung von Metastasen gibt (Marten, Auer et al. 2006). Vergleiche zwischen ein-, zwei- und dreidimensionalen Messungen von Tumoren verdeutlichen die Präzision dieser Methode (Revel, Lefort et al. 2004; Tran, Brown et al. 2004; Goodman, Gulsun et al. 2006; Jaffe 2006; Marten, Auer et al. 2006; Fabel and Bolte 2008). Insbesondere in Bezug auf den Erfahrungswert in der Befundung von Computertomographien und der Messung von Rundherden testeten Bolte et al. an 46 artifiziellen Lungenrundherden in einer ex-vivo Studie zwei Gruppen von Radiologen mit jeweils 0 Jahren und 5 Jahren an Berufserfahrung. Verglichen mit der manuellen Messung des Durchmessers erreichte die Volumetrie signifikant niedrigere Interobserverunterschiede (Bolte, Jahnke et al. 2007).

Marten et al. untersuchten in einer weiteren Studie aus dem Jahr 2007, ob und wie die Volumetrie in die Responsekategorien nach RECIST eingegliedert werden kann (Marten, Auer et al. 2007). Wenn die Grenzwerte der RECIST-Kriterien bezüglich Tumorprogression (Wachstum von 20%) und -regression (Größenminderung von 30%) in Volumenequivalente umgesetzt werden, entspricht dies Volumenänderungen von 73%, respektive 65% (siehe Tabelle 5).

Tabelle 5: Vergleich zwischen Volumenänderung und Änderung des Durchmessers nach RECIST.

| Klassifikation | Complete response | Partial response | Stable disease | Progressive disease |
|--|--------------------------|-------------------------|--------------------------------|----------------------------|
| Durchmesseränderung nach RECIST | Verschwinden der Läsion | $\leq 30\%$ | Keine signifikanten Änderungen | $\geq 20\%$ |
| Volumenänderung | 100% | $\leq 65\%$ | | $\geq 73\%$ |

Ziel der Studie war es, die kleinstmögliche Volumenänderung zur Klassifizierung von Tumorprogression und partieller Remission zu bestimmen und des Weiteren die minimale Anzahl an Targetläsionen festzulegen um weiterhin eine hohe Reproduzierbarkeit der geänderten Grenzwerte zu haben. Es wurden 50 Patienten mit insgesamt 202 Lungenmetastasen evaluiert und es konnte gezeigt werden, dass der Grenzwert für eine Volumenänderung von 70% auf 35% gesenkt werden konnte. Mindestens drei Läsionen müssen gemessen werden um die Stabilität der Methode zu gewährleisten.

Wiemker et al. berichten, dass die semiautomatische Volumenbestimmung mit Hilfe von Computerprogrammen zunehmend in die klinische Routine integriert wird (Wiemker, Rogalla et al. 2005). Sie liefert zuverlässige und reproduzierbare Ergebnisse und kann somit zur Beurteilung bezüglich des Wachstumsverhaltens von Tumoren herangezogen werden (Sohaib, Turner et al. 2000; Wormanns, Kohl et al. 2004; Marten, Rummeny et al. 2005; Marten, Auer et al. 2006; Graser, Becker et al. 2008). Sogar die Nutzung von low-dose CT mit signifikanter Einsparung der Strahlendosis im Rahmen des Strahlenschutzes bei Screening-Trials, sowie klinischen Follow-Ups, zeigt die Unabhängigkeit der Volumetrie von der eingesetzten Dosis (Hein, Romano et al. ; Hein, Romano et al. 2009).

1.7 Fragestellung

In Studien mit Tumorpatienten, und allgemeiner bei deren Therapiemonitoring, sind die Messkriterien durch die Mess- und Bewertungsrichtlinien der WHO und die RECIST-Kriterien näher bestimmt und gelten als Standard.

Die vorliegende Studie beschäftigt sich mit der Auswertung von CT-Datensätzen im Krankheitsverlauf von Patienten mit Lungentumoren. Das angestrebte Ziel ist die Befundung der Datensätzen bei höherer Aussagekraft zu verkürzen und ggf. fehlendes Ansprechen einer nebenwirkungsreichen und teuren Chemotherapieform zuverlässiger und früher erkennbar zu machen. Entwickelt wurde ein semiautomatisches Tumolvolumetrietool für Lungentumoren CTSV-Prototype (MeVis, Bremen). Die Auswertung der Lungentumore erfolgt nach bisherigen Kriterien der WHO und RECIST sowie mit Hilfe der softwareunterstützten semiautomatischen Volumetrie.

Hierbei sollen folgende Aspekte bearbeitet werden:

- Vergleich der Einteilungen nach RECIST-Schwellenwerten (30/20%) und an RECIST-angepassten Schwellenwerten (65/73%) zur Existenzabschätzung von Schwellenwerten/Referenzbereichen und zur Beurteilung des Unterschieds zwischen den Klassifikationen.
- Bewertung der Übereinstimmung der Messwerte bei gleicher Technik und Methode aber wechselnden Beobachtern (Reproduzierbarkeit).
- Bewertung der Übereinstimmung bei unterschiedlicher Technik und Methode aber gleichen Beobachtern.
- Interaktionszeitanalyse mit Hinblick auf die Praktikabilität.

Zusammenfassend soll diese Studie Informationen über die Zuverlässigkeit und Genauigkeit des Volumetrietools unter klinischen Arbeitsbedingungen, eine Einschätzung der Änderung des klinischen Behandlungsverlaufs bei Verwendung der Volumetrie im Vergleich zu den RECIST-Kriterien ermöglichen und eine Abschätzung der Praktikabilität und Relevanz der Tumolvolumetrie im Rahmen des klinischen Alltags liefern.

2 Material und Methoden

2.1 Patientenkollektiv

Grundlage dieser Studie sind CT-Untersuchungen des Thorax von 23 Patienten (16 Männer, 7 Frauen) im Alter von 45 – 87 Jahren (mittleres Alter 65.7 Jahre), die im Zeitraum November 2006 bis Januar 2009 an der Uniklinik Marburg wegen unterschiedlicher Primärtumoren chemotherapeutisch behandelt wurden. Bei den Patienten wurde zur Verlaufsbeobachtung der Lungentumore bzw. -metastasen unter Therapie routinemäßig Computertomographien durchgeführt. Zwölf Patienten haben mindestens ein Follow-Up, acht Patienten mindestens zwei Follow-Ups, vier Patienten mindestens drei Follow-Ups und zwei Patienten haben vier Follow-Ups. Elf Patienten weisen keine Follow-Up-Untersuchung auf. Unter den Primärtumoren befanden sich zehn Nichtkleinzellige-Lungenkarzinome (NSCLC), drei Adenokarzinome, drei Kolonkarzinome, zwei Rektumkarzinome, ein Kleinzelliges-Lungenkarzinom (SCLC), ein Hypopharynxkarzinom, ein Harnblasenkarzinom, ein Plattenepithelkarzinom und ein Larynxkarzinom (siehe Abbildung 7).

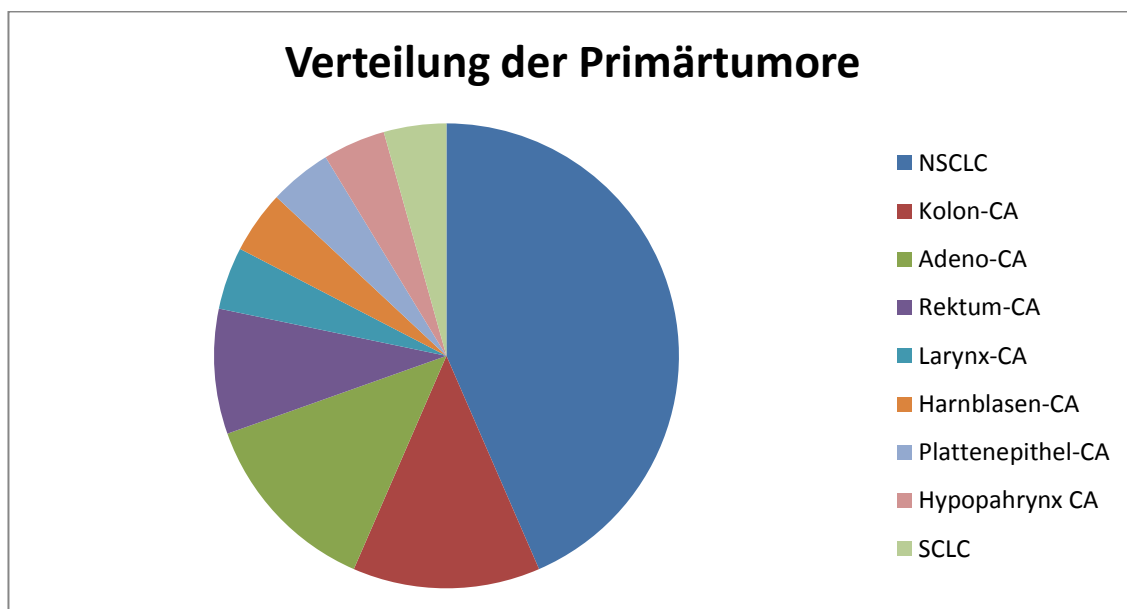


Abbildung 7: Verteilung der Primärtumore der Studienpatienten.

2.2 Computertomographie

Alle Computertomographien wurden im Rahmen der Verlaufskontrolle unter Chemotherapie durchgeführt. Die Bilder wurden mit einem CT (Siemens, Somatom Definition) und mit standardisiertem Untersuchungsprotokoll akquiriert: Schichtdicke 2 mm, 500 ms Rotationszeit, 120 mAs, 120 kV. Die Thoraxaufnahme umfasste einen Scanbereich vom Jugulum bis zur Leberkuppel und wurde in End-Inspiration des Patienten in Rückenlage angefertigt. Nach Ausschluss von Kontraindikationen (z.B. Allergie gegen jodhaltiges Kontrastmittel, Schilddrüsenüberfunktion, Niereninsuffizienz) startete der Scan 60 Sekunden nach intravenöser Injektion von 120 ml nieder-osmolarem jodhaltigem Kontrastmittel (Imeron 300) mit einer Injektionsgeschwindigkeit von 2.5 ml/s.

Zur Bildanalyse wurden die Computertomographien aus dem Picture Archiving and Communication System (PACS) in das Programm CTSV-Prototype (MeVis, Bremen) geladen. Zur Auswertung jedes Bildes wurde der Grauwertbereich auf das Lungenfenster eingestellt. Das Zentrum des Fensters lag bei -500 HU, die Fensterbreite bei 1600 HU.

2.3 Methoden zur Bestimmung der Tumorlast

2.3.1 Manuelle Messung

Alle 215 Läsionen wurden jeweils von 2 Untersuchern (Reader1, Reader2) mit unterschiedlicher Befundungserfahrung sowohl manuell, als auch mit Hilfe der semiautomatischen Volumetrie vermessen. Bei jedem Patienten wurden die Läsionen zu Beginn der Studie (Zeitpunkt t0) sowie zum Zeitpunkt weiterer Verlaufskontrollen (t1, t2, t3 und t4) vermessen, sodass es sich insgesamt um 430 Messungen handelte.

Die eindimensionale manuelle Messung wurde mit Hilfe eines digitalen Lineals, integriert in das Volumetrie-Programm CTSV-Prototype (MeVis, Bremen), entsprechend der RECIST-Kriterien ausgeführt.

Zuerst wurde eine Läsion innerhalb der Lunge ausgewählt und die transversale Schnittebene festgelegt, in der der Tumor die größte Ausdehnung zeigte. Als max. Durchmesser wurde die größte Ausdehnung in der transversalen Schnittebene der Metastase definiert und in Millimetern (mm) angegeben. Dies wurde für jede messbare Läsion sowohl im Baseline-Scan als auch in den Follow-Up-Untersuchungen durchgeführt. Sowohl intrapulmonale, als auch mediastinal- und pleuraständige Läsionen wurden berücksichtigt.

Anhand der zwei größten Läsionen innerhalb der Lunge konnte für jede Untersuchung der entsprechende RECIST-Wert berechnet werden und die Follow-Up-Untersuchungen in die jeweilige Responsekategorie eingeteilt werden.

2.3.2 Semiautomatische Volumetrie

Die Berechnung des Volumens erfolgte semiautomatisch durch das Programm CTSV-Prototype Version 1.0.0 (MeVis, Bremen), indem, wie oben beschrieben, manuell der Durchmesser einer Läsion bestimmt wurde. Eine weitere Möglichkeit zur Messung des Volumens besteht in dem einfachen Anklicken der Läsion mit der linken Maustaste. Der Segmentierungsalgorithmus wurde ausführlich von Bornemann et al. beschrieben (Bornemann, Kuhnigk et al. 2005). Es ist ein Hybrid-Algorithmus, der in mehreren Teilschritten erfolgt:

Nachdem man die zu untersuchende Läsion identifiziert hat, wird diese mit einem Mausklick markiert und somit zum „volume of interest“ (VOI) gewählt. Die

Segmentierung und Volumetrie der Läsion folgen automatisch. Von dem gewählten Punkt aus erfolgt initial eine grobe Segmentierung des Herdes durch einen sog. „Region-growing-Algorithmus“. Anhand festgesetzter Dichteschwellenwerte wächst ein Bereich um den markierten Punkt an, bis am Rande der Läsion die Dichte wieder abfällt. Die Segmentierung unterteilt den Herd in einzelne dreidimensionale Würfel, die als Voxel bezeichnet werden. Da in der so entstehenden groben Darstellung des Herdes auch noch Brustwand enthalten sein kann (wie beispielsweise bei pleuraassoziierten Prozessen) oder Anteile von Blutgefäßen, werden diese Strukturen anschließend durch spezielle Berechnungsverfahren ausgeschlossen.

Durch die Anwendung eines „Convex-hull-Algorithmus“ kann eine Separation von Brustwand und Läsion erreicht werden. Die konvexe Hülle eines definierten Voxel-Sets ist das kleinste konvexe Objekt, das das komplette Voxel-Set beinhaltet. Das Resultat eines solchen Algorithmus ist eine Maske die die ungefähre Parenchymkontur beschreibt, wie sie ohne eine dort ansässige Läsion erscheinen würde. Diese Kontur wird anschließend benutzt um Pleura und Herd zu separieren.

Bei der Trennung von Läsion und Blutgefäßen wird ebenfalls mit einem Schwellenwert-basiertem „Region-growing-Algorithmus“ begonnen. Anhand von Erosions- und Dilatations-Operationen wird eine weitere Maske mit Separation der Blutgefäße erstellt. Die Kombination beider Masken ergibt das endgültige Bild der Läsion.

Durch das Vorhandensein eines partiellen Volumeneffekts ist es nicht adäquat, die Volumetrie durch Zählen der einzelnen Voxel durchzuführen. Auf der Basis der Segmentation werden drei verschiedene Areale automatisch definiert. Mittlere Schwächungswerte werden für den Kern der Läsion und ein parenchymales Areal berechnet und erlauben so die Berechnung einer gewichteten Beteiligung der Voxel innerhalb der Region mit partiellem Volumeneffekt zum Gesamtvolumen der Läsion. Insbesondere bei hohen Dichteunterschieden erlaubt diese Vorgehensweise der Segmentierung und Volumetrie eine exakte Bestimmung der Objektgrenzen unter Verwendung eines vergleichsweise geringen Zeitaufwandes.

Sollte der Herd jedoch nicht zur vollen Zufriedenheit des Untersuchers segmentiert worden sein, so sind weitere Möglichkeiten der manuellen Nachkorrektur vorhanden. In einem Abschluss-Report wird neben dem Volumen des Herdes in ml, der effektive Durchmesser in mm (als Durchmesser einer Kugel mit dem Volumen der Läsion), automatisch berechnete RECIST- und WHO-Werte, manuelle RECIST-Werte, die

Ausdehnung in Z-Richtung, der maximale 3D-Durchmesser, sowie durch Mittelung der Dichte der einzelnen Voxel die durchschnittliche Dichte des Herdes angegeben. Nach Starten der Volumetrie wurde außerdem die Zeit berechnet, die vom ersten Anklicken der Läsion bis zur fertigen Berechnung der Parameter benötigt wurde.

2.4 Statistische Auswertung

Für alle oben genannten manuell und automatisch ermittelten Parameter wurden folgende statistische Werte berechnet: Minimum, Maximum, Mittelwert, mittlere prozentuale Abweichung der Differenzen von den Mittelwerten und die Standardabweichung. Die p-Werte wurden mit Hilfe des Student t-Tests ermittelt.

Die statistische Auswertung erfolgte mittels Microsoft Excel 2007 und das Erstellen von Diagrammen durch die statistische Software MedCalc (MedCalc Software bvba, Version 11.3.3., Belgien).

Für die Bewertung einer Therapieänderung wurde das RECIST-Kriterium auf Durchmesser und Volumen übertragen. Eine relevante Zunahme der Metastase war dann gegeben, wenn der jeweilige Parameter (max. Durchmesser oder Volumen) um 20% bzw. 73% oder mehr zunahm. Eine relevante Größenabnahme lag vor, wenn sich der jeweilige Wert (max. Durchmesser oder Volumen) um 30% bzw. 65% oder mehr im zeitlichen Verlauf verringerte. Die Anzahl der divergierenden Therapieentscheidung durch beide Verfahren wurde als absoluter und als prozentualer Wert angegeben. Mit Hilfe der Bland-Altman-Analyse (Bland and Altman 1986) wurde die Übereinstimmung von zwei Methoden bzw. zwei Untersuchern getestet. Hierbei werden die Differenzen der Messwertpaare zweier miteinander zu vergleichender Messverfahren gegen den Mittelwert der beiden Messverfahren (Bias) aufgetragen, wobei idealerweise Bias = 0 ist. Diese Darstellung ist sinnvoll, wenn der wahre Messwert nicht bekannt ist. Zusätzlich zur medianen Differenz als Maß der Übereinstimmung zwischen den diagnostischen Verfahren werden noch die sogenannten „limits of agreement“ abgetragen. Diese umschließen üblicherweise 95% der beobachteten Abweichungen (± 1.96 fache Standardabweichung) zwischen den beiden Messmethoden.

Das Bland-Altman Diagramm gestattet eine optimale Einschätzung systematischer Fehler. So lässt sich schon visuell einschätzen, ob eine Messmethode prinzipiell höhere oder tiefere Werte misst als die andere, oder ob die Abweichung der Methode oder die

Streuung der Abweichungen von der Höhe der Werte abhängig ist. Liegen nicht 95% der Punkte innerhalb des Toleranzbereichs der Paardifferenzen, sind die Methoden nicht vergleichbar.

Grundsätzlich wurden alle Mittelwerte im Zusammenhang mit der Standardabweichung angegeben, das Konfidenzintervall auf 95% und das Signifikanzniveau auf 5% festgelegt.

3 Ergebnisse

3.1 Durchmesser und Volumen der Lungenläsionen

Der Mittelwert des manuell gemessenen maximalen Durchmessers aller Messungen ($n = 430$) betrug 16.88 mm (Tabelle 6).

Tabelle 6: Ergebnisse der manuellen Messung des max. Durchmessers der Lungenläsionen.

| Anzahl der Metastasen $n = 430$ | Manuelle Messung |
|---------------------------------|------------------|
| Mittelwert [mm] | 16.88 |
| Standardabweichung [mm] | 23.17 |
| Minimum [mm] | 2.09 |
| Maximum [mm] | 141.84 |

Der Mittelwert des semiautomatisch gemessenen Volumens aller Messungen ($n = 430$) betrug 19.52 mm³ (Tabelle 7).

Tabelle 7: Ergebnisse der semiautomatischen Volumetrie.

| Anzahl der Metastasen $n = 430$ | Semiautomatische Volumetrie |
|---------------------------------------|-----------------------------|
| Mittelwert [mm ³] | 19.52 |
| Standardabweichung [mm ³] | 90.63 |
| Minimum [mm ³] | 0.02 |
| Maximum [mm ³] | 818.71 |

3.2 Vergleich der Einteilung nach RECIST-Schwellenwerten (30/20%) und an RECIST angepassten Schwellenwerten (65/73%) zur Existenzabschätzung von Schwellenwerten/Referenzbereichen und zur Beurteilung des Unterschieds zwischen den Klassifikationen

Von den 23 Patienten verfügen 12 Patienten über mindestens eine Follow-Up-Untersuchung. Insgesamt wurden 26 Follow-Up-Untersuchungen durchgeführt. Zu jedem Follow-Up wurde der entsprechende RECIST-Wert berechnet, indem die manuell gemessene Summe der Durchmesser der zwei größten Läsionen mit den jeweiligen Werten der vorherigen Messung verglichen wurde. Ebenso wurde für die zwei größten Läsionen semiautomatisch das Volumen ermittelt und ein entsprechender RECIST-Wert berechnet (siehe Kapitel 2.3). Das Ergebnis einer Progression oder Regression ist in Prozent ausgedrückt und für jedes Follow-Up ($t_1 - t_4$) wurden die Ergebnisse in Tabelle 8 dargestellt.

Tabelle 8: Prozentuale Änderungen der Tumorgröße der einzelnen Follow-Up-Untersuchungen (t1 – t4) manuell gemessen nach RECIST-Kriterien und durch semiautomatische Volumetrie.

| Patient | | t1 | t2 | t3 | t4 |
|----------------|------------|-----------|-----------|-----------|-----------|
| 1 | RECIST | +1.96 | -2.12 | -3.67 | -2.23 |
| | Volumetrie | -54.86 | +65.19 | -22.16 | +26.89 |
| 2 | RECIST | -13.48 | +4.07 | -15.29 | |
| | Volumetrie | -19.36 | -1.76 | +1.05 | |
| 3 | RECIST | +459.17 | | | |
| | Volumetrie | +4826.07 | | | |
| 4 | RECIST | +14.90 | +18.41 | +16.99 | -18.07 |
| | Volumetrie | +149.21 | +27.64 | +65.63 | -26.73 |
| 5 | RECIST | -30.93 | | | |
| | Volumetrie | -30.64 | | | |
| 6 | RECIST | +5.44 | +48.19 | +31.53 | |
| | Volumetrie | +10.95 | +203.76 | +39.03 | |
| 7 | RECIST | +0.75 | -24.49 | | |
| | Volumetrie | +118.72 | -64.31 | | |
| 8 | RECIST | -26.64 | -7.89 | | |
| | Volumetrie | -33.70 | -35.09 | | |
| 9 | RECIST | +21.57 | | | |
| | Volumetrie | +17.87 | | | |
| 10 | RECIST | +29.53 | +1318.42 | | |
| | Volumetrie | +5.47 | +1072.80 | | |
| 11 | RECIST | -16.38 | +15.27 | | |
| | Volumetrie | +565.76 | -84.46 | | |
| 12 | RECIST | -36.05 | | | |
| | Volumetrie | -31.09 | | | |

Die RECIST-Werte für jede Follow-Up-Untersuchung (n = 26) wurden in die entsprechenden Responsekategorien eingeteilt (Abbildung 8). Die semiautomatisch ermittelten Volumetriewerte wurden ebenfalls unter Verwendung der an das Volumen

angepassten Schwellenwerte von 65/73% nach RECIST in Responsekategorien gegliedert (Abbildung 9).

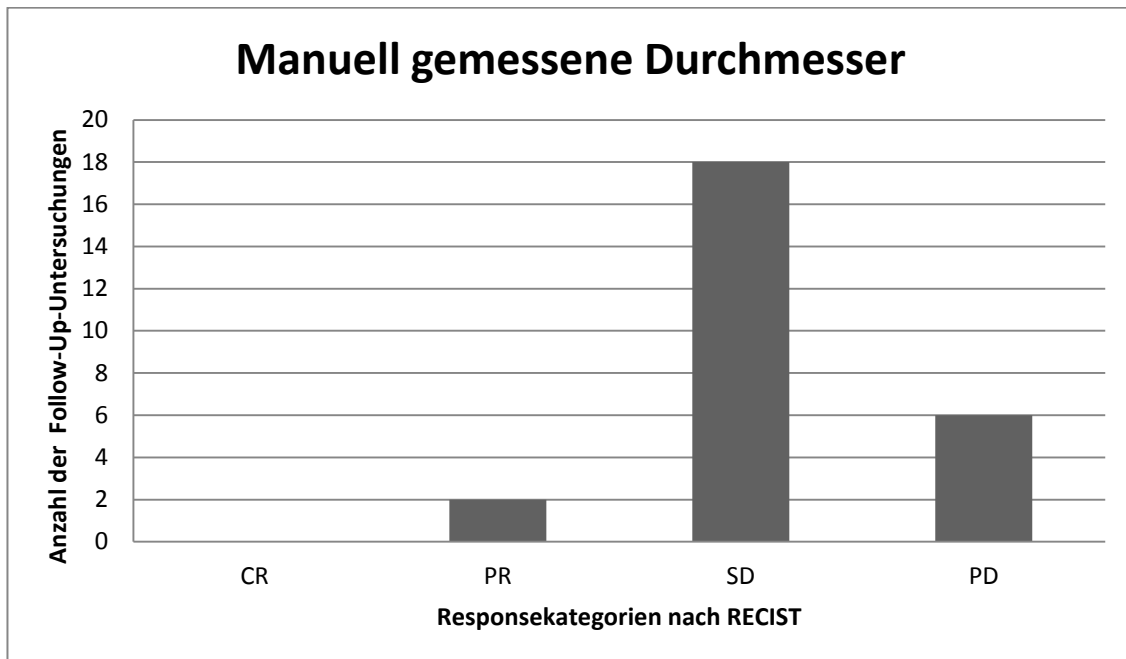


Abbildung 8: Einteilung der Follow-Up-Untersuchungen (n = 26) mit manuell gemessenem Durchmesser in Responsekategorien nach RECIST.

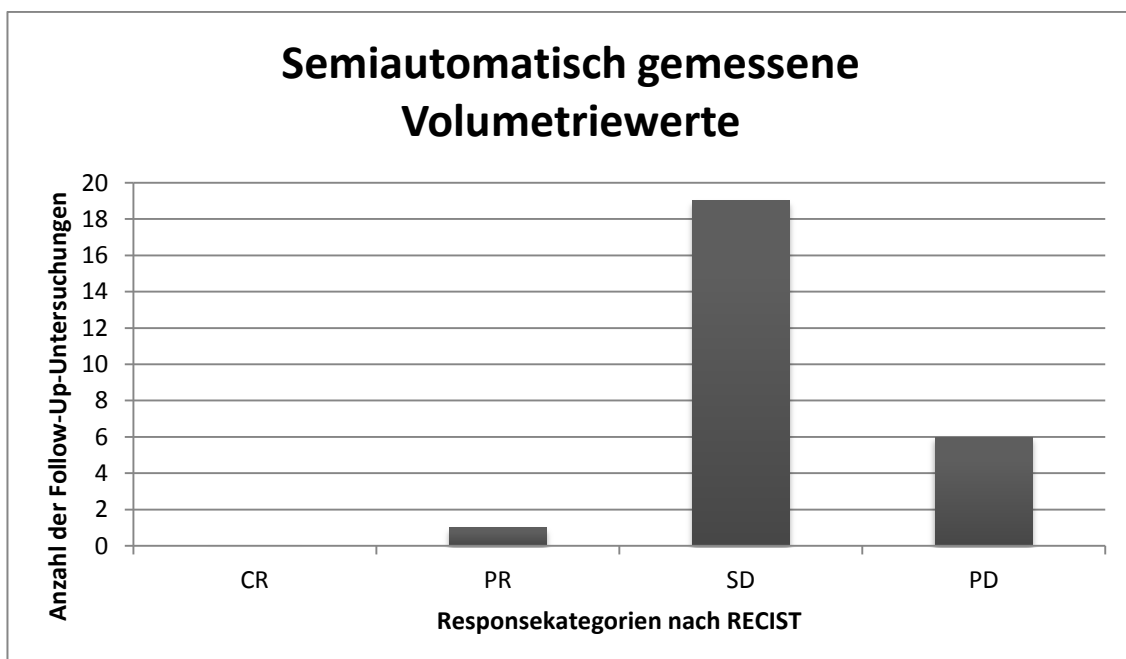


Abbildung 9: Einteilung der Follow-Up-Untersuchungen mit semiautomatisch gemessenen Volumetriewerten in Responsekategorien nach RECIST.

Beim Vergleich der Verteilung der Volumetriewerte zwischen den per RECIST-Kriterien festgelegten Responsekategorien und den RECIST-Werten der manuell gemessenen Durchmesser zeigten sich folgende Unterschiede, die in den Tabelle 9 -Tabelle 11 dargestellt sind. Für die jeweilige Responsekategorie der manuell gemessenen Durchmesser wurden die dazugehörigen semiautomatischen Volumetriewerte aufgeschlüsselt. In der Kategorie „stable disease“ wurden von 18 Follow-Up-Untersuchungen mit manuell gemessenen Durchmessern 14 Volumetriewerte ebenfalls in diese Kategorie eingeordnet. Daraus ergab sich eine Abweichung von 33.3%. Ein Volumetriewert fiel in die Kategorie „partial response“ und drei weitere mussten volumetrisch gemessen als „progressive disease“ gewertet werden, was in einer Abweichung von jeweils 50% resultierte. Sechs der manuell gemessenen Follow-Up-Untersuchungen wurden nach RECIST als „progressive disease“ eingestuft, wovon in der Volumetrie drei der Untersuchungen gleichfalls progressiv waren, die anderen drei jedoch als „stable disease“ gewertet wurden. Dies zeigte eine Abweichung von 50%. Als „partial response“ wurden zwei manuell gemessene Follow-Ups festgelegt, die sich in der Volumetrie durchweg als „stable disease“ präsentierten, womit sich hier Abweichungen der volumetrischen von den manuellen Messungen von 100% darstellten.

Werden die jeweils manuell gemessenen Durchmesser und die semiautomatischen Volumetriewerte, beide klassifiziert durch die entsprechenden RECIST-Kriterien, als Basis herangezogen, so würden sich die daraus resultierenden Therapiemaßnahmen zwischen beiden Arten der Auswertung in 32% der Fälle unterscheiden.

Tabelle 9: Follow-Up-Untersuchungen mit manuell gemessenem Durchmesser (n = 18) in der Responsekategorie SD nach RECIST aufgeschlüsselt in der semiautomatischen Volumetrie.

| | Semiautomatische Volumetrie |
|----|-----------------------------|
| CR | 0 |
| PR | 1 |
| SD | 14 |
| PD | 3 |

Tabelle 10: Follow-Up-Untersuchungen mit manuell gemessenem Durchmesser (n = 6) in der Responsekategorie PD nach RECIST aufgeschlüsselt in der semiautomatischen Volumetrie.

| | Semiautomatische Volumetrie |
|----|------------------------------------|
| CR | 0 |
| PR | 0 |
| SD | 3 |
| PD | 3 |

Tabelle 11: Follow-Up-Untersuchungen mit manuell gemessenem Durchmesser (n = 2) in der Responsekategorie PR nach RECIST aufgeschlüsselt in der semiautomatischen Volumetrie.

| | Semiautomatische Volumetrie |
|----|------------------------------------|
| CR | 0 |
| PR | 0 |
| SD | 2 |
| PD | 0 |

3.3 Bewertung der Übereinstimmung der Messwerte bei gleicher Technik und Methode aber wechselnden Beobachtern (Reproduzierbarkeit)

Alle Messungen wurden jeweils von zwei unabhängigen Untersuchern (Reader1 und Reader2) durchgeführt um die Reproduzierbarkeit der Ergebnisse bewerten zu können.

Nach Auswertung der manuellen Messungen der Follow-Up-Untersuchungen ergaben sich für die beiden Untersucher geringe Unterschiede in der Einteilung der Responsekategorien nach RECIST. In den Kategorien PR und SD lagen die Differenzen bei jeweils einer Messung (PR: 50%, SD: 6%), in der Kategorie PD bei zwei Messungen (50%). Die Reproduzierbarkeit der Messungen der Tumorresponse durch zwei Untersucher lag somit bei 84.6%. Dementsprechend würde das in 4 von 26 Fällen (15.4%) eine divergente Therapieänderung bedeuten (Abbildung 10).

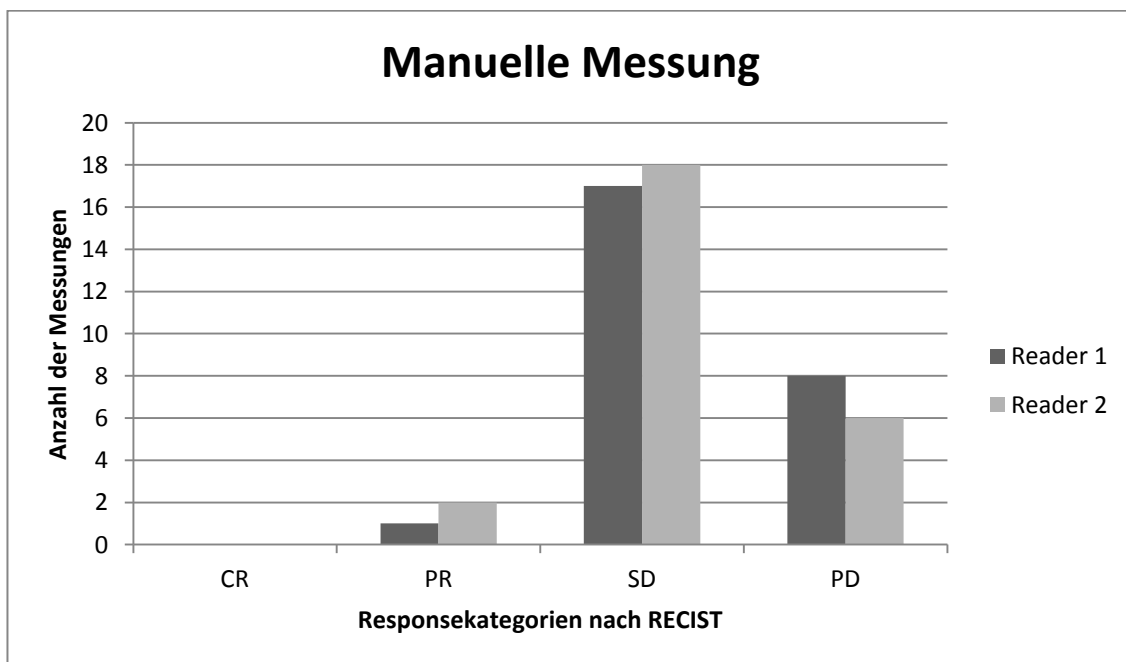


Abbildung 10: Vergleich der Einteilung der manuellen Follow-Up-Messungen durch zwei unabhängige Untersucher (Reader1, Reader2) in Responsekategorien nach RECIST.

Die computerunterstützte semiautomatische Volumetrie der Follow-Up-Untersuchungen, ebenfalls durch die zwei Untersucher initiiert, ergab jeweils eine unterschiedliche Einteilung in den Responsekategorien SD und PD. Dies zeigte eine

Reproduzierbarkeit von 92.3% und eine geringere Abweichung in nur 2 von 26 Fällen bezüglich der Tumorresponse und konsekutiver Therapieänderung (Abbildung 11).

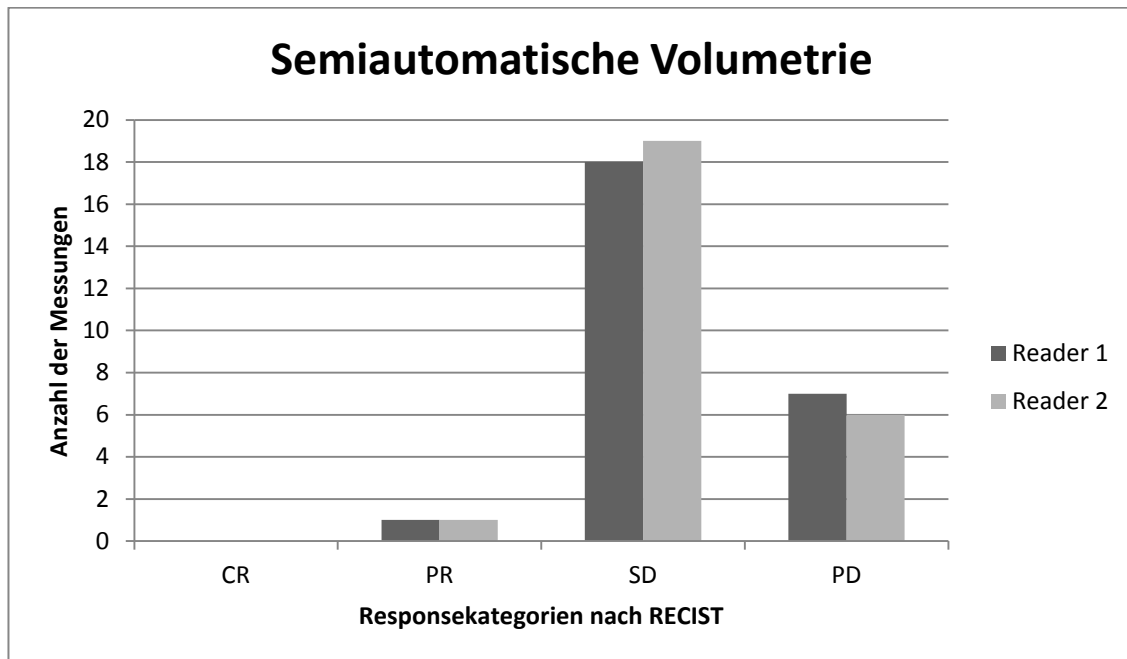


Abbildung 11: Vergleich der Einteilung der semiautomatisch gemessenen Follow-Up-Volumina durch zwei unabhängige Untersucher (Reader1, Reader2) in Responsekategorien nach RECIST.

3.3.1 Vergleich der manuellen Messung des größten Durchmessers von zwei Untersuchern

Zur Überprüfung der Reproduzierbarkeit wurden zunächst die manuellen Messungen des größten Durchmessers der Läsionen nach RECIST durch zwei unabhängige Untersucher verglichen. Der Mittelwert der Differenzen der Messungen betrug 1.84 mm, die Standardabweichung 2.25 mm. Bei einem p-Wert von 0.058 waren die Unterschiede nicht signifikant. Zu oben genanntem Mittelwert wurde jeweils die prozentuale Abweichung der Einzelmessungen der beiden Untersucher berechnet. Ein t-Test dieser prozentualen Abweichung zum Mittelwert wies ein Signifikanzniveau von $p < 0.001$ auf.

Werden die manuellen Messungen des größten Durchmessers einer Läsion nach RECIST von Reader1 und Reader2 ($n = 430$) graphisch durch ein Bland-Altman-Diagramm (Bland and Altman 1986) dargestellt, ergibt sich Abbildung 12. Das

Diagramm zeigt, dass die Messungen der beiden Untersucher sich bei Werten unter 25 mm zum überwiegenden Teil innerhalb der „limits of agreement“ und nahe am Mittelwert gruppieren, bei Werten größer als 50 mm jedoch stärker auseinanderdivergieren.

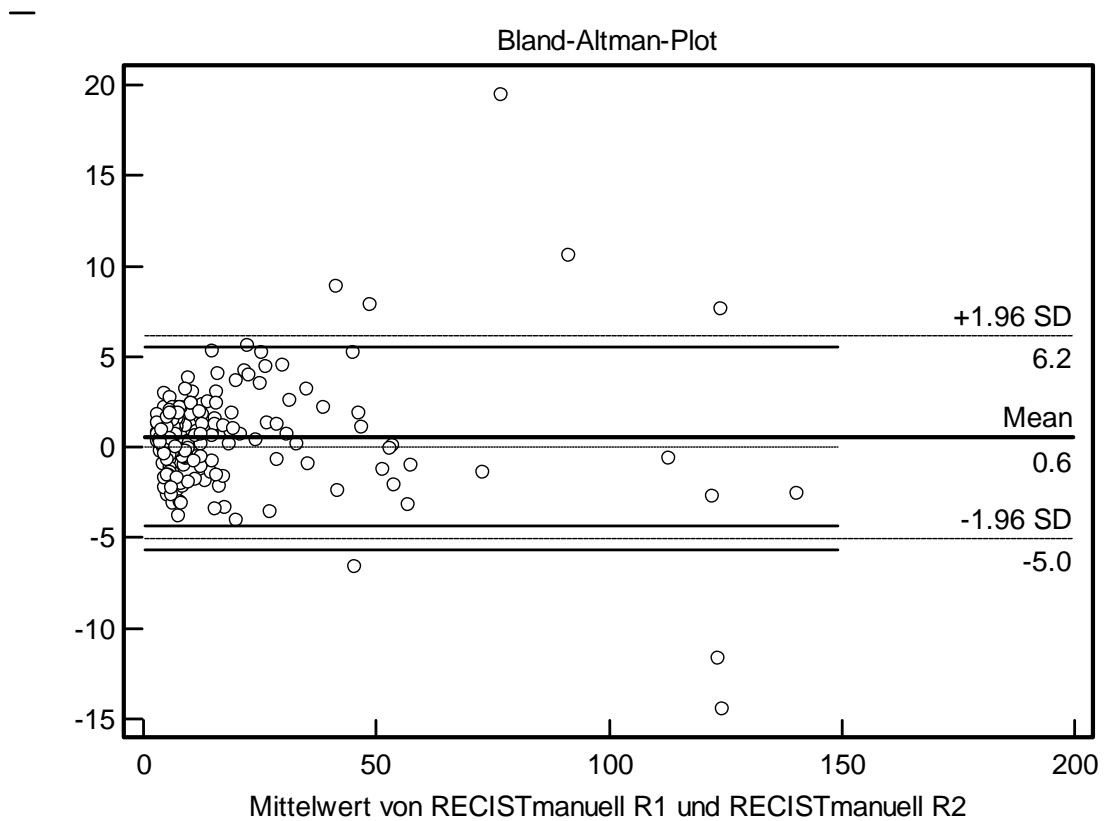


Abbildung 12: Bland-Altman-Diagramm zur manuellen Messung nach RECIST durch zwei Untersucher (R1, R2). Differenz zwischen manuell ermitteltem Durchmesser von R1 und R2 in Relation zur Summe des Mittelwertes der manuell ermittelten Durchmesser von R1 und R2 für jede Messung.

Weiterhin war zu untersuchen, ob eine Änderung in der Übereinstimmung der beiden Messmethoden (nach Bland-Altman) erkennbar ist, wenn die Messwerte bei kleinen und größeren Läsionen getrennt untersucht wurden.

Um die Unterschiede in Bezug auf die Größe der Läsionen verdeutlichen zu können, wurden zunächst alle Läsionen, die kleiner als 10 mm waren, aus den Berechnungen herausgenommen. Bei somit 176 Messungen ergab sich ein Mittelwert der Differenzen der Messungen von zwei Untersuchern von 2.75 mm mit einer Standardabweichung

von 3.22 mm bei einem Signifikanzniveau von $p = 0.08$. Eine Berechnung der mittleren prozentualen Abweichung zum Mittelwert ergab ein Signifikanzniveau von $p < 0.05$. Die graphische Darstellung erfolgte durch ein Bland-Altman-Diagramm, in welchem erkennbar ist, dass der Ausschluss von Läsionen kleiner als 10 mm, die Ergebnisse von Abbildung 11 nochmal verdeutlicht. Je größer die Läsionen sind, desto höher sind die Differenzen der Messwerte von zwei Untersuchern (Abbildung 13).

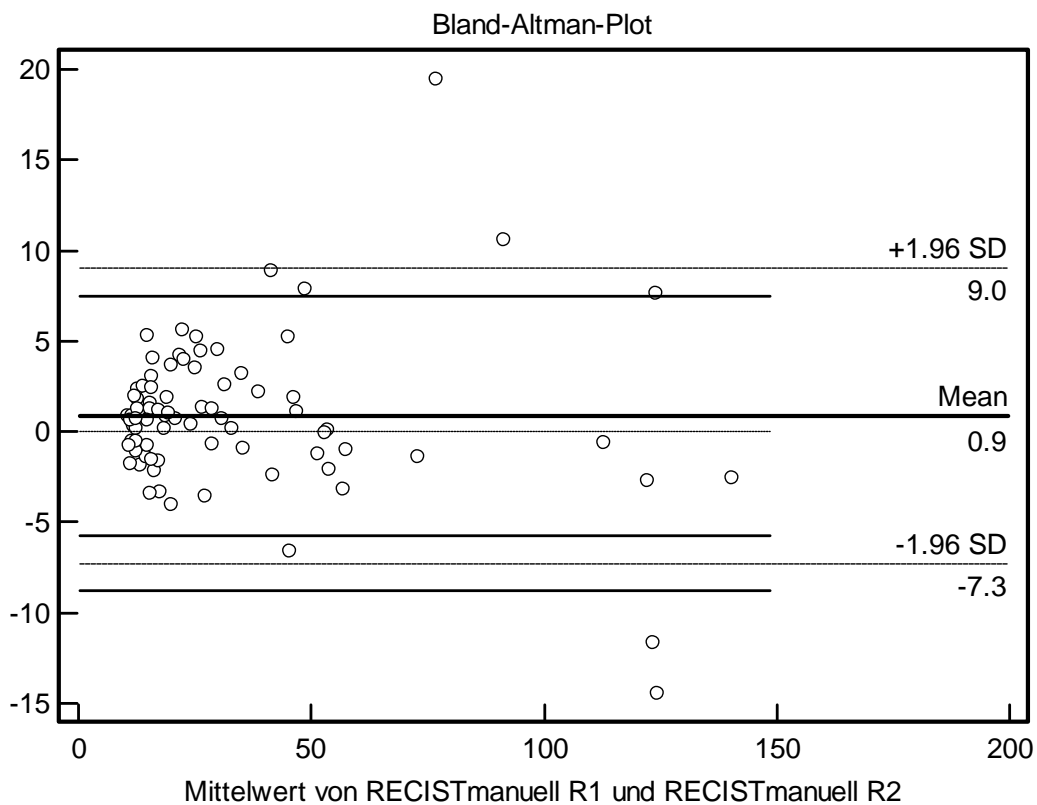


Abbildung 13: Bland-Altman-Diagramm zur manuellen Messung nach RECIST durch zwei Untersucher (R1, R2) unter Ausschluss aller Läsionen kleiner als 10 mm. Differenz zwischen manuell ermitteltem Durchmesser von R1 und R2 in Relation zur Summe des Mittelwertes der manuell ermittelten Durchmesser von R1 und R2 für jede Messung.

Ebenso wurden die Auswirkungen für alle Messungen unter Ausschluss von Läsionen, die größer als 10 mm waren, untersucht ($n = 254$). Der Mittelwert der Differenz der Messwerte war 1.17 mm bei einer Standardabweichung von 0.83 mm. Diese Unterschiede zwischen den Untersuchern waren bei einem p-Wert von 0.07 nicht signifikant, konnten jedoch durch Berechnung der prozentualen Abweichung vom oben genannten Mittelwert als signifikant dargestellt werden ($p < 0.05$).

In der graphischen Darstellung durch ein Bland-Altman-Diagramm zeigen sich bei der Untersuchung der kleinen Läsionen unter 10 mm starke Abweichungen vom Mittelwert, überwiegend im positiven Bereich von 3 mm - 8 mm (Abbildung 14).

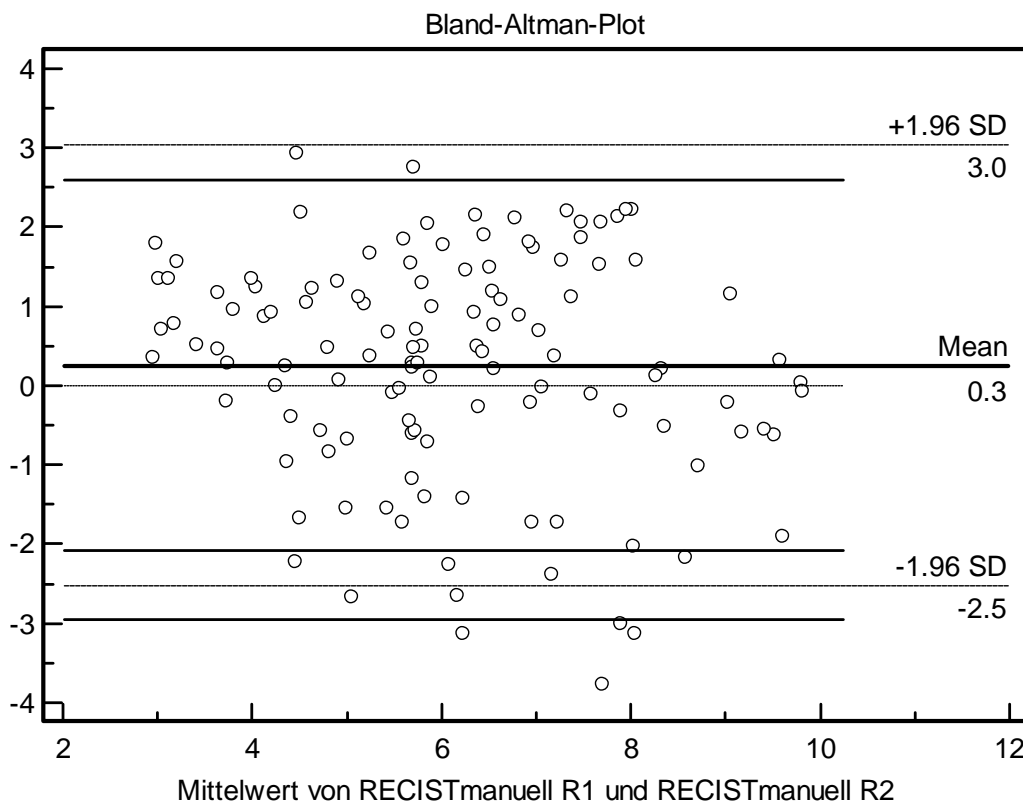


Abbildung 14: Bland-Altman-Diagramm zur manuellen Messung nach RECIST durch zwei Untersucher (R1, R2) unter Ausschluss aller Läsionen größer als 10 mm. Differenz zwischen manuell ermitteltem Durchmesser von R1 und R2 in Relation zur Summe des Mittelwertes der manuell ermittelten Durchmesser von R1 und R2 für jede Messung.

3.3.2 Vergleich der semiautomatischen Messung des größten Durchmessers von zwei Untersuchern

Der Vergleich der semiautomatisch ermittelten größten Durchmesser der Läsionen durch zwei Untersucher ($n = 430$) resultierte in folgenden Ergebnissen: Der Mittelwert der Differenzen der Messwerte der Untersucher lag bei 0.89 mm mit einer Standardabweichung von 1.93 mm. Die Unterschiede waren sowohl hier nicht statistisch signifikant ($p = 0.14$), als auch bei der Berechnung der mittleren prozentualen Abweichungen der Differenzen von den Mittelwerten ($p = 0.49$).

Im Bland-Altman-Diagramm zeigt sich eine hohe Konzentration der Messpunkte um den Mittelwert, besonders im Bereich der kleineren Läsionen. Stärkere Abweichungen vom Mittelwert befinden sich im Bereich größerer Läsionen (Abbildung 15).

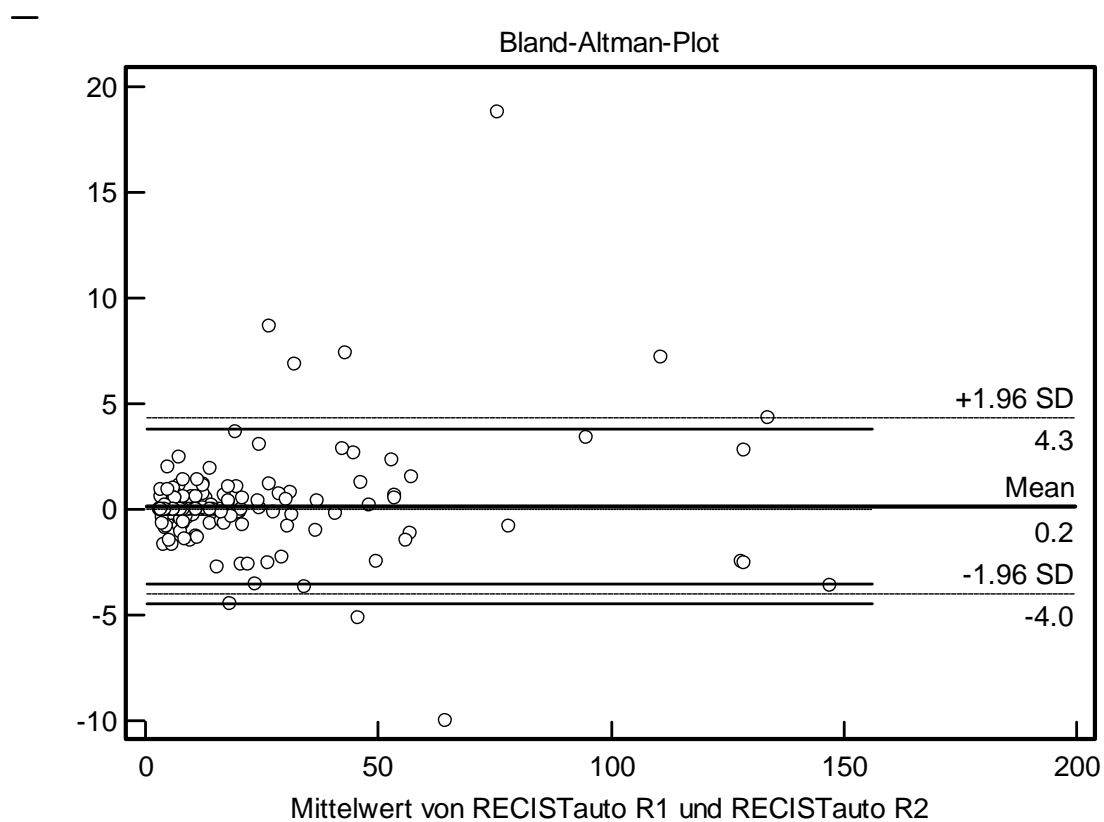


Abbildung 15: Bland-Altman-Diagramm zur semiautomatischen Messung des größten Durchmessers der Läsion durch zwei Untersucher (R1, R2). Differenz zwischen semiautomatisch ermitteltem Durchmesser von R1 und R2 in Relation zur Summe des Mittelwertes der semiautomatisch ermittelten Durchmesser von R1 und R2 für jede Messung.

Unter Ausschluss aller Läsionen kleiner als 10 mm war der Mittelwert der Differenzen der Messergebnisse ($n = 176$) 1.64 mm, die Standardabweichung 2.66 mm und die statistische Signifikanz änderte sich nicht im Vergleich zu den obigen Werten ($p = 0.15$).

Das Diagramm zeigt die Abweichungen der Werte aufgeschlüsselt für größere Läsionen. Im Bereich der Größe von 10 mm bis 50 mm lässt sich hier eine hohe Übereinstimmung der Werte der beiden Untersucher mit nur geringen Abweichungen vom Mittelwert erkennen (Abbildung 16).

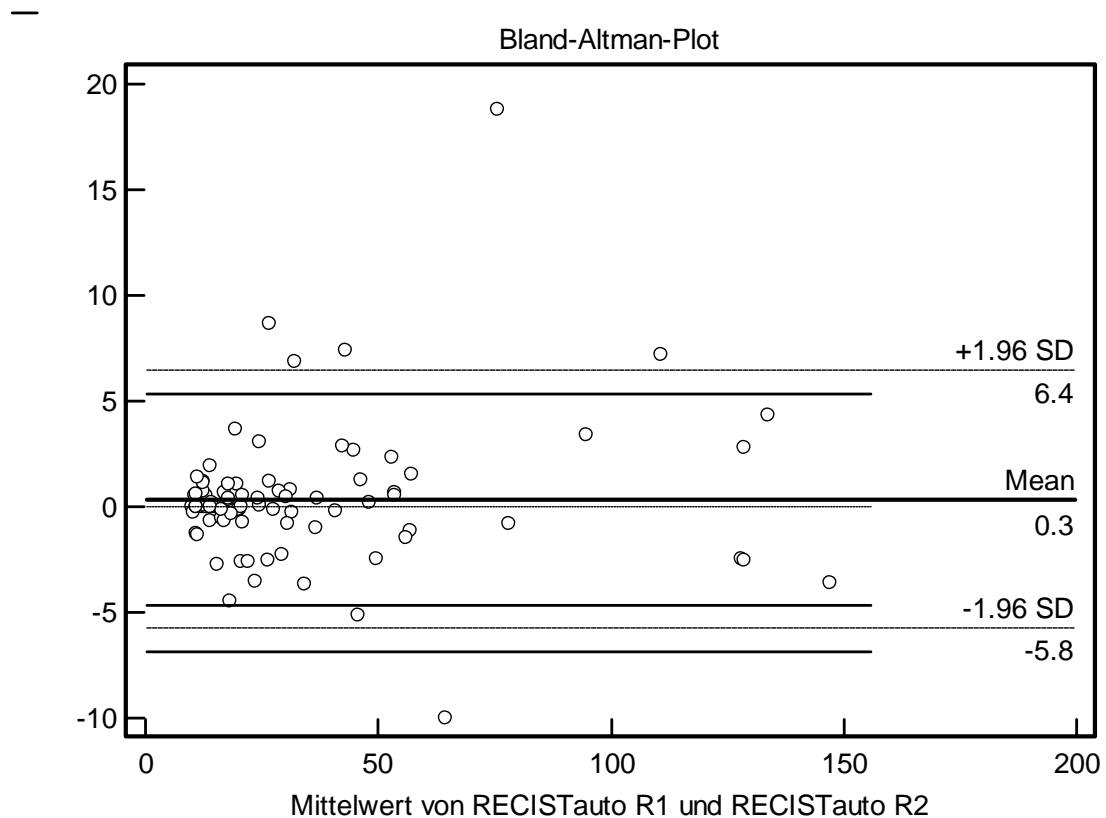


Abbildung 16: Bland-Altman-Diagramm zur semiautomatischen Messung des größten Durchmessers der Läsion durch zwei Untersucher (R1, R2) unter Ausschluss aller Läsionen kleiner als 10 mm. Differenz zwischen semiautomatisch ermitteltem Durchmesser von R1 und R2 in Relation zur Summe des Mittelwertes der semiautomatisch ermittelten Durchmesser von R1 und R2 für jede Messung.

Eine Berechnung aller Läsionen kleiner als 10 mm ($n = 254$) ergab einen Mittelwert der Differenzen von 0.28 mm mit einer Standardabweichung von 0.49 mm. Die Unterschiede zwischen den Untersuchern sind als nicht signifikant zu betrachten ($p = 0.82$).

Die graphische Darstellung der Messwerte durch das Bland-Altman-Diagramm lässt erkennen, dass sich gerade im Bereich der kleinen Läsionen viele Werte der beiden Untersucher gleichen, da sie genau auf der horizontalen Gerade des Mittelwertes liegen. Ansonsten liegt eine gleichmäßige Streuung in beide Richtungen vor (Abbildung 17).

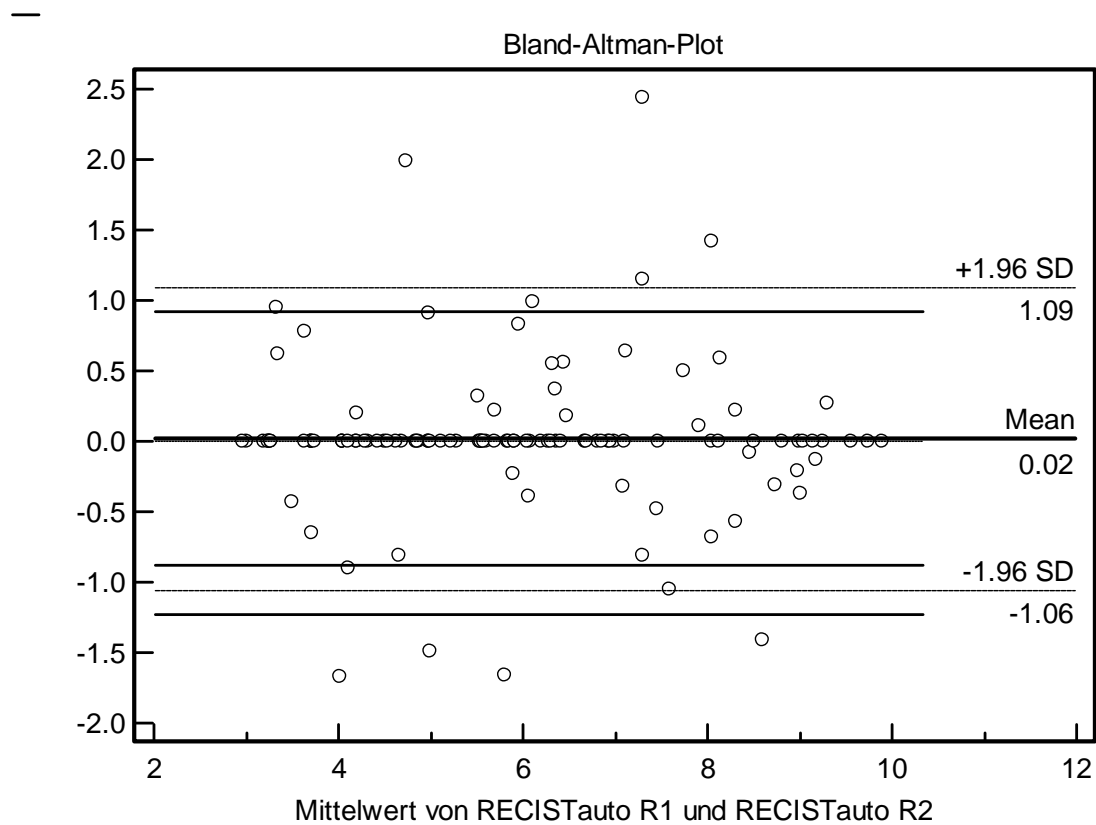


Abbildung 17: Bland-Altman-Diagramm zur semiautomatischen Messung des größten Durchmessers der Läsion durch zwei Untersucher (R1, R2) unter Ausschluss aller Läsionen größer als 10 mm. Differenz zwischen semiautomatisch ermitteltem Durchmesser von R1 und R2 in Relation zur Summe des Mittelwertes der semiautomatisch ermittelten Durchmesser von R1 und R2 für jede Messung.

Insgesamt zeigen die semiautomatisch ermittelten Messwerte keine statistisch signifikanten Unterschiede zwischen den beiden Untersuchern auf.

3.3.3 Tabellarische Übersicht der Ergebnisse

Tabelle 12: Übersicht der Ergebnisse der Messwerte bei gleicher Technik und Methode aber wechselnden Beobachtern. Aufgeführt sind die Ergebnisse des Vergleichs der manuellen Messung des größten Durchmessers der Läsionen von zwei Untersuchern (R1, R2) und der semiautomatischen Messung des größten Durchmessers der Läsionen von beiden Untersuchern.

| | Anzahl der gemessenen Werte | Mittelwert der Differenzen | Standardabweichung vom Mittelwert | P- Wert des Mittelwerts der Differenzen | P-Wert der mittleren prozentualen Abweichung der Differenzen von den Mittelwerten |
|---|-----------------------------|----------------------------|-----------------------------------|---|---|
| Manuelle Messung R1 vs. R2 | | | | | |
| Alle Läsionen | 430 | 1.85 | 2.25 | 0.06 | < 0.001 |
| Läsionen > 10 mm | 176 | 2.75 | 3.22 | 0.08 | < 0.05 |
| Läsionen < 10 mm | 254 | 1.17 | 0.83 | 0.07 | < 0.05 |
| Semiautomatische Messung R1 vs. R2 | | | | | |
| Alle Läsionen | 430 | 0.89 | 1.93 | 0.14 | 0.49 |
| Läsionen > 10 mm | 176 | 1.64 | 2.66 | 0.15 | 0.40 |
| Läsionen < 10 mm | 254 | 0.28 | 0.49 | 0.82 | 0.84 |

3.4 Bewertung der Übereinstimmung der Messwerte bei unterschiedlicher Technik und Methode und gleichen Beobachtern

3.4.1 Vergleich der manuellen mit der semiautomatischen Messung von Untersucher 1

Um den Vergleich der manuellen Messung durch einen geschulten Befunder mit der softwareunterstützten semiautomatischen Messung durch das Volumetrieprogramm CTSV-Prototype anzustellen, wurden die manuellen und semiautomatisch ermittelten Messwerte ($n = 430$) des ersten Untersuchers einander gegenübergestellt.

Der Mittelwert der Differenzen der manuell und semiautomatisch ermittelten Messwerte lag bei 1.35 mm bei einer Standardabweichung von 2.04 mm. Die statistische Auswertung mittels eines t-Tests zeigte eine hohe statistische Signifikanz bei einem p-Wert < 0.001 .

In dem Bland-Altman-Diagramm ist zu sehen, dass besonders im Bereich der größeren Läsionen die meisten Werte im negativen Bereich anzufinden sind. Das bedeutet, dass die manuellen Durchmesser im Vergleich zum semiautomatisch ermittelten Durchmesser kleiner gemessen wurden. Im Bereich der kleineren Läsionen unterhalb von 25 mm gibt es nur sehr wenige Übereinstimmungen der beiden Messmethoden, die Werte streuen in beide Richtungen (Abbildung 18).

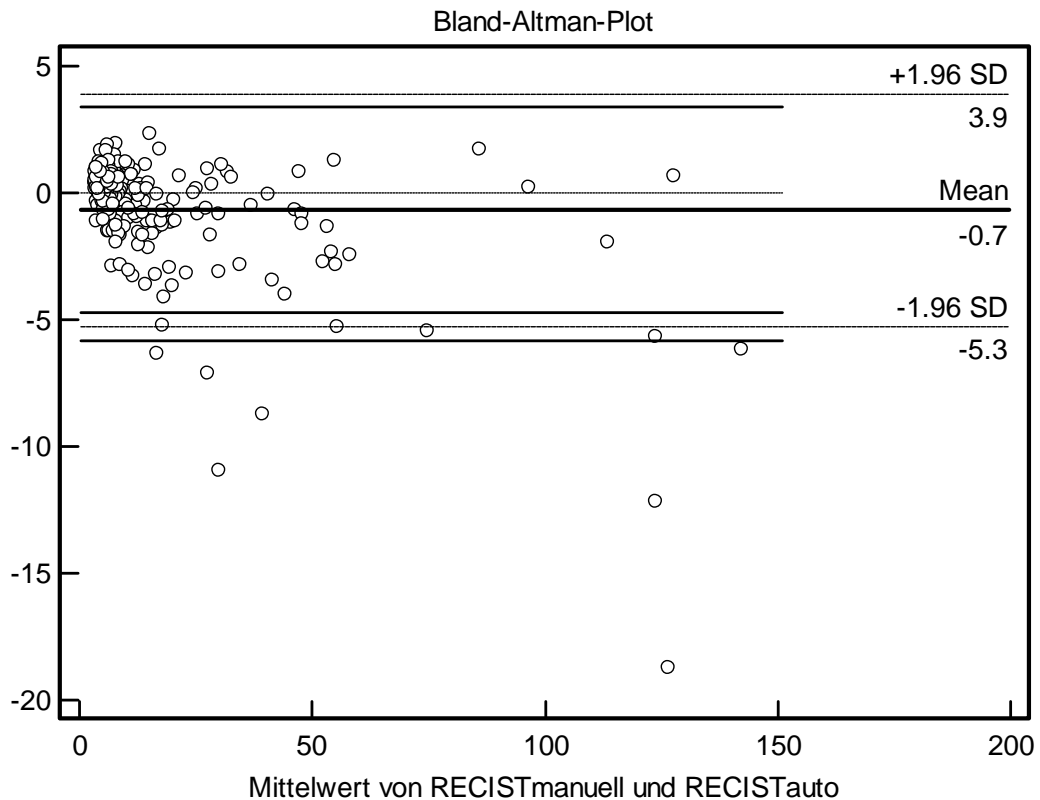


Abbildung 18: Bland-Altman-Diagramm zum Vergleich der manuellen und semiautomatischen Messung des größten Durchmessers einer Läsion durch den ersten Untersucher (R1). Differenz zwischen manuell und semiautomatisch ermitteltem Durchmesser in Relation zur Summe des Mittelwertes des manuell und semiautomatisch ermittelten Durchmessers für jede Messung.

Anhand eines Bland-Altman-Diagramms mit allen Läsionen größer als 10 mm ($n = 176$) können die oben beschriebenen Ergebnisse verdeutlicht werden. Besonders größere Läsionen oberhalb von 3 cm wurden durch die manuelle Messung kleiner bewertet als durch die semiautomatische computergestützte Methode (Abbildung 19). Der Mittelwert der Differenzen ergab 2.12 mm und die Standardabweichung lag bei 2.88 mm. Die Unterschiede waren bei einem p-Wert von < 0.001 hoch signifikant.

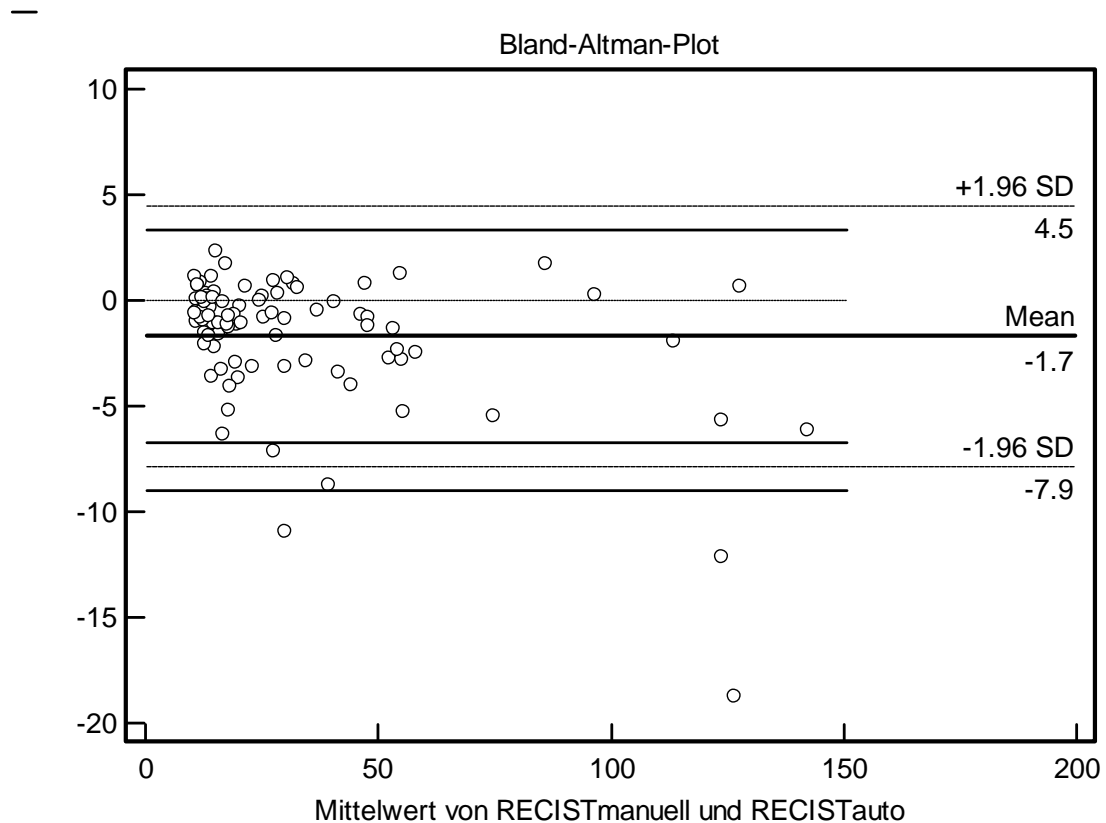


Abbildung 19: Bland-Altman-Diagramm zum Vergleich der manuellen und semiautomatischen Messung des größten Durchmessers einer Läsion durch den ersten Untersucher (R1) unter Ausschluss aller Läsionen kleiner als 10 mm. Differenz zwischen manuell und semiautomatisch ermitteltem Durchmesser in Relation zur Summe des Mittelwertes des manuell und semiautomatisch ermittelten Durchmessers für jede Messung.

Bei der Betrachtung aller Werte kleiner als 10 mm ($n = 254$) durch ein Bland-Altman-Diagramm lässt sich erkennen, dass nur wenige der manuell und semiautomatisch ermittelten Werte auf oder sehr nahe der Linie des Mittelwerts liegen. Es gibt eine breite Streuung der Werte in beide Richtungen (Abbildung 20).

Der Mittelwert der Differenzen der beiden Messwerte war 0.71 mm, die Standardabweichung 0.53 mm. Die Berechnung der prozentualen Abweichung der jeweils manuell und semiautomatisch gemessenen Werte zum oben genannten Mittelwert wies eine statistische Signifikanz von $p < 0.05$ auf.

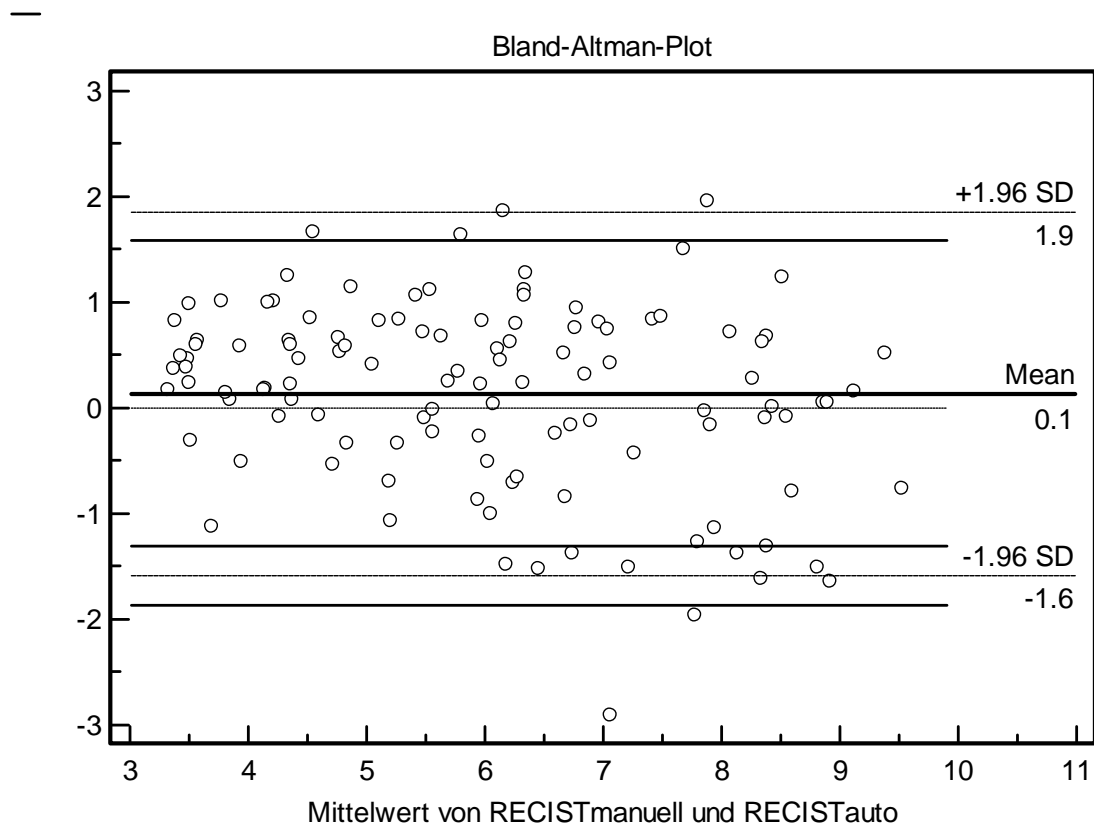


Abbildung 20: Bland-Altman-Diagramm zum Vergleich der manuellen und semiautomatischen Messung des größten Durchmessers einer Läsion durch den ersten Untersucher (R1) unter Ausschluss aller Läsionen größer als 10 mm. Differenz zwischen manuell und semiautomatisch ermitteltem Durchmesser in Relation zur Summe des Mittelwertes des manuell und semiautomatisch ermittelten Durchmessers für jede Messung.

3.4.2 Vergleich der manuellen mit der semiautomatischen Messung von Untersucher 2

Die manuell und semiautomatisch ermittelten Messwerte ($n = 430$) des größten Durchmessers der Läsionen durch den zweiten Untersucher (R2) zeigten im Vergleich folgende Ergebnisse: Der Mittelwert der Differenzen der manuellen und semiautomatischen Messwerte ergab 1.99 mm bei einer Standardabweichung von 1.87 mm. Bei einem p -Wert < 0.001 zeigten die Unterschiede zwischen den beiden Messmethoden eine hohe statistische Signifikanz.

In der Darstellung durch ein Bland-Altman-Diagramm lässt sich besonders bei den kleinen Werten eine vertikale Anordnung der Messwerte sowohl über- als auch unterhalb des Mittelwerts zeigen. Im Verlauf der größeren Messwerte divergieren die Ergebnisse stark vom Mittelwert (Abbildung 21).

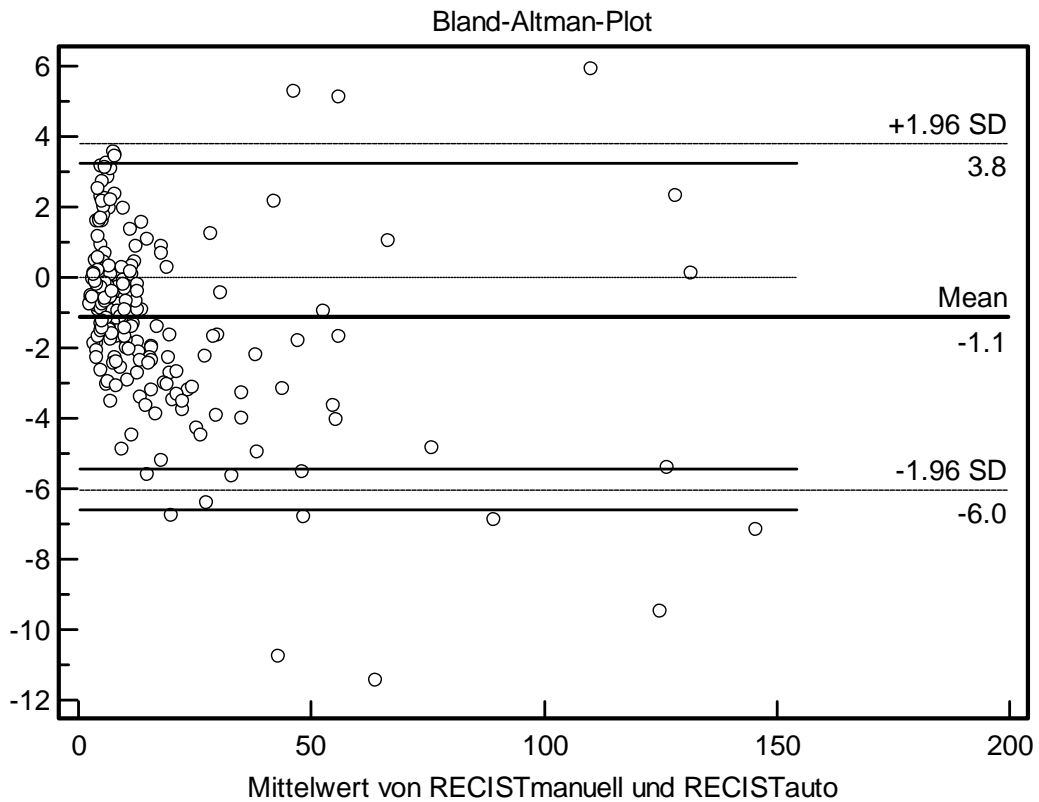


Abbildung 21: Bland-Altman-Diagramm zum Vergleich der manuellen und semiautomatischen Messung des größten Durchmessers einer Läsion durch den zweiten Untersucher (R2). Differenz zwischen manuell und semiautomatisch ermitteltem Durchmesser in Relation zur Summe des Mittelwertes des manuell und semiautomatisch ermittelten Durchmessers für jede Messung.

Bei Betrachtung aller Werte ($n = 176$) über 10 mm in einem Bland-Altman-Diagramm zeigen sich besonders bei den höheren Messwerten große Unterschiede (Abbildung 22). Auffallend ist der hohe Mittelwert der Differenzen von 3.03 mm bei einer Standardabweichung von 2.33 mm. Im Rahmen eines t-Tests waren die Differenzen stark signifikant ($p < 0.001$).

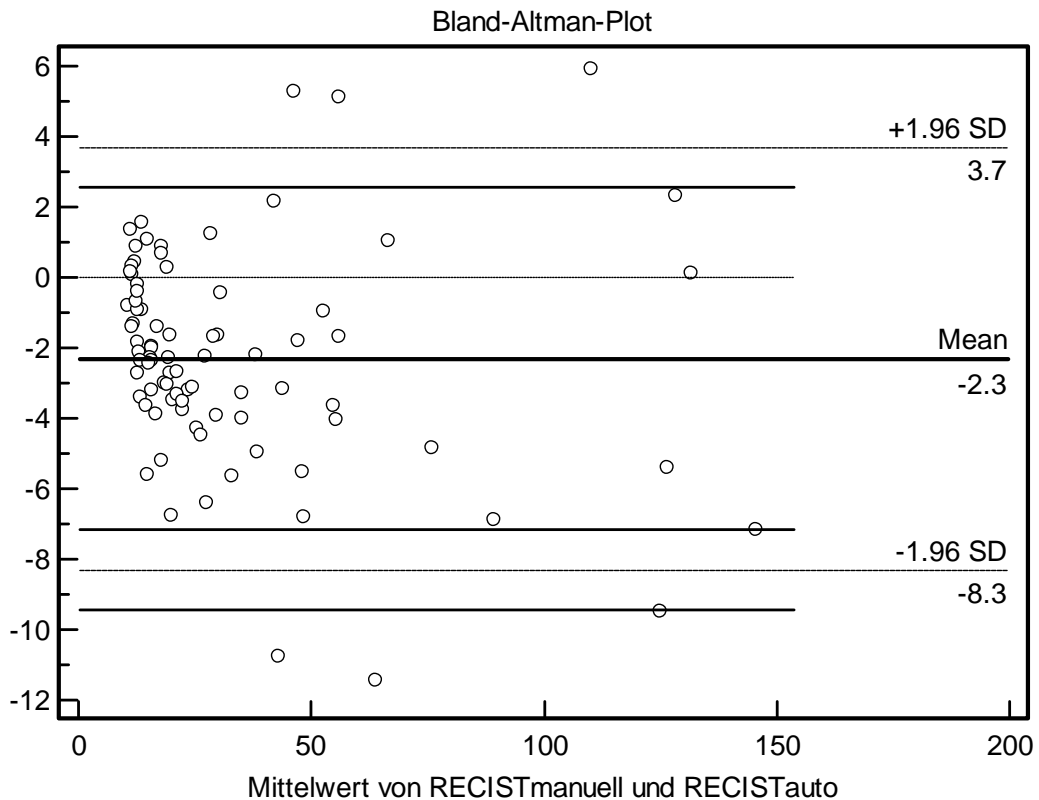


Abbildung 22: Bland-Altman-Diagramm zum Vergleich der manuellen und semiautomatischen Messung des größten Durchmessers einer Läsion durch den zweiten Untersucher (R2) unter Ausschluss aller Läsionen kleiner als 10 mm. Differenz zwischen manuell und semiautomatisch ermitteltem Durchmesser in Relation zur Summe des Mittelwertes des manuell und semiautomatisch ermittelten Durchmessers für jede Messung.

Unter Ausschluss der Läsionen, die größer als 10 mm sind, errechnete sich ein Mittelwert der Differenzen der Messwerte ($n = 254$) von 1.23 mm mit einer Standardabweichung von 0.97 mm. Der t-Test ergab bei diesen geringen Differenzen keine statistische Signifikanz ($p = 0.22$), bei Betrachtung der prozentualen Abweichung der jeweiligen manuellen und semiautomatischen Messwerte vom Mittelwert zeigte sich jedoch ein p-Wert von < 0.05 .

Die graphische Darstellung mittels eines Bland-Altman-Diagramms verdeutlicht, dass sich nur einzelne Messwerte in der Nähe des Mittelwerts bewegen. Im Vergleich zur semiautomatischen Messung wurde der manuelle Wert der Läsionen unter 10 mm in den meisten Fällen kleiner gemessen, was sich anhand der negativen Messpunkte unterhalb des Mittelwerts zeigt (Abbildung 23).

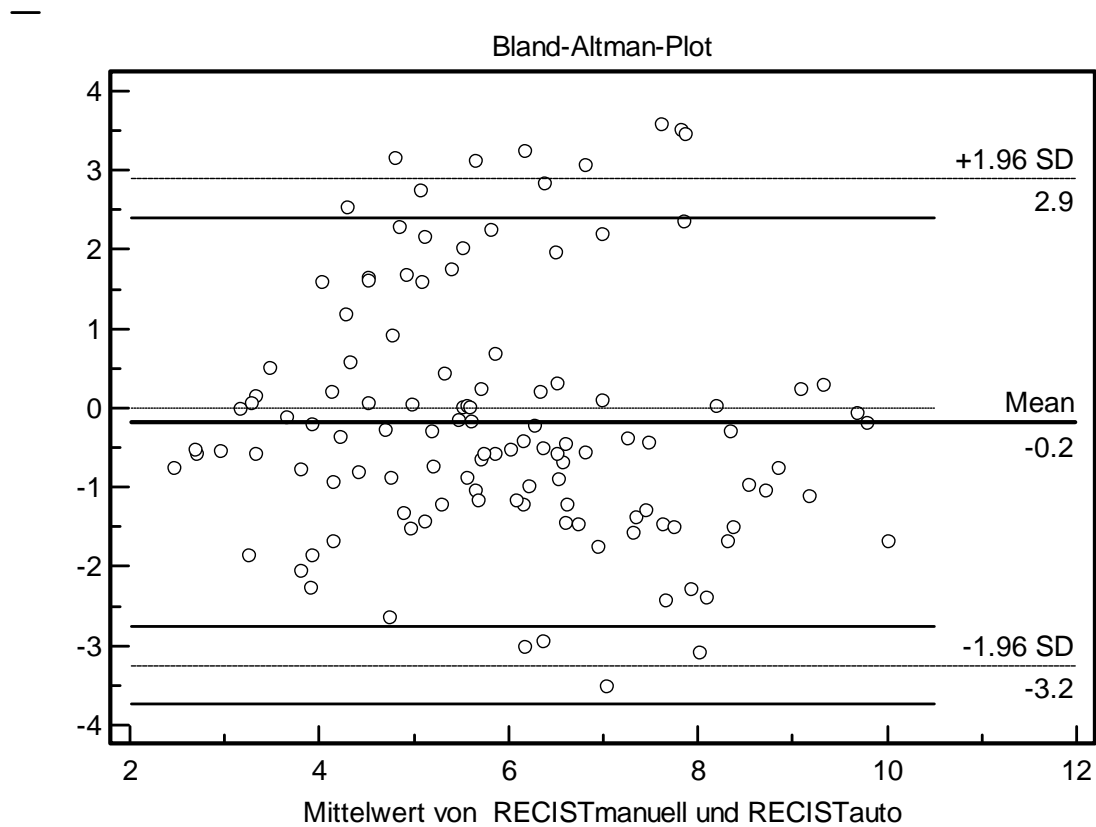


Abbildung 23: Bland-Altman-Diagramm zum Vergleich der manuellen und semiautomatischen Messung des größten Durchmessers einer Läsion durch den zweiten Untersucher (R2) unter Ausschluss aller Läsionen größer als 10 mm. Differenz zwischen manuell und semiautomatisch ermitteltem Durchmesser in Relation zur Summe des Mittelwertes des manuell und semiautomatisch ermittelten Durchmessers für jede Messung.

3.4.3 Tabellarische Übersicht der Ergebnisse

Tabelle 13: Übersicht der Messwerte bei unterschiedlicher Technik und Methode und gleichen Beobachtern. Aufgeführt sind die Ergebnisse des Vergleichs zwischen manueller und semiautomatischer Messung des größten Durchmessers der Läsionen jeweils von Untersucher 1 und Untersucher 2.

| | Anzahl der gemessenen Werte | Mittelwert der Differenzen | Standardabweichung vom Mittelwert | P-Wert des Mittelwerts der Differenzen | P-Wert der mittleren prozentualen Abweichung der Differenzen von den Mittelwerten |
|---|-----------------------------|----------------------------|-----------------------------------|--|---|
| Manuell vs. Semiautomatisch Reader 1 | | | | | |
| Alle Läsionen | 430 | 1.35 | 2.04 | < 0.001 | 0.42 |
| Läsionen > 10 mm | 176 | 2.12 | 2.88 | < 0.001 | < 0.001 |
| Läsionen < 10 mm | 254 | 0.71 | 0.53 | 0.1 | < 0.05 |
| Manuell vs. Semiautomatisch Reader 2 | | | | | |
| Alle Läsionen | 430 | 1.99 | 1.87 | < 0.001 | < 0.001 |
| Läsionen > 10 mm | 176 | 3.03 | 2.33 | < 0.001 | < 0.001 |
| Läsionen < 10 mm | 254 | 1.23 | 0.97 | 0.22 | < 0.05 |

3.5 Interaktionszeitanalyse mit Hinblick auf die Praktikabilität

Für alle 430 Messungen wurde durch die Volumetrie-Software CTSV-Prototype (MeVis, Bremen) automatisch die Zeit (in Sekunden) berechnet, die vom ersten Identifizieren der Läsion durch Mausklicken bis zur vollständigen Berechnung sämtlicher Parameter benötigt wurde.

Der Mittelwert betrug 3.32 s, die Standardabweichung 3.27 s. Die schnellste Messung dauerte 0.84 s, die längste Messdauer betrug 23.92 s (Tabelle 14).

Tabelle 14: Parameter der Zeitanalyse durch das Volumetrie-Programm CTSV-Prototype. Dargestellt sind der Mittelwert, die Standardabweichung, die schnellste und die längste Messdauer für alle Messungen (n = 430).

| Anzahl der Metastasen n = 430 | Zeitanalyse |
|-------------------------------|-------------|
| Mittelwert [s] | 3.32 |
| Standardabweichung [s] | 3.27 |
| Minimum [s] | 0.84 |
| Maximum [s] | 23.92 |

Im Vergleich zu den automatisch ermittelten Zeitwerten bei Nutzung des Volumetrie-Programms dauert die manuelle Untersuchung einer Läsion durch einen erfahrenen Befunder dennoch ein Vielfaches davon. Je nach Größe, Lage und Form des Tumors kann im günstigsten Fall von etwa 30 Sekunden ausgegangen werden. In der Computertomographie muss zunächst der größte Durchmesser der Läsion durch scrollen durch die einzelnen Schnitte identifiziert werden und der Durchmesser der Läsion mittels eines digitalen Lineals gemessen werden. Im Falle einer Follow-Up-Untersuchung müssen die entsprechenden Werte nach RECIST berechnet und in eine Responsekategorie eingeteilt werden. Im ungünstigsten Fall ist eine Läsion für das menschliche Auge anhand ähnlicher Dichtewerte der Umgebung so schwer abgrenzbar, dass der größte Durchmesser nur zu schätzen ist und die Befundung des Tumors dementsprechend ungenau ist und lange dauert.

4 Diskussion

Ziel der vorliegenden Studie war ein Vergleich unterschiedlicher Messmethoden zur Quantifizierung von Lungentumoren.

Die Reliabilität eines Messverfahrens von Tumoren bedingt die klinischen Konsequenzen, die aus dieser Größenbestimmung resultieren. Zur Beurteilung der Messmethoden wurde der Vergleich der Methoden an sich, sowie der Schwellenwerte zwischen manuell und volumetrisch bestimmter Tumorgröße angestellt und die Reproduzierbarkeit der Messwerte anhand eines Interobserver-Vergleichs bestimmt.

Der momentane Goldstandard zur Quantifizierung der Tumorlast sind die RECIST-Kriterien, die aus den modifizierten WHO-Kriterien entstanden. Gegenstand der aktuellen Forschung ist bereits die Verbesserung von automatisierten Volumetrie-Softwareprogrammen und Etablierung im klinischen Alltag um ein genaueres Evaluieren der Therapiereponse zu ermöglichen. Häufig zeigen sich intrapulmonale Läsionen irregulär konfiguriert und nicht streng rund oder ellipsoid. Uni- oder bidimensionale Verfahren können somit nicht sicher die Tumorausdehnung in allen drei Ebenen erfassen und sich als ungenau erweisen. Veränderungen in der Tumorausdehnung in Verlaufskontrollen werden mit der Volumetrie akkurater und schneller erfasst als mit linearen Messungen. Insgesamt zeigt die aktuelle Studienlage zum Vergleich von uni-, bi- und dreidimensionalen Messverfahren den Vorteil der Messgenauigkeit der Volumetrie (Revel, Bissery et al. 2004; Tran, Brown et al. 2004; Goodman, Gulsun et al. 2006; Marten, Auer et al. 2006; Zhao, Schwartz et al. 2006; Fabel and Bolte 2008). Eine Durchmesseränderung einer Läsion von ca. 25% geht mit einer vergleichsweise hohen Volumenänderung von bis zu 70 - 100% einher, je nach Form der Läsion. Um eine sichere Einteilung in Responsekategorien zu gewährleisten, die das Therapieregime des Patienten entscheidend beeinflussen, ist eine möglichst exakte Quantifizierung der Läsionen notwendig.

Zur Bewertung der Übereinstimmung der Messwerte bei unterschiedlicher Technik und Methode, aber gleichem Beobachter wurden die manuellen Messwerte mit den semiautomatisch ermittelten Messwerten miteinander verglichen. Der Mittelwert der Differenzen der manuellen und semiautomatischen Messwerte lag bei Untersucher 1 bei 1.35 mm, SD 2.04 mm und bei Untersucher 2 bei 1.99 mm mit einer

Standardabweichung von 1.87 mm. Der mittels t-Test berechnete p-Wert wies eine statistisch hohe Signifikanz auf ($p < 0.001$). Deutlich werden die Unterschiede zwischen den beiden Messmethoden in der graphischen Aufstellung nach Bland-Altman. Bei beiden Untersuchern zeigen die manuellen Messergebnisse im Vergleich zu den semiautomatisch ermittelten Werten, insbesondere bei größeren Läsionen, insgesamt kleiner gemessene Werte. Im Bereich der kleinen Läsionen ergeben sich wenige Übereinstimmungen.

Da bei dem untersuchten Patientenkollektiv die exakten Volumina der Lungenläsionen nicht bekannt sind, war es nicht möglich die Präzision der Messverfahren zu evaluieren. Anhand verschiedener Phantomstudien in den letzten Jahren konnte jedoch gezeigt werden, dass eine Volumetrie zu einer präziseren Bestimmung des Phantomvolumens führt als eine Berechnung durch den maximalen Durchmesser (Breiman, Beck et al. 1982; Van Hoe, Haven et al. 1997; Yankelevitz, Reeves et al. 2000).

Beim Vergleich der manuell gemessenen Durchmesser und der korrespondierenden volumetrischen Werte unserer Studie und nach ihrer Einteilung in die jeweilige Responsekategorie nach RECIST, zeigte sich insgesamt eine Differenz in 32% der Fälle. 18 manuell gemessene Durchmesser fielen in die Kategorie „stable disease“, wovon volumetrisch erfasst drei Patienten der Kategorie „progressive disease“ und ein Patient eine „partial response“ zeigte. Sechs der manuell gemessenen Untersuchungen wurden der Kategorie „progressive disease“ zugeteilt, in der Volumetrie zeigten drei der Patienten jedoch eine „stable disease“. Die manuell gemessenen Durchmesser in der Kategorie „partial response“ zeigten sich in der Volumetrie durchweg als „stable disease“. Durch unzureichende Erfassung der Tumorausdehnung können demnach folgenreiche falsche Einteilungen und Änderungen des Therapieregimes zu unnötigem Wechsel einer Chemotherapie bei tatsächlichem Therapieansprechen bis hin zu unzureichender Therapie bei Progression der Krankheit entstehen.

Zu einer ähnlichen Einschätzung der Situation kommen Marten et al. in einer vergleichenden Studie zur Therapiereponse mit manuell und automatisch gemessenen pulmonalen Metastasen. In 14 von 50 Patienten (28%) zeigte sich eine unterschiedliche Einteilung in Responsekategorien basierend auf manueller oder automatischer Messung (Marten, Auer et al. 2006).

In der Studie von Sohaib et al. wurde die Tumorummetrie von Keimzelltumoren und Lymphomen mit Anwendung der RECIST- und WHO-Kriterien bei der Verlaufsbeurteilung verglichen. Erstaunlicherweise zeigte sich hier eine 90%ige Übereinstimmung der Einteilung in Resonsekategorien. Das Team stellte die These auf, dass diese hohe Konkordanz an den untersuchten Tumorentitäten liegen müsse, die ein eher gleichmäßigeres Wachstum zeigen als pulmonale Läsionen (Sohaib, Turner et al. 2000).

Die Ergebnisse dieser beiden Studien verdeutlichen die klinisch potentiell relevanten Unterschiede der Tumorquantifizierung und Tumorresponse-Kategorisierung in Abhängigkeit der Konfiguration der Läsion. Die Volumetrie ermöglicht die vollständige Erfassung unregelmäßig wachsender Tumorkläsionen, eine frühere Detektion eines Progresses und eine dementsprechend angepasste Therapiestrategie.

Ein weiterer wichtiger Parameter, der in dieser Studie untersucht wurde, ist die Reproduzierbarkeit der Ergebnisse. Hierzu wurden alle Messungen ($n = 430$) durch zwei unabhängige Untersucher sowohl manuell, als auch semiautomatisch vermessen und die Ergebnisse in Responsekategorien gegliedert. Die Reproduzierbarkeit der manuellen Messung der Durchmesser lag bei 84.6%. In 4 von 26 Fällen (15.4%) würde dies in einer divergenten Therapieentscheidung resultieren. Für die semiautomatische Messung zeigte sich eine Reproduzierbarkeit von 92.3% mit einer geringeren Abweichung der Tumorresponseeinteilung bei nur 2 von 26 Follow-Up-Untersuchungen (7.7 %).

Der Interobserver-Vergleich der manuell gemessenen Durchmesser ergab einen Mittelwert der Differenzen von 1.84 mm und eine Standardabweichung von 2.25 mm. Die prozentuale Abweichung der Einzelmessungen zum Mittelwert wies ein hohes Signifikanzniveau auf ($p < 0.001$). Bei der semiautomatisch durchgeführten Untersuchung ergab sich ein geringerer Mittelwert der Differenzen von 0.89 mm mit einer Standardabweichung von 1.93 mm. Die geringen Messunterschiede zwischen den Untersuchern waren nicht signifikant. Bei der graphischen Darstellung der Untersuchungsergebnisse durch Bland-Altman-Diagramme wurden bei den manuellen Messungen besonders die differenten Messwerte der größeren Läsionen deutlich. Eindrücklich war auch die große Übereinstimmung der semiautomatisch gemessenen Werte bei der Aufschlüsselung der Läsionen unter 10 mm.

Diese Ergebnisse werden in der Literatur bestätigt. Marten et al. zeigten in einer Vergleichsstudie der RECIST-Kriterien mit volumetrischen Messungen durch zwei unabhängige Radiologen an 50 Patienten einen signifikant höheren interobserverbedingten relativen Messfehler zwischen den beiden Methoden (Marten, Auer et al. 2006). Revel et al. untersuchten 54 pulmonale Läsionen durch drei unabhängige Untersucher und kamen zu dem Ergebnis, dass zweidimensionale Messungen, durch hohe Intra- und Interobserverunterschiede, nicht in dem Maße zur Verlaufsbeobachtung vor allem kleiner Rundherde geeignet sind, wie die Volumetrie (Revel, Bissery et al. 2004). Die softwareunterstützte semiautomatische Volumenbestimmung, insbesondere intrapulmonaler Läsionen, weist eine hohe Reliabilität und Reproduzierbarkeit bei der Bestimmung der Tumorlast und Follow-Up-Untersuchungen auf (Wormanns, Kohl et al. 2004; Bolte, Riedel et al. 2006). Auch bei Nutzung von Computertomographen verschiedener Firmen, wie es beispielsweise bei einem Follow-up eines Patienten im klinischen Alltag mit externen Vorbefunden möglich wäre, erwies sich die Volumetrie im Vergleich zur manuellen Messung als akkurater (Das, Ley-Zaporozhan et al. 2007).

Im Hinblick auf die Praktikabilität eines semiautomatischen Volumetrieprogramms in der klinischen Routine wurde im Rahmen dieser Studie eine Interaktionszeitanalyse der volumetrischen Messungen durchgeführt. Es wurde die Zeit vom ersten Anklicken der Läsion bis zur vollständigen Berechnung sämtlicher Parameter berechnet. Im Mittel betrug die Zeit zur semiautomatischen Quantifizierung einer Läsion 3.32 s, mit einer Standardabweichung von 3.27 s. Die kürzeste Messdauer lag bei 0.84 s, die längste Messung dauerte 23.92 s.

Fabel et al. verglichen in ihrer Studie die semiautomatische Volumetrie mit den RECIST-Kriterien an Lymphknotenmetastasen, ebenfalls mit einem Software-Programm von MeVis. Die Messdauer der volumetrischen Analyse lag im Mittel deutlich unter der manuellen Messung, 57 Sekunden für die Volumetrie im Vergleich zu 108 Sekunden bei manuell gemessenen Durchmesser (Fabel, von Tengg-Kobligk et al. 2008).

Ebenfalls wichtige Parameter in der volumetrischen Messung, die nicht Teil der Fragestellung dieser Studie waren, sind die Einflüsse verschiedener Rekonstruktionsparameter, wie Schichtdicke oder Fenstereinstellungen, sowie

Differenzierung der Läsionen in juxtavaskuläre, juxtapleurale oder intrapulmonale Lungenrundherden.

Ravenel et al. untersuchten den Einfluss von Rekonstruktionsparametern anhand einer Phantomstudie mit 27 sphärischen und 29 lobulierten Phantomen mit einem Durchmesser von 3 - 16 mm. Es ließ sich kein klinisch signifikanter Unterschied der Messungen in Bezug auf die Fenstereinstellung finden. Die Schichtdicke wies jedoch unterschiedliche Ergebnisse, je nach Größe der Läsion, auf. Die genauesten Ergebnisse ließen sich bei einer Schichtdicke von 0.625 mm nachweisen. Für Läsionen größer als 8 mm erwies sich eine Schichtdicke von 2.50 mm als akzeptabel. Die Gruppe riet von der Nutzung von 5 mm - Schichtdicken ab (Ravenel, Leue et al. 2008).

Vogel et al. untersuchten 101 Lungenrundherde bei insgesamt 28 Patienten mit Schichtdicken-Rekonstruktionen von 1 -, 3 - und 5 mm. Die volumetrischen Messungen der Läsionen zeigten bei 3 mm eine vergleichbare Genauigkeit wie bei 1 mm und erwiesen sich sogar bei 5 mm genauer als ein- oder zweidimensionale Messungen (Vogel, Vonthein et al. 2008).

Die Untersuchung der Läsionen in dieser Studie orientiert sich an den aktuellen Forschungsergebnissen und erfolgte unter standardisierten Protokolleinstellungen mit einer Schichtdicke von 2 mm und Durchführung der Messung im Lungenfenster.

Wang et al. evaluierten in einer großen retrospektiven Studie mit 4225 Lungenrundherden bei insgesamt 2239 Patienten den Einfluss von Morphologie, Größe und Sitz der Läsionen. Die Interreader-Übereinstimmung lag bei 86%. Alle drei Parameter beeinflussten die volumetrische Messung signifikant. Die größte Variabilität in der Volumenmessung hatten irregulär konfigurierte, juxtavaskulär gelegene Läsionen (Wang, van Klaveren et al. 2008). Auch Gietema et al. zeigten an 218 Lungenrundherden, dass die Präzision der Segmentierung einer Läsion signifikant von ihrer Form abhängig ist (Gietema, Schaefer-Prokop et al. 2007). Obwohl der Fortschritt der Volumetrie-Programme innerhalb der letzten Jahre deutlich ist, ist eine manuelle Nachkorrektur im Algorithmus bei schwer segmentierbaren Läsionen als sinnvoll anzusehen.

Die Schwellenwerte, sowie die Anzahl der zu messenden Läsionen zur Einteilung der Tumorresponse wurden für die Volumetrie von Lungenrundherden angepasst an die RECIST-Kriterien. Mit 65%, respektive 73% Volumenänderung im Vergleich zur

Änderung des Durchmessers von 30%, respektive 20%, sind die Schwellenwerte für eine volumetrisch erfasste Änderung als sehr hoch gesetzt anzusehen. Marten et al. suchten in ihrer Studie den geringsten Schwellenwert der Volumenänderung, der weiterhin eine hohe Reproduzierbarkeit der Ergebnisse aufweist, sowie die minimale Anzahl an Target-Läsionen bei diesem Schwellenwert. Es zeigte sich eine potentielle Reduzierung des Schwellenwerts für die Volumetrie auf 35% mit reproduzierbaren Ergebnissen bei einer Messung von mindestens drei Läsionen.

Sollte die Volumetrie als Goldstandard in den klinischen Alltag integriert werden, werden diese Ergebnisse durch weitere Studien zur Präzisierung genauer Schwellenwerte vervollständigt werden.

Die Volumetrie von Lungentumoren erweist sich im Vergleich zur manuellen Messung des größten Durchmessers nach RECIST-Kriterien als vorrangig.

Durch die Entwicklung spezieller und optimierter Software-Algorithmen bietet die Volumetrie bei einer Einführung in die klinische Routine den Vorteil eines niedrigeren Zeit- und Untersuchungsaufwandes. Die Detektion und automatische Berechnung des Volumens bereits bekannter intrapulmonaler Rundherde in Follow-Up-Untersuchungen führt zu einer wesentlichen Erleichterung im klinischen Alltag.

Die volumetrische Messung kann asymmetrisch konfigurierte Tumorkläsionen genauer erfassen und eine frühere Detektion eines Tumorprogresses/-regresses ermöglichen. Besonders kleinere Läsionen werden durch eine manuelle Messung unterschätzt. Die rechtzeitige Erkennung einer Wachstumsänderung des Tumors zieht wichtige klinische Entscheidungen nach sich. Bei einem verminderten Therapieansprechen kann so eine zeitnahe Änderung des Therapieregimes durchgeführt werden, die möglicherweise bessere Heilungschancen hat.

Im Forschungsbereich, wenn es um die Überlegenheit einer neuen Therapie im Vergleich mit der Standardtherapie geht, wird die Bewertung der Wirksamkeit anhand der manuell gemessenen RECIST-Kriterien durchgeführt. Die Nutzung der Tumorumetrie könnte hier zu einer besseren Einschätzung der Therapieresponse und somit zu einer genaueren Beurteilung des zu testenden Medikaments führen.

Die Möglichkeit der Einführung der Tumorumetrie als Goldstandardmethode zur Quantifizierung der Tumorklast im klinischen Alltag, wie auch in der Forschung, wird in Zukunft eingehend betrachtet werden müssen.

5 Zusammenfassung

5.1 Zusammenfassung

Die Quantifizierung der Tumorlast eines Patienten spielt sowohl in Studien, als auch im klinischen Alltag, eine große Rolle beim Therapiemonitoring. Als Goldstandard gelten die Messkriterien nach RECIST mit der Bestimmung des längsten Tumordurchmessers. In der vorliegenden Studie wurde ein Vergleich der Tumolvolumetrie mit der Messung nach RECIST-Kriterien aufgeführt. Anhand von CT-Untersuchungen von Patienten mit Lungenläsionen wurden die unterschiedlichen Messmethoden miteinander verglichen, die Reproduzierbarkeit der beiden Messmethoden durch zwei Untersucher dargestellt, sowie die Einteilung in Tumorresponsekategorien nach RECIST-Schwellenwerten zur Beurteilung des Unterschieds zwischen den Messmethoden aufgeführt. Eine Interaktionszeitanalyse zur Beurteilung der Praktikabilität im klinischen Alltag wurde durchgeführt.

- Es konnte gezeigt werden, dass ein Vergleich der Einteilung der manuell gemessenen Durchmesser mit den volumetrischen Daten in korrespondierende Tumorresponsekategorien nach RECIST zu einer Differenz von 32% der Fälle führte. Dies bedeutet unter Umständen eine divergente Therapieentscheidung für den Patienten, je nach Messmethode.
- Die Untersuchung der Reproduzierbarkeit der beiden Messmethoden durch zwei verschiedene Untersucher ergab eine Reproduzierbarkeit der manuellen Messung der Durchmesser von 84.6%. Bei der volumetrischen Messung lag die Reproduzierbarkeit zwischen zwei Untersuchern bei 92.3%.
- Die Bewertung der Übereinstimmung bei unterschiedlicher Technik und Methode, aber gleichem Untersucher zeigte bei beiden Untersuchern Mittelwerte der Differenzen zwischen manueller und semiautomatischer Messung, die bei 1.35 mm, respektive 1.99 mm lagen. Die Unterschiede wiesen eine hohe Signifikanz auf ($p < 0.001$) In der graphischen Darstellung mittels Bland-Altman-Diagrammen wurde deutlich, dass die manuell gemessenen

Werte insgesamt kleiner gemessen werden als in der semiautomatischen Evaluierung.

- Mittels einer Interaktionszeitanalyse wurde die Volumetrie-Software auf ihre Praktikabilität im klinischen Alltag untersucht. Mit Untersuchungszeiten von 0.84 s im schnellsten, 23.92 s im längsten Fall und im Durchschnitt mit einem Wert von 3.32 s erwies sich die Volumetrie als ein schnelles Verfahren, durch das die Zeit der Untersuchungs-Befundung herabgesetzt werden kann.

Zusammenfassend zeigt die vorliegende Studie, dass sich durch die Nutzung einer semiautomatischen Volumetrie-Software Lungenläsionen in allen drei Ebenen vollständiger, schneller und mit einer besseren Reproduzierbarkeit erfasst werden können als durch den momentanen Goldstandard der eindimensionalen manuellen Messung nach RECIST-Kriterien.

5.2 Summary

The quantification of a patient's tumor burden plays an important role in clinical studies, as well as in the clinical routine for tumor response assessment. Current gold standards are the RECIST criteria based on the measurement of the maximum axial diameter of a tumor lesion. This study compares volumetric assessment of a tumor with manual measurement following the RECIST guidelines. CT-scans of patients with lung nodules were used for comparison of the different measurement methods, evaluation of the reproducibility of the two measurement methods through two readers, as well as arranging the tumor lesions into different tumor response categories to evaluate the differences between the measurement methods. A time analysis was conducted for practicability purposes in the clinical practice.

- Comparison of manually measured diameter and volumetric data of a tumor lesion led to different classification in the RECIST tumor response categories. In 32% of the studied cases it could have caused divergent therapy decisions depending on the measurement method.
- The investigation of reproducibility of both methods by two different readers showed a correspondence of 84.6% for manual diameter measurement and 92.3% for volumetry, respectively.
- Assessment of agreement using different technique and method but the same reader revealed mean values of the differences between manual and semi-automatic measurement of 1.35 mm, and 1.99 mm respectively, with a high significance ($p < 0.001$). By means of Bland-Altman-Plots it could be demonstrated that manually collected data was generally measured smaller than the semi-automatic.
- The volumetric software was tested for practicability in the daily routine by conducting a time analysis. With 0.84 s in the fastest and 23.92 s in the longest case and a mean value of 3.32 s it demonstrated to be a quick method. Time needed for a CT-Evaluation can be reduced.

In summary, this study shows that by using a semi-automatic volumetric software lung lesions can be measured in all three dimensions, less time consuming and with better reproducibility than the current gold standard of manual one dimensional measurement by RECIST.

6 Literaturverzeichnis

- Bland, J. M. and D. G. Altman (1986). "Statistical methods for assessing agreement between two methods of clinical measurement." Lancet **1**(8476): 307-10.
- Bogaerts, J., R. Ford, et al. (2009). "Individual patient data analysis to assess modifications to the RECIST criteria." Eur J Cancer **45**(2): 248-60.
- Bolte, H., T. Jahnke, et al. (2007). "Interobserver-variability of lung nodule volumetry considering different segmentation algorithms and observer training levels." Eur J Radiol **64**(2): 285-95.
- Bolte, H., C. Riedel, et al. (2006). "Reproducibility of computer-aided volumetry of artificial small pulmonary nodules in ex vivo porcine lungs." Invest Radiol **41**(1): 28-35.
- Bornemann, L., J. M. Kuhnigk, et al. (2005). "Informatics in radiology (infoRAD): new tools for computer assistance in thoracic CT part 2. Therapy monitoring of pulmonary metastases." Radiographics **25**(3): 841-8.
- Breiman, R. S., J. W. Beck, et al. (1982). "Volume determinations using computed tomography." AJR Am J Roentgenol **138**(2): 329-33.
- Collins, L. G., C. Haines, et al. (2007). "Lung cancer: diagnosis and management." Am Fam Physician **75**(1): 56-63.
- Dachman, A. H., P. M. MacEaney, et al. (2001). "Tumor size on computed tomography scans: is one measurement enough?" Cancer **91**(3): 555-60.
- Das, M., J. Ley-Zaporozhan, et al. (2007). "Accuracy of automated volumetry of pulmonary nodules across different multislice CT scanners." Eur Radiol **17**(8): 1979-84.
- Disler, D. G., D. S. Marr, et al. (1994). "Accuracy of volume measurements of computed tomography and magnetic resonance imaging phantoms by three-dimensional reconstruction and preliminary clinical application." Invest Radiol **29**(8): 739-45.
- Eisenhauer, E. A., P. Therasse, et al. (2009). "New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1)." Eur J Cancer **45**(2): 228-47.
- Erasmus, J. J. and B. S. Sabloff (2008). "CT, positron emission tomography, and MRI in staging lung cancer." Clin Chest Med **29**(1): 39-57, v.
- Fabel, M. and H. Bolte (2008). "[Automated procedure for volumetric measurement of metastases: estimation of tumor burden]." Radiologe **48**(9): 857-62.
-

- Fabel, M., H. von Tengg-Koblighk, et al. (2008). "Semi-automated volumetric analysis of lymph node metastases in patients with malignant melanoma stage III/IV--a feasibility study." Eur Radiol **18**(6): 1114-22.
- Fraioli, F., L. Bertolotti, et al. (2006). "Volumetric evaluation of therapy response in patients with lung metastases. Preliminary results with a computer system (CAD) and comparison with unidimensional measurements." Radiol Med **111**(3): 365-75.
- Gietema, H. A., C. M. Schaefer-Prokop, et al. (2007). "Pulmonary nodules: Interscan variability of semiautomated volume measurements with multisection CT--influence of inspiration level, nodule size, and segmentation performance." Radiology **245**(3): 888-94.
- Goodman, L. R., M. Gulsun, et al. (2006). "Inherent variability of CT lung nodule measurements in vivo using semiautomated volumetric measurements." AJR Am J Roentgenol **186**(4): 989-94.
- Gottlieb, R. H., A. Litwin, et al. (2008). "Qualitative radiology assessment of tumor response: does it measure up?" Clin Imaging **32**(2): 136-40.
- Graser, A., C. R. Becker, et al. (2008). "[Volumetry of metastases from renal cell carcinoma: comparison with the RECIST criteria]." Radiologe **48**(9): 850-6.
- Hauth, E. A., J. Stattaus, et al. (2007). "[Comparison of unidimensional and bidimensional measurement to assess therapeutic response in the treatment of solid tumors]." Radiologe **47**(7): 628, 630-4.
- Hein, P. A., V. C. Romano, et al. "Variability of semiautomated lung nodule volumetry on ultralow-dose CT: comparison with nodule volumetry on standard-dose CT." J Digit Imaging **23**(1): 8-17.
- Hein, P. A., V. C. Romano, et al. (2009). "Linear and volume measurements of pulmonary nodules at different CT dose levels - intrascan and interscan analysis." Rofo **181**(1): 24-31.
- Hensing, T. A. (2005). "Clinical evaluation and staging of patients who have lung cancer." Hematol Oncol Clin North Am **19**(2): 219-35, v.
- Hofer, M. (2006). CT-Kursbuch. Düsseldorf, Didamed Verlag GmbH.
- Hopper, K. D., C. J. Kasales, et al. (1996). "The impact of 2D versus 3D quantitation of tumor bulk determination on current methods of assessing response to treatment." J Comput Assist Tomogr **20**(6): 930-7.

- Jaffe, C. C. (2006). "Measures of response: RECIST, WHO, and new alternatives." J Clin Oncol **24**(20): 3245-51.
- James, K., E. Eisenhauer, et al. (1999). "Measuring response in solid tumors: unidimensional versus bidimensional measurement." J Natl Cancer Inst **91**(6): 523-8.
- Klein, J. S. and S. Braff (2008). "Imaging evaluation of the solitary pulmonary nodule." Clin Chest Med **29**(1): 15-38, v.
- Levy, M. A. and D. L. Rubin (2008). "Tool support to enable evaluation of the clinical response to treatment." AMIA Annu Symp Proc: 399-403.
- Marten, K., F. Auer, et al. (2006). "Inadequacy of manual measurements compared to automated CT volumetry in assessment of treatment response of pulmonary metastases using RECIST criteria." Eur Radiol **16**(4): 781-90.
- Marten, K., F. Auer, et al. (2007). "Automated CT volumetry of pulmonary metastases: the effect of a reduced growth threshold and target lesion number on the reliability of therapy response assessment using RECIST criteria." Eur Radiol **17**(10): 2561-71.
- Marten, K., E. J. Rummeny, et al. (2005). "[Computer-aided diagnosis and volumetry of pulmonary nodules: current concepts and future perspectives]." Rofo **177**(2): 188-96.
- Mazumdar, M., A. Smith, et al. (2004). "A statistical simulation study finds discordance between WHO criteria and RECIST guideline." J Clin Epidemiol **57**(4): 358-65.
- Park, J. O., S. I. Lee, et al. (2003). "Measuring response in solid tumors: comparison of RECIST and WHO response criteria." Jpn J Clin Oncol **33**(10): 533-7.
- Prasad, S. R., S. Saini, et al. (2003). "Radiological measurement of breast cancer metastases to lung and liver: comparison between WHO (bidimensional) and RECIST (unidimensional) guidelines." J Comput Assist Tomogr **27**(3): 380-4.
- Ravenel, J. G., W. M. Leue, et al. (2008). "Pulmonary nodule volume: effects of reconstruction parameters on automated measurements--a phantom study." Radiology **247**(2): 400-8.
- Revel, M. P., A. Bissery, et al. (2004). "Are two-dimensional CT measurements of small noncalcified pulmonary nodules reliable?" Radiology **231**(2): 453-8.
- Revel, M. P., C. Lefort, et al. (2004). "Pulmonary nodules: preliminary experience with three-dimensional evaluation." Radiology **231**(2): 459-66.
-

- Sohaib, S. A., B. Turner, et al. (2000). "CT assessment of tumour response to treatment: comparison of linear, cross-sectional and volumetric measures of tumour size." Br J Radiol **73**(875): 1178-84.
- Therasse, P., S. G. Arbuck, et al. (2000). "New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada." J Natl Cancer Inst **92**(3): 205-16.
- Therasse, P., E. A. Eisenhauer, et al. (2006). "RECIST revisited: a review of validation studies on tumour assessment." Eur J Cancer **42**(8): 1031-9.
- Therasse, P., A. Le Cesne, et al. (2005). "RECIST vs. WHO: prospective comparison of response criteria in an EORTC phase II clinical trial investigating ET-743 in advanced soft tissue sarcoma." Eur J Cancer **41**(10): 1426-30.
- Tran, L. N., M. S. Brown, et al. (2004). "Comparison of treatment response classifications between unidimensional, bidimensional, and volumetric measurements of metastatic lung lesions on chest computed tomography." Acad Radiol **11**(12): 1355-60.
- Uffmann, M. and M. Prokop (2001). "[Multislice CT of the lung. Technique and clinical applications]." Radiologe **41**(3): 240-7.
- Van Hoe, L., F. Haven, et al. (1997). "Factors influencing the accuracy of volume measurements in spiral CT: a phantom study." J Comput Assist Tomogr **21**(2): 332-8.
- Vogel, M. N., R. Vonthein, et al. (2008). "[Automated pulmonary nodule volumetry with an optimized algorithm. Accuracy at different slice thicknesses compared to unidimensional and bidimensional measurements]." Rofo **180**(9): 791-7.
- Wang, Y., R. J. van Klaveren, et al. (2008). "Effect of nodule characteristics on variability of semiautomated volume measurements in pulmonary nodules detected in a lung cancer screening program." Radiology **248**(2): 625-31.
- Wegener, O. H. (1996). Grundkurs Computertomographie. Berlin, Blackwell Wissenschafts-Verlag.
- Werner-Wasik, M., Y. Xiao, et al. (2001). "Assessment of lung cancer response after nonoperative therapy: tumor diameter, bidimensional product, and volume. A serial CT scan-based study." Int J Radiat Oncol Biol Phys **51**(1): 56-61.
- WHO (1979). WHO Handbook For Reporting Results For Cancer Treatment. Geneva, Switzerland.
-

- WHO (2008). World Cancer Report. Lyon.
- Wiemker, R., P. Rogalla, et al. (2005). "Aspects of computer-aided detection (CAD) and volumetry of pulmonary nodules using multislice CT." Br J Radiol **78 Spec No 1**: S46-56.
- Wormanns, D., G. Kohl, et al. (2004). "Volumetric measurements of pulmonary nodules at multi-row detector CT: in vivo reproducibility." Eur Radiol **14**(1): 86-92.
- Yankelevitz, D. F., A. P. Reeves, et al. (2000). "Small pulmonary nodules: volumetrically determined growth rates based on CT evaluation." Radiology **217**(1): 251-6.
- Zhao, B., L. H. Schwartz, et al. (2006). "Lung cancer: computerized quantification of tumor response--initial results." Radiology **241**(3): 892-8.

7 Abbildungsverzeichnis/Tabellenverzeichnis

- Abbildung 1:** Bidimensionale Messung eines Bronchialkarzinoms im Verlauf nach WHO-Kriterien. Der Vergleich der bidimensionalen Messergebnisse der Ausgangs-CT (1. Spalte) und der Verlaufs-CT (2. Spalte) ergab mit einer Reduktion der TumorgroÙe von 34.5% nach WHO-Kriterien SD (stable disease) (Hauth, Stattaus et al. 2007). **2**
- Abbildung 2:** Unidimensionale Messung des Bronchialkarzinoms im Verlauf nach RECIST. Der Vergleich des unidimensionalen Messergebnisses der Ausgangs-CT (1. Spalte) und der Verlaufs-CT (2. Spalte) ergab mit einer Reduktion der TumorgroÙe von 18% nach RECIST-Kriterien SD (stable disease) (Hauth, Stattaus et al. 2007). **6**
- Abbildung 3:** Prinzip der Datenerfassung bei der Computertomographie Projektionen einer Schicht werden aus vielen verschiedenen Richtungen senkrecht zur Körperachse durch ein Detektor-System registriert (Wegener 1996). **11**
- Abbildung 4:** Die Schichtebene besteht aus einer Matrix aus gleich großen Quadern (Voxel) mit einer quadratischen Grundfläche und einer definierten Schichtdicke (d_s) (Hofer 2006). **12**
- Abbildung 5:** Dichtwerte unterschiedlicher Gewebearten angegeben in Hounsfield-Einheiten (Hofer 2006). **13**
- Abbildung 6:** Bei der einzeiligen (SSCT = single slice CT) oder mehrzeiligen (MDCT = multi detector CT) Spiral-CT erfolgt die Abtastung kontinuierlich während die Liege gefahren wird (Hofer 2006). **14**
- Abbildung 7:** Verteilung der Primärtumore der Studienpatienten. **19**
- Abbildung 8:** Einteilung der Follow-Up-Untersuchungen ($n = 26$) mit manuell gemessenem Durchmesser in Responsekategorien nach RECIST. **28**
- Abbildung 9:** Einteilung der Follow-Up-Untersuchungen mit semiautomatisch gemessenen Volumeriewerten in Responsekategorien nach RECIST. **28**
- Abbildung 10:** Vergleich der Einteilung der manuellen Follow-Up-Messungen durch zwei unabhängige Untersucher (Reader1, Reader2) in Responsekategorien nach RECIST. **31**
- Abbildung 11:** Vergleich der Einteilung der semiautomatisch gemessenen Follow-Up-Volumina durch zwei unabhängige Untersucher (Reader1, Reader2) in Responsekategorien nach RECIST. **32**
- Abbildung 12:** Bland-Altman-Diagramm zur manuellen Messung nach RECIST durch zwei Untersucher (R1, R2). Differenz zwischen manuell ermitteltem Durchmesser von

R1 und R2 in Relation zur Summe des Mittelwertes der manuell ermittelten Durchmesser von R1 und R2 für jede Messung. **33**

Abbildung 13: Bland-Altman-Diagramm zur manuellen Messung nach RECIST durch zwei Untersucher (R1, R2) unter Ausschluss aller Läsionen kleiner als 10 mm. Differenz zwischen manuell ermitteltem Durchmesser von R1 und R2 in Relation zur Summe des Mittelwertes der manuell ermittelten Durchmesser von R1 und R2 für jede Messung. **34**

Abbildung 14: Bland-Altman-Diagramm zur manuellen Messung nach RECIST durch zwei Untersucher (R1, R2) unter Ausschluss aller Läsionen größer als 10 mm. Differenz zwischen manuell ermitteltem Durchmesser von R1 und R2 in Relation zur Summe des Mittelwertes der manuell ermittelten Durchmesser von R1 und R2 für jede Messung. **35**

Abbildung 15: Bland-Altman-Diagramm zur semiautomatischen Messung des größten Durchmessers der Läsion durch zwei Untersucher (R1, R2). Differenz zwischen semiautomatisch ermitteltem Durchmesser von R1 und R2 in Relation zur Summe des Mittelwertes der semiautomatisch ermittelten Durchmesser von R1 und R2 für jede Messung. **36**

Abbildung 16: Bland-Altman-Diagramm zur semiautomatischen Messung des größten Durchmessers der Läsion durch zwei Untersucher (R1, R2) unter Ausschluss aller Läsionen kleiner als 10 mm. Differenz zwischen semiautomatisch ermitteltem Durchmesser von R1 und R2 in Relation zur Summe des Mittelwertes der semiautomatisch ermittelten Durchmesser von R1 und R2 für jede Messung. **37**

Abbildung 17: Bland-Altman-Diagramm zur semiautomatischen Messung des größten Durchmessers der Läsion durch zwei Untersucher (R1, R2) unter Ausschluss aller Läsionen größer als 10 mm. Differenz zwischen semiautomatisch ermitteltem Durchmesser von R1 und R2 in Relation zur Summe des Mittelwertes der semiautomatisch ermittelten Durchmesser von R1 und R2 für jede Messung. **38**

Abbildung 18: Bland-Altman-Diagramm zum Vergleich der manuellen und semiautomatischen Messung des größten Durchmessers einer Läsion durch den ersten Untersucher (R1). Differenz zwischen manuell und semiautomatisch ermitteltem Durchmesser in Relation zur Summe des Mittelwertes des manuell und semiautomatisch ermittelten Durchmessers für jede Messung. **41**

Abbildung 19: Bland-Altman-Diagramm zum Vergleich der manuellen und semiautomatischen Messung des größten Durchmessers einer Läsion durch den ersten

Untersucher (R1) unter Ausschluss aller Läsionen kleiner als 10 mm. Differenz zwischen manuell und semiautomatisch ermitteltem Durchmesser in Relation zur Summe des Mittelwertes des manuell und semiautomatisch ermittelten Durchmessers für jede Messung. **42**

Abbildung 20: Bland-Altman-Diagramm zum Vergleich der manuellen und semiautomatischen Messung des größten Durchmessers einer Läsion durch den ersten Untersucher (R1) unter Ausschluss aller Läsionen größer als 10 mm. Differenz zwischen manuell und semiautomatisch ermitteltem Durchmesser in Relation zur Summe des Mittelwertes des manuell und semiautomatisch ermittelten Durchmessers für jede Messung. **43**

Abbildung 21: Bland-Altman-Diagramm zum Vergleich der manuellen und semiautomatischen Messung des größten Durchmessers einer Läsion durch den zweiten Untersucher (R2). Differenz zwischen manuell und semiautomatisch ermitteltem Durchmesser in Relation zur Summe des Mittelwertes des manuell und semiautomatisch ermittelten Durchmessers für jede Messung. **44**

Abbildung 22: Bland-Altman-Diagramm zum Vergleich der manuellen und semiautomatischen Messung des größten Durchmessers einer Läsion durch den zweiten Untersucher (R2) unter Ausschluss aller Läsionen kleiner als 10 mm. Differenz zwischen manuell und semiautomatisch ermitteltem Durchmesser in Relation zur Summe des Mittelwertes des manuell und semiautomatisch ermittelten Durchmessers für jede Messung. **45**

Abbildung 23: Bland-Altman-Diagramm zum Vergleich der manuellen und semiautomatischen Messung des größten Durchmessers einer Läsion durch den zweiten Untersucher (R2) unter Ausschluss aller Läsionen größer als 10 mm. Differenz zwischen manuell und semiautomatisch ermitteltem Durchmesser in Relation zur Summe des Mittelwertes des manuell und semiautomatisch ermittelten Durchmessers für jede Messung. **46**

| | |
|--|-----------|
| Tabelle 1: Responsekategorien nach WHO (1979). | 3 |
| Tabelle 2: Responsekategorien nach RECIST 1.0 (2000). | 6 |
| Tabelle 3: Beurteilung des Gesamtansprechens. | 7 |
| Tabelle 4: Vergleich alter und neuer RECIST-Kriterien. | 10 |
| Tabelle 5: Vergleich zwischen Volumenänderung und Änderung des Durchmessers nach RECIST. | 17 |
| Tabelle 6: Ergebnisse der manuellen Messung des max. Durchmessers der Lungenläsionen. | 25 |
| Tabelle 7: Ergebnisse der semiautomatischen Volumetrie. | 25 |
| Tabelle 8: Prozentuale Änderungen der Tumorgröße der einzelnen Follow-Up-Untersuchungen (t1 – t4) manuell gemessen nach RECIST-Kriterien und durch semiautomatische Volumetrie. | 27 |
| Tabelle 9: Follow-Up-Untersuchungen mit manuell gemessenem Durchmesser (n = 18) in der Responsekategorie SD nach RECIST aufgeschlüsselt in der semiautomatischen Volumetrie. | 29 |
| Tabelle 10: Follow-Up-Untersuchungen mit manuell gemessenem Durchmesser (n = 6) in der Responsekategorie PD nach RECIST aufgeschlüsselt in der semiautomatischen Volumetrie. | 30 |
| Tabelle 11: Follow-Up-Untersuchungen mit manuell gemessenem Durchmesser (n = 2) in der Responsekategorie PR nach RECIST aufgeschlüsselt in der semiautomatischen Volumetrie. | 30 |
| Tabelle 12: Übersicht der Ergebnisse der Messwerte bei gleicher Technik und Methode aber wechselnden Beobachtern. Aufgeführt sind die Ergebnisse des Vergleichs der manuellen Messung des größten Durchmessers der Läsionen von zwei Untersuchern (R1, R2) und der semiautomatischen Messung des größten Durchmessers der Läsionen von beiden Untersuchern. | 39 |
| Tabelle 13: Übersicht der Messwerte bei unterschiedlicher Technik und Methode und gleichen Beobachtern. Aufgeführt sind die Ergebnisse des Vergleichs zwischen manueller und semiautomatischer Messung des größten Durchmessers der Läsionen jeweils von Untersucher 1 und Untersucher 2. | 47 |
| Tabelle 14: Parameter der Zeitanalyse durch das Volumetrieprogramm CTSV-Prototype. Dargestellt sind der Mittelwert, die Standardabweichung, die schnellste und die längste Messdauer für alle Messungen (n = 430). | 48 |

8 Verzeichnis der akademischen Lehrer

Meine akademischen Lehrer in Marburg waren die Damen und Herren Professoren bzw. Privat-Dozenten in Marburg:

Aumüller, Barth, Basler, Bauer, Baum, Becker, Behr, Berger, Bien, Christiansen, Czubayko, Daut, Dietrich, Feuser, Görg, Gress, Grimm, Hertl, Herzum, Höffken, Hofbauer, Hoffmann, Hoyer, Kann, Kanngiesser, Klose, Kretschmer, Lill, Löffler, Lohoff, Maier, Maisch, Mandrek, Meyer, Moll, Mutters, Mueller, Müller, Oertel, Neubauer, Plant, Renz, Richter, Röhm, Rothmund, Ruchholtz, Schäfer, Schäfer, Schulze, Sesterhenn, Steininger, Stiletto, Vogelmeier, Vogt, Wagner, Weihe, Werner, Westermann, Wulf.

9 Danksagung

Die vorliegende Dissertation wurde in der Klinik für Strahlendiagnostik der Universitätskliniken Gießen und Marburg, Standort Marburg, Philipps-Universität Marburg unter der Leitung von Herrn Prof. Dr. Dr. Johannes Heverhagen angefertigt.

Herrn Prof. Dr. Dr. Johannes Heverhagen möchte ich für die Möglichkeit, die Dissertationsarbeit in der Klinik für Strahlendiagnostik durchführen zu können, die ausgezeichnete Betreuung, sowie für die kritische Durchsicht des Manuskripts, sehr herzlich danken.

Herrn Dr. Volker Dicken aus Bremen möchte ich für die Hilfe bei der Zusammenstellung der Ergebnisse, der statistischen Auswertung, und der Firma MeVis für die Bereitstellung der Volumetriesoftware danken. Herzlichen Dank auch den Medizinisch Technischen Röntgen Assistentinnen und Assistenten der Abteilung für die freundliche Aufnahme und stetige Hilfe.

Besonderer Dank gilt meinen verstorbenen Großeltern, meiner Großmutter, meinen Geschwistern, meinen Freunden und insbesondere meinem Freund Oliver, der sich immer Zeit genommen hat aufkommende Fragen zu diskutieren und mich zu motivieren.

Zuletzt, aber ganz besonders, möchte ich meinen Eltern danken, ohne deren fortwährende Unterstützung weder mein Studium noch diese Arbeit möglich gewesen wären und die mich stets in der Verwirklichung meiner Pläne und Wünsche mit Rat und Tat unterstützt haben.