

Aus dem  
Universitätsklinikum Münster  
Institut für Klinische Radiologie  
- Direktor: Univ.-Prof. Dr. med. Walter L. Heindel -

**Differentialdiagnostik pulmonaler Rundherde mit  
dynamischer CT: Wertigkeit eines volumenbasierten  
Messverfahrens**

INAUGURAL – DISSERTATION

zur

Erlangung des doctor medicinae

der Medizinischen Fakultät

der Westfälischen Wilhelms-Universität Münster

vorgelegt von  
Dregger, Uwe  
aus Arnsberg

2006

Gedruckt mit Genehmigung der Medizinischen Fakultät  
der Westfälischen Wilhelms-Universität Münster

Dekan: Univ. Prof. Dr. med. Heribert Jürgens

1. Berichterstatter: PD Dr. med. Dag Wormanns

2. Berichterstatter: Prof. Dr. med. Stefan Diederich

Tag der mündlichen Prüfung: 17.01.2006

Aus dem Universitätsklinikum Münster

Institut für Klinische Radiologie

- Direktor: Univ.-Prof. Dr. med. Walter L. Heindel -

Referent: PD Dr. med. Dag Wormanns

Koreferent: Prof. Dr. med. Stefan Diederich

### **Zusammenfassung**

*Differentialdiagnostik pulmonaler Rundherde mit dynamischer CT: Wertigkeit eines volumenbasierten Messverfahrens*

Uwe Dregger

Ziel der Arbeit war der Vergleich zweier Messverfahren des Kontrastmittelenhancements im dynamischen CT, das zur Differentialdiagnostik von Lungenrundherden eingesetzt wird. Dabei werden Aufnahmen vor und nach Kontrastmittelgabe gemacht und die Differenz der mittleren CT-Dichte des Herdes gebildet. Man geht davon aus, dass maligne Herde Kontrastmittel anreichern, so dass immer eine Dichteerhöhung nachweisbar ist. Reicht ein Herd nicht an, ist von Benignität auszugehen. Die Dichtemessung erfolgt bei einem Verfahren durch einen zweidimensionalen Schnitt, bei dem anderen Verfahren in einer 3D-Darstellung des ganzen Herdes.

Der Vergleich beider Methoden ergab keine Übereinstimmung ihrer Anreicherungsweite. Sie sind in der klinischen Anwendung nicht austauschbar. Bei den Testparametern zeigte das 3D-Verfahren bessere Werte für die Spezifität (35% ggü. 20 %) und ähnliche Resultate für den positiv prädiktiven Wert (48,0% zu 44,8%) und die Akkuranz (57,6 zu 51,5%). Im Gegensatz zum 2D-Verfahren erbrachte die 3D-Methode aber eine falsch negative Diagnose, was die Sensitivität (92,3% zu 100%) und den negativ prädiktiven Wert (87,5% zu 100%) verschlechterte. Eine ROC-Analyse zeigte einen Vorteil des 3D-Verfahrens (AUC 0,70 zu 0,52). Insgesamt wies die 3D-Methode in unseren Daten einen geringeren Messfehler auf als die 2D-Methode. Zum möglichst sicheren Ausschluss von Malignität sollten jedoch beide Verfahren kombiniert eingesetzt werden, gegebenenfalls unterstützt von Enhancement-Maps.

Tag der mündlichen Prüfung: 17.01.2006

Meinen Eltern gewidmet.

## **Inhaltsverzeichnis**

|  |           |
|--|-----------|
| <b>Inhaltsverzeichnis</b>  | <b>5</b>  |
| <br>   |           |
| <b>1. Einführung</b>   | <b>7</b>  |
| 1.1. <b>Notwendigkeit einer verbesserten Dignitätsbestimmung</b> | <b>7</b>  |
| 1.2. <b>Prinzipien der radiologischen Dignitätsbestimmung</b>    | <b>10</b> |
| 1.3. <b>Radiologische Untersuchungsverfahren</b>                 | <b>13</b> |
| 1.3.1. Positronen-Emissions-Tomographie (PET)                    | 13        |
| 1.3.2. Computertomographie (CT)                                  | 14        |
| 1.3.3. Magnetresonanztomographie (MRT)                           | 14        |
| 1.3.4. Weitere Verfahren   | 15        |
| 1.4. <b>Hintergrund dieser Arbeit</b>                            | <b>15</b> |
| <br>   |           |
| <b>2. Problemstellung und Zielsetzung</b>                        | <b>17</b> |
| <br>   |           |
| <b>3. Material und Methoden</b>                                  | <b>20</b> |
| 3.1. <b>Patientenstichprobe</b>                                  | <b>20</b> |
| 3.2. <b>Untersuchungsprotokoll</b>                               | <b>24</b> |
| 3.3. <b>Auswertung der Daten</b>                                 | <b>26</b> |
| 3.3.1. Auswertung nach Swensen                                   | 27        |
| 3.3.2. Auswertung nach dem 3D-Verfahren                          | 29        |
| 3.3.3. Statistische Auswertung der Daten                         | 31        |
| 3.3.3.1. <i>Übereinstimmung (Agreement) beider Messmethoden</i>  | 31        |
| 3.3.3.2. <i>Testgütekriterien beider Verfahren</i>               | 34        |
| 3.3.3.3. <i>Statistische Auswertung der Doppeluntersuchungen</i> | 36        |

|  |           |
|--|-----------|
| <b>4. Ergebnisse</b>                                 | <b>39</b> |
| 4.1. Übereinstimmung (Agreement) beider Methoden     | 39        |
| 4.2. Testgüteparameter                               | 41        |
| 4.3. Kontrastmitteldynamik                           | 45        |
| 4.4. Doppelmessungen                                 | 48        |
| <br>   |           |
| <b>5. Diskussion</b>                                 | <b>53</b> |
| 5.1. Diskussion der Ergebnisse                       | 53        |
| 5.1.1. Übereinstimmung (Agreement) beider Verfahren  | 53        |
| 5.1.2. Ergebnisse der Doppelmessungen                | 55        |
| 5.1.3. Testgüteparameter                             | 56        |
| 5.1.3.1. <i>Sensitivität</i>                         | 56        |
| 5.1.3.2. <i>Spezifität</i>                           | 56        |
| 5.2. Probleme bei der Datenauswertung                | 58        |
| 5.2.1. Problemanalyse                                | 58        |
| 5.2.2. Lösungsansätze                                | 62        |
| 5.3. Vergleich mit bisherigen Publikationen          | 64        |
| 5.3.1. Publikationen von Swensen et al.              | 64        |
| 5.3.2. Publikation von Ohno und Schäfer et al.       | 67        |
| 5.4. Klinische Anwendungsmöglichkeiten der Verfahren | 69        |
| 5.5. Fazit   | 71        |
| <br>   |           |
| <b>6. Quellenverzeichnis</b>                         | <b>73</b> |
| <br>   |           |
| <b>7. Danksagung</b>                                 | <b>90</b> |
| <br>   |           |
| <b>8. Lebenslauf</b>                                 | <b>91</b> |
| <br>   |           |
| <b>9. Anhang</b>                                     | <b>I</b>  |

## 1. Einführung

### 1.1. Notwendigkeit einer verbesserten Dignitätsbestimmung

Im Jahr 2002 standen nach den Herz-Kreislauf-Erkrankungen mit 46,8% die Krebserkrankungen mit 25,0% auf dem zweiten Platz der Häufigkeit der Todesursachen in Deutschland (87). Eine besondere Bedeutung unter den bösartigen Neubildungen nehmen die malignen Tumoren der Lunge ein. Im Jahr 2002 verstarben daran in Deutschland 39105 Menschen, darunter 75% Männer und 25% Frauen. Das Bronchialkarzinom war im Jahre 2002 somit die fünfthäufigste Todesursache in Deutschland insgesamt. Bei den Frauen stand es bei den häufigsten Todesursachen auf Platz 10, bei den Männern auf Platz 3 und war bei ihnen dabei der vierthäufigste Grund für einen stationären Krankenhausaufenthalt (88).

Während bei Männern die Sterberate an Lungenkrebs seit 1988 stagniert bzw. sogar leicht rückläufig ist, ist bei den Frauen ein rascher Anstieg zu verzeichnen (59). Setzt sich diese Entwicklung fort, so könnte bei den Frauen in Deutschland das Bronchialkarzinom bald das Colonkarzinom als zweithäufigste krebsbedingte Todesursache nach dem Brustkrebs ablösen.

Doch nicht nur in der Bundesrepublik sind die malignen Lungentumoren ein bedeutender Faktor in der Mortalitätsstatistik. Auch in den USA ist er mit 169500 Neuerkrankungen die zweithäufigste diagnostizierte Krebserkrankung und führte 2001 mit 157400 Todesfällen die Statistik der Krebstodesursachen an (1). Es wurde geschätzt, dass weltweit im Jahr 2000 etwa 1,3 Millionen Menschen am Bronchialkarzinom verstorben sind (72).

Die Ursache für diese Entwicklung wird vornehmlich im Hauptrisikofaktor, nämlich dem Tabakrauchen gesucht (37). Das Risiko eines Rauchers ein Bronchialkarzinom zu entwickeln ist bei männlichen Rauchern 22-mal, und bei Frauen 12-mal so hoch wie bei Nichtrauchern (81). Insgesamt geht man davon aus, dass 85% aller malignen Lungentumoren auf Tabakrauch zurückzuführen sind (82).



Der Tabakkonsum ist in der Bevölkerung weiterhin sehr verbreitet. Der zu versteuernde Verkaufswert von Zigaretten und anderen Tabakprodukten lag im Jahr 2004 bei 22,9 Mrd. Euro. Nach wiederholten Erhöhungen der Tabaksteuer bedeutete dies zwar einen Rückgang von 1,2 % gegenüber dem Vorjahr (89), die bösartigen Tumoren der Lunge werden aber auch in Zukunft eine große Bedeutung in der Medizin haben.

Die Prognose eines malignen Lungentumors hängt in hohem Maße vom Stadium ab, in dem die Diagnose gestellt wurde. Während die mittlere 5-Jahres-Überlebensrate bei allen Stadien nur 14 % beträgt, konnte gezeigt werden, dass bei einer Diagnose im Stadium 1, in dem noch keine Metastasierung erfolgt ist und somit noch ein lokalisierter Prozess vorliegt, 70% der Patienten kurativ therapiert und geheilt werden können (23, 45, 80).

Viele Tumoren sind Zufallsbefunde und werden auf routinemäßigen Röntgenaufnahmen des Thorax entdeckt, oft schon, wenn der Patient noch gar keine Beschwerden hat. Die radiologischen Bilder, die sich dabei bieten können sind mannigfaltig, denn hinter vielen Anomalien einer Thoraxaufnahme (z.B. Atelektasen, Infiltrate) kann ein Bronchialkarzinom als Ursache stecken.

Die bei weitem häufigste Manifestation ist allerdings der pulmonale Rundherd (Solitary pulmonary nodule = SPN), also eine umschriebene Dichteerhöhung im Lungengewebe. Wenn man einen solchen Herd entdeckt, stellt sich das Problem, wie man ihn bestmöglich als gut- oder bösartig klassifizieren kann, da auch benigne Prozesse das Bild eines pulmonalen Rundherdes bieten können. Insgesamt stellten sich etwa 50% der Herde bei in den USA durchgeführten Biopsien und Thorakotomien als benigne heraus (6, 49, 64).

Die an die Medizin gestellte Anforderung ist also, Verfahren zu entwickeln, mit denen pulmonale Rundherde möglichst früh erkannt und identifiziert werden können. Dabei sollte das Untersuchungsverfahren den Patienten möglichst wenig belasten sowie bestmöglich zwischen benignen und malignen Prozessen

unterscheiden können, da von dieser Differenzierung das weitere therapeutische Vorgehen abhängig gemacht wird.

Wird bei einem Patienten auf einer Röntgenaufnahme des Thorax ein verdächtiger Rundherd entdeckt, wird zunächst versucht mittels einer Computertomographie den Prozess genauer darzustellen, die genaue Lokalisation zu bestimmen, sowie Hinweise für Malignität (z.B. Einwachsen in umgebende Strukturen, typische Morphologie im Sinne von unregelmäßigem Wachstum, Nekrosen) oder Benignität (z.B. Verkalkungen) zu finden.

Weiterhin verdächtige Herde versucht man bronchoskopisch oder perkutan zu biopsieren, um mit der histopathologischen Untersuchung eine definitive Dignitätsbestimmung durchführen zu können.

Insbesondere kleine Herde, deren frühe Diagnose dem Patienten eine besonders gute Chance auf Heilung eröffnen würde, sind leider sehr schwierig zu biopsieren. Ein weiterer Grund, warum auch für größere Herde eine Biopsie unmöglich werden kann, ist eine schwierige Lokalisation, z.B. im Bereich des Lungenhilus in Nachbarschaft zu den großen Lungengefäßen, die bei der Untersuchung verletzt werden könnten. Eine weitere und relativ häufige Komplikation bei der Biopsie ist ein Pneumothorax, der in einem kleinen Teil der Fälle mittels Thoraxdrainage therapiebedürftig ist und somit zu einer durch die Biopsie bedingten Morbidität beiträgt (4, 8, 9, 20, 39, 46, 60, 61, 67, 102).

Ist eine Biopsie aufgrund der Lokalisation oder Größe des Herdes nicht möglich, verbleibt die Thorakoskopie oder eine Thorakotomie, mit dem Ziel, den verdächtigen Herd vollständig zu resezieren und einer histopathologischen Untersuchung zuzuführen (6, 49, 64).

Nicht nur, dass so eine Operation für den Patienten mit ungleich höheren Belastungen, Schmerzen und Risiken verbunden ist als z.B. eine radiologische Untersuchung, darüber hinaus sind auch die entstehenden Kosten für das Gesundheitssystem eminent hoch. So gibt St. Anthony's DRG guidebook die Kosten für den Klinikaufenthalt eines Patienten und der chirurgischen

Entfernung eines verdächtigen pulmonalen Rundherdes in den USA mit 18.000 bis 30.000 US\$ an (86).

In der aktuellen Literatur wird allerdings der Grundsatz vertreten, dass ein entdeckter pulmonaler Rundherd bis zum Beweis des Gegenteiles als maligne zu gelten hat, mit der Konsequenz, dass ohne weitere diagnostische Maßnahmen direkt biopsiert oder reseziert werden sollte (99).

Da sich aber etwa 50% der resezierten Herde in der Histologie als benigne erweisen (6, 49, 64, 84), wäre es sehr wünschenswert, diese vorher mit hinreichender Sicherheit erkennen zu können, um eine solche Operation zu vermeiden.

Aus diesen Gründen sind in den letzten Jahren Anstrengungen unternommen worden, um mit modifizierten radiologischen Verfahren eine bessere Unterscheidung zwischen benignen und malignen Prozessen zu ermöglichen, wobei diese Unterscheidung schon in möglichst frühen Stadien durchführbar sein sollte.

## **1.2. Prinzipien der radiologischen Dignitätsbestimmung**

Im Folgenden soll kurz auf die Ansätze eingegangen werden, mit denen versucht wurde, die radiologische Diagnostik bei der Dignitätsbestimmung pulmonaler Rundherde zu verbessern.

Es hat sich gezeigt, dass bisher kein radiologisches Verfahren alleine in der Lage ist, hinreichende Sicherheit über die Dignität eines Prozesses zu liefern. Meist werden die Verfahren in Kombination eingesetzt, um klinisch verwertbare Aussagen machen zu können.

Im Wesentlichen gibt es drei Prinzipien, die bei der Dignitätsbestimmung relevant sind und bei unterschiedlichen radiologischen Verfahren eingesetzt werden. Sie wurden nicht nur an Tumoren der Lunge, sondern auch an benignen und malignen Prozessen anderer Organe untersucht.

Das erste Prinzip basiert auf den metabolischen Unterschieden von malignen und benignen Tumoren, die insbesondere im Energiestoffwechsel bestehen. Aufgrund des wesentlich schnelleren Wachstums von bösartigen Geweben, sind diese auf eine vermehrte Energieerzeugung angewiesen und zwar viel stärker als langsamer wachsende gutartige Prozesse. Zudem entsteht durch die grenzwertige Blutversorgung im malignen Tumor eine anaerobe Stoffwechsellage, die wegen der schlechteren Energieausbeute zu einem zusätzlichen Glucoseverbrauch führt.

Der durch diese Mechanismen intensivierte Glucosemetabolismus wird durch einen vermehrten Glucosetransport in die Tumorzellen gewährleistet. (24).

Eine Differenzierung zwischen benignen und malignen Tumoren kann also mittels einer Darstellung der Intensität des Glucosemetabolismus versucht werden.

Neben diesem Ansatz wird bei einem weiteren Prinzip ein morphologischer Unterschied zur Dignitätsbestimmung eines Tumors genutzt.

Das schnelle Wachstum maligner Tumoren erfordert zur Ermöglichung eines intensivierten Energiestoffwechsels auch eine Steigerung der Zufuhr von Sauerstoff und Nährstoffen. Dieses wird durch eine eigene Gefäßversorgung des Tumors ermöglicht. Diese sogenannte Neoangiogenese wird durch vom Tumor sezernierte Substanzen (z.B. Tumorangiogenesefaktor, TAF) stimuliert und gewährleistet eine vermehrte Perfusion des Tumorgewebes (12). Die Gefäßneubildung wird zur Sicherstellung des weiteren Wachstums ab einem Tumordurchmesser von etwa 1 mm notwendig. Dies entspricht etwa einer Zellformation von 1 Millionen Zellen und ist gleichzeitig die Nachweisgrenze derartiger Läsionen in der CT.

Bei der Dignitätsbestimmung wird nun der Umstand genutzt, dass die erhöhte Tumorperfusion als Anreicherung von Kontrastmittel (= Enhancement) radiologisch nachweisbar ist (50, 62).

Die Menge an Kontrastmittel, die in einem Tumor gemessen werden kann, ergibt sich als Funktion der Kontrastmittelmenge, die durch die Tumorgefäße

strömt (Perfusion), der Menge, die von dort in das Tumorgewebe übertritt (Diffusion) sowie aus der Größe des interstitiellen Verteilungsraumes (13, 62).

Daraus folgt, dass ein maligner Tumor mehr Kontrastmittel anreichert als viele der benignen Herde. Benigne Läsionen wie Hamartome oder fibrotische, vernarbte Strukturen weisen eine geringere Mikrogefäßdichte auf, manche Tuberkulome sind zentral komplett nekrotisch, so dass hier im Gegensatz zu malignen Läsionen kein wesentliches Enhancement zu erwarten ist (52).

Zum Nachweis der Anreicherung werden computer- oder kernspintomographisch Aufnahmen vor und nach Kontrastmittelgabe angefertigt und anschließend die Dichte (CT) bzw. Signalintensität (MRT) des zu untersuchenden Herdes gemessen. Eine Kontrastmittelanreicherung würde einen nachweisbaren Dichte- bzw. Signalintensitätsanstieg verursachen und so einen Hinweis auf Malignität geben.

Allerdings gibt es auch eine Reihe von gutartigen Prozessen, die ein ausgeprägtes Enhancement zeigen können. So besteht beispielsweise bei floriden Entzündungen ebenfalls eine vermehrte Gewebsperfusion (10), die durch von Entzündungsmediatoren induzierter Vasodilatation und Permeabilitätserhöhung verursacht wird. Eine Anreicherung von Kontrastmittel ist demnach keinesfalls spezifisch für Malignität. Umgekehrt allerdings würde eine fehlende Anreicherung einen bösartigen Tumor ausschließen, da alle malignen Prozesse essentiell mit erhöhter Perfusion und damit Kontrastmittel-Enhancement einhergehen.

Dieses Prinzip ist also nutzbar, um durch einen Ausschluss von Kontrastmittelanreicherung einen bösartigen Prozess sehr unwahrscheinlich zu machen und statt einer Biopsie oder sogar einer Resektion weitere Verlaufskontrollen durchzuführen.

Das dritte Prinzip schließlich soll direkt das schnellere Wachstum maligner im Gegensatz zu benignen Tumoren nachweisen.

Dazu bedient man sich der computergestützten Bestimmung des Tumolvolumens. Bei dieser sogenannten Volumetrie wird der Tumor im CT

dreidimensional dargestellt, so dass entsprechende Programme das Volumen berechnen können. Eine Aussage über die Wachstumsgeschwindigkeit erhält man, wenn man wiederholt in bestimmten Zeitabständen das Volumen und daraus die Tumorverdopplungszeit berechnet, also die Zeit, die ein Tumor braucht, um sein Volumen zu verdoppeln. Diese Größe kann zur Bestimmung der Dignität genutzt werden.

### **1.3. Radiologische Untersuchungsverfahren**

Im Folgenden soll kurz auf die radiologischen Untersuchungsmethoden eingegangen werden, die sich den beschriebenen Prinzipien bedienen.

#### **1.3.1. Positronen-Emissions-Tomographie (PET)**

Die Positronen-Emissions-Tomographie (14, 15, 30, 32-35, 54, 63, 76, 78), ein nuklearmedizinisches Verfahren, ist in der Lage, einen erhöhten Glucose-Stoffwechsel zu detektieren. Dazu wird den Patienten markierte Glucose ([<sup>18</sup>F] 2-Deoxy-2-Fluoro-D-Glucose) verabreicht, die wie natürliche Glucose metabolisiert wird, die im Unterschied dazu allerdings mittels PET im Körper lokalisiert werden kann. Besonders intensive Anreicherung erfolgt an Stellen, an denen ein intensiver Glucosestoffwechsel besteht, so zum Beispiel in malignen Tumoren.

Die PET bietet die höchste Spezifität aller radiologischen Verfahren. Problematisch ist allerdings, dass bei kleinen Herden, die einen Durchmesser von weniger als 1 cm haben, bei Karzinoiden, bei bronchoalveolären Karzinomen sowie bei hochdifferenzierten Adenokarzinomen die Sensitivität eingeschränkt ist.

Eine Metaanalyse von Reske et al. (78) ermittelte für die PET einen negativ prädiktiven Wert von 90%. Das bedeutet, dass bei 10 % aller negativ getesteten Patienten dennoch ein maligner Tumor vorhanden ist, was die klinische Anwendung limitiert.

### 1.3.2. Computertomographie (CT)

Die Computertomographie ist das Verfahren, das die höchste Ortsauflösung bei der Darstellung eines Lungenherdes bietet (29, 31, 55-57, 71, 77, 83, 91-94, 97, 98, 106, 108, 109). Mittels CT kann die Kontrastmittelanreicherung von Herden untersucht werden. Dazu wird ein Röntgenkontrastmittel verabreicht und vor und nach Injektion die Dichte des zu untersuchenden Herdes bestimmt. Weiterhin gibt es die Möglichkeit der Volumetrie. Mit speziellen Programmen können anhand der CT-Daten Durchmesser- und Volumenbestimmungen von Tumoren durchgeführt werden.

Im Gegensatz zur Magnetresonanztomographie (siehe Kapitel 1.3.3.) ist für den Patienten bei der Computertomographie eine Strahlenbelastung vorhanden. Die Vorteile liegen allerdings darin, dass eine bessere Kontrastierung zwischen lufthaltigem Lungengewebe und einem soliden Herd (beispielsweise einem Tumor) möglich ist und dadurch viele Herde besser dargestellt werden können. Weiterhin ist eine schnellere und kostengünstigere Untersuchung möglich. Aus diesen Gründen ist die Computertomographie das Verfahren, was in diesem Zusammenhang am häufigsten eingesetzt wird.

### 1.3.3. Magnetresonanztomographie (MRT)

Das dritte relevante Verfahren in diesem Zusammenhang ist die Magnetresonanztomographie (3, 13, 18, 28, 36, 40-44, 48, 51, 58, 74, 79, 85). Sie bietet wie die Computertomographie die Möglichkeit der Dichtemessung vor und nach Gabe eines paramagnetischen Kontrastmittels (z.B. Gadolinium-DTPA) mit Bestimmung der Anreicherung. Studien wurden unter anderem von Ohno et al. (74) durchgeführt. Sie kamen zu dem Ergebnis, dass dieses Verfahren zwar gute Ergebnisse hinsichtlich der Sensitivität zeigte (100% bei Ohno bei einer relativ kleinen Fallzahl), jedoch die Spezifität eingeschränkt war (70% bei Ohno).

Ohno untersuchte neben dem Enhancement noch zusätzlich die Steilheit des Kontrastmittelanstieges im Tumor, den sogenannten Slope. Ob dieser

allerdings als eigenständiger Parameter bei der Dignitätsbestimmung dienen kann ist unklar, da das Ausmaß der Anreicherung mit der Steilheit des Anstieges unmittelbar korreliert, wenn die Zeit bis zum Erreichen des Dichtemaximums vergleichbar ist.

Gückel und Schäfer et al. (36, 79) untersuchten charakteristische Kurvenverläufe des Enhancements und konnte dabei vier Verläufe unterscheiden, die sich mit einer bestimmten Wahrscheinlichkeit einem Herdtyp zuordnen ließen.

#### 1.3.4. Weitere Verfahren

In der Literatur werden noch weitere radiologische differentialdiagnostische Verfahren besprochen. Dazu gehören z.B. die Angiographie (68, 69, 70, 73, 100) und die Dopplersonographie (47, 107). Da deren Aussagekraft sich aber in dieser Fragestellung als begrenzt erwiesen hat, soll hier nicht näher darauf eingegangen werden.

### 1.4. Hintergrund dieser Arbeit

Diese Arbeit soll das Prinzip der computertomographischen Differentialdiagnostik durch den Nachweis oder Ausschluss von Kontrastmittel-Enhancement in der dynamischen CT näher untersuchen.

Entscheidende Arbeiten zu diesem Thema wurden von Swensen (91-98) vorgelegt. Er fand heraus, dass das Fehlen einer signifikanten Anreicherung ein deutlicher Hinweis für die Benignität eines Prozesses ist. Umgekehrt ist jede nicht durch Messfehler erklärbare Anreicherung bis zum Beweis des Gegenteils als malignitätsverdächtig einzustufen.

Die Problematik bei diesem Verfahren liegt darin, dass Messfehler einen großen Einfluss haben und noch nicht geklärt ist, ab welcher Dichteerhöhung



nach Kontrastmittelgabe nicht mehr von einem Messfehler ausgegangen werden kann, sondern von einem signifikanten Effekt, der durch Kontrastmittelaufnahme des Tumors verursacht wird.

Swensen ermittelte retrospektiv einen empirischen Schwellenwert von 15 Hounsfield-Einheiten (HU) und erzielte damit eine Sensitivität von 98%, eine Akkuranz von 77% und eine Spezifität von 58%.

Zur Reduktion des Messfehlers wurde das von Swensen vorgeschlagene Verfahren modifiziert. Anstatt die mittlere Dichte nur eines Schnittbildes durch den Tumor zu bestimmen, vermochte ein Segmentierungsverfahren, welches von Fan et al. (19) beschrieben wurde, eine dreidimensionale Darstellung des Tumors und somit eine Messung der mittleren Dichte des kompletten Herdes.

Die Unterschiede beider Verfahren hinsichtlich des Messfehlers wurden von Wormanns et al. (103) beschrieben. Es zeigte sich, dass die Dichtebestimmung mit dem 3D-Verfahren wesentlich genauer erfolgen konnte.

Als Hauptgründe dafür werden angeführt, dass bei Erfassung des ganzen Herdes der Einfluss unterschiedlicher Atemzugtiefen auf die Schichthöhe wegfällt, keine Inhomogenitätsdifferenzen auftreten, sowie aufgrund des größeren Messvolumens eine bessere Mittelung der Dichtewerte und somit ein besserer Fehlerausgleich erzielt werden kann, als bei dem von Swensen angewendeten 2D-Verfahren.

Insgesamt ist zu erwarten, dass der Messfehler signifikant reduziert werden kann.

Weitere Studien über dreidimensionale Darstellung von Lungenherden und deren Segmentierung sind in der Literatur zu finden (2, 11, 53, 101, 110, 111).

Mit dem Vergleich von Swensen's 2D-Verfahren und dem neuen 3D-Ansatz beschäftigt sich diese Arbeit.

## **2. Problemstellung und Zielsetzung**

Ziel dieser Arbeit ist ein Vergleich des von Swensen (91-98) angewandten 2D-Verfahrens mit dem neuen 3D-Ansatz.

In einem ersten Schritt soll die Frage geklärt werden, inwiefern die Ergebnisse, die beide Verfahren liefern, übereinstimmen. Daraus kann dann beantwortet werden, ob sie in der klinischen Anwendung austauschbar sind, sofern sie annähernd zu demselben Ergebnissen gelangen. Sind die Messunterschiede, die bei einem Vergleich beider Methoden festgestellt werden zu groß, so können sie nicht nebeneinander eingesetzt werden, und es muss geklärt werden, welches Verfahren sich besser für den klinischen Einsatz eignet.

Um dies zu beantworten werden Testgütekriterien berechnet. Anhand der bereits histologisch oder über den langfristigen Verlauf sicher als benigne oder maligne diagnostizierten Patienten, kann für beide Methoden festgestellt werden, wie viele maligne Herde mit dem jeweiligen Verfahren als solche erkannt werden können (Sensitivität), wie viele benigne Herde richtig diagnostiziert werden können (Spezifität) und somit wie viele Patienten insgesamt bezüglich der Dignität ihrer Rundherde richtig erkannt werden (Akkuranz).

Darüber hinaus ist eine Aussage darüber möglich, wie viele der als maligne getesteten Herde (also Herde, die über einen bestimmten Schwellenwert hinaus in ihrer Dichte nach Kontrastmittelgabe zunehmen) tatsächlich auch bösartig sind (positiver prädiktiver Wert) und wie viele von den als gutartig getesteten (Herde ohne signifikante Anreicherung) gutartig sind (negativer prädiktiver Wert).

Diese Indizes sollen für beide Verfahren ermittelt und miteinander verglichen werden.

Die Entscheidung, ob ein Herd als maligne oder benigne eingestuft wird, hängt vom Ausmaß der gemessenen Dichteerhöhung ab. Da hierbei der Messfehler des Verfahrens eine große Rolle spielt, stellt sich die Frage, ab welcher Dichteerhöhung (Schwellenwert) tatsächlich von einer Kontrastmittelanreicherung und somit von Malignitätsverdacht auszugehen ist.

Einmal soll das Ausmaß des Messfehlers bei beiden Verfahren abgeschätzt werden, indem Doppelmessungen ohne Kontrastmittel an jeweils einem Herd durchgeführt werden, anhand derer man Aussagen über die Reproduzierbarkeit der beiden Methoden machen kann. Die Kenntnis der Reproduzierbarkeit und somit des Messfehlers ist für die Festlegung des Schwellenwertes wichtig.

Swensen hatte in seiner Arbeit empirisch einen Schwellenwert von 15 Hounsfield-Einheiten ermittelt. Auch dieser Schwellenwert soll im Folgenden untersucht werden.

Neben der Beurteilung des empirischen Wertes von Swensen hinsichtlich Sensitivität, Spezifität, Akkuranz und Prädiktion sollen alternative Schwellenwerte darauf untersucht werden, ob mit ihnen bessere Ergebnisse erzielt werden können. Dazu wird eine ROC-Analyse vorgenommen, die für jeden Schwellenwert einen entsprechenden Wert für die Sensitivität und Spezifität zuordnet und so für beide Verfahren einen Überblick über die mit jedem Schwellenwert zu erzielenden Resultate gestattet.

Die ROC-Analysen können dann genutzt werden, um zu beurteilen, welches der beiden Verfahren bei allen Schwellenwerten die besseren Testgütekriterien liefert.

Weiterhin wird die Fragestellung untersucht, ob die besten Resultate bei den Testgütekriterien durch Nutzung der jeweils maximalen Anreicherungen erzielt wird, wie von Swensen vorgeschlagen, oder ob eine weitere Verbesserung möglich ist, wenn man stattdessen die durchschnittliche Anreicherung nach Kontrastmitteldosis oder sogar nur jeweils einzelne Messungen für die Beurteilung zu Grunde legt.

Zusätzlich wird eine Analyse des Dichteanstieges durchgeführt. Es soll geklärt werden, ob ein Zusammenhang zwischen dem Verlauf des Dichteanstieges und der Dignität des Tumors zu finden ist.

Dazu wird ermittelt, bei welcher Messung nach Kontrastmittelgabe die jeweiligen Herde ihr Dichtemaximum erreichen und ob daraus ein weiterer Ansatz herleitbar ist, wie man benigne und maligne Herde differenzieren kann. Für diese Analyse werden die mit dem 2D-Verfahren ausgewerteten Daten herangezogen.

Abschließend soll anhand der gefundenen Daten beurteilt werden, welche der beiden untersuchten Methoden sich besser für die klinische Beurteilung der Dignität von pulmonalen Rundherden eignet und ob die Ergebnisse, die das bessere von beiden Verfahren liefert ausreichend genau sind, um unter bestimmten Umständen den Einsatz invasiver diagnostischer Verfahren zu vermeiden.

### **3. Material und Methoden**

In diesem Kapitel wird besprochen, an welchem Patientenkollektiv die Untersuchungen durchgeführt wurden, wie diese erfolgten und auf welche Weise die gewonnenen Daten ausgewertet wurden.

#### **3.1. Patientenstichprobe**

Die Untersuchungen wurden an einer Stichprobe von 47 Patienten durchgeführt, bei denen unklare pulmonale Herde entdeckt wurden. 43 Patienten hatten solitäre Herde, während 4 Patienten 2 Herde aufwiesen. Insgesamt konnten so 51 Herde in die Auswertung mit einbezogen werden. Bei einem Patienten konnte aufgrund der geringen Größe seines Herdes keine Auswertung mit dem 3D-Verfahren erfolgen, so dass für diesen Herd weder Messwerte, die das Herdvolumen betreffen, noch 3D-Dichtewerte ermittelt werden konnten. Dieser Herd wurde für die Analyse genutzt, die untersuchte, bei welcher Messung die Herde ihr maximales Enhancement erreichen, und die ausschließlich auf Daten beruhte, die mit dem 2D-Verfahren ausgewertet wurden. Für alle Untersuchungen, die mit dem 3D-Verfahren durchgeführt wurden, blieben somit 50 Herde übrig.

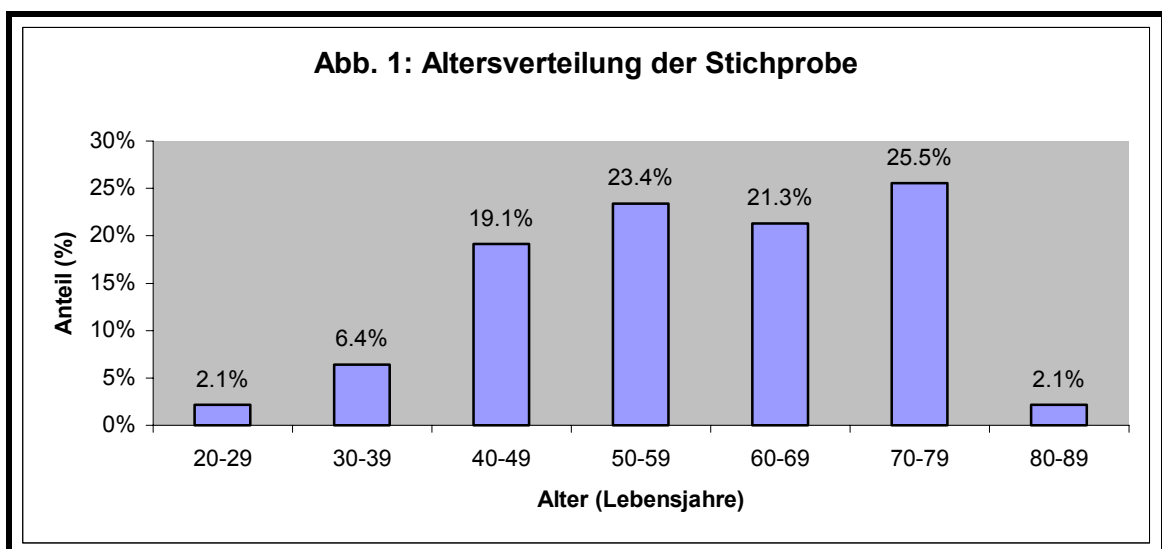


Abb. 1: Altersverteilung der Stichprobe (n=47 Patienten)

Unter den 47 Patienten befinden sich 20 Frauen (42,6%) und 27 Männer (57,4%). Der jüngste Patient war 23 Jahre, der älteste 80 Jahre alt. Der Altersdurchschnitt lag bei 58,6 Jahren. Die Altersverteilung der Stichprobe kann Abb. 1 entnommen werden.

Unter den untersuchten Herden hatte der kleinste einen Durchmesser von 3,1 mm (Minimum) bis 4,8 mm (Maximum) bei einem Volumen von 34 mm<sup>3</sup>, der größte einen Durchmesser von 19,3 mm (Minimum) bis 26,8 mm (Maximum) bei einem Volumen von 2955 mm<sup>3</sup>. Der mittlere Durchmesser aller untersuchten Prozesse liegt bei 10 mm, das mittlere Volumen beträgt 553 mm<sup>3</sup>. Durchmesser und Volumen der Herde wurden mit einer Prototypenversion des unten näher beschriebenen Programms syngo LungCare CT™ (Siemens Medical Solutions, Forchheim, Deutschland) auf einer Wizard™-Workstation bestimmt.

Die Größenverteilung der untersuchten Knoten wird in Abb. 2 dargestellt.

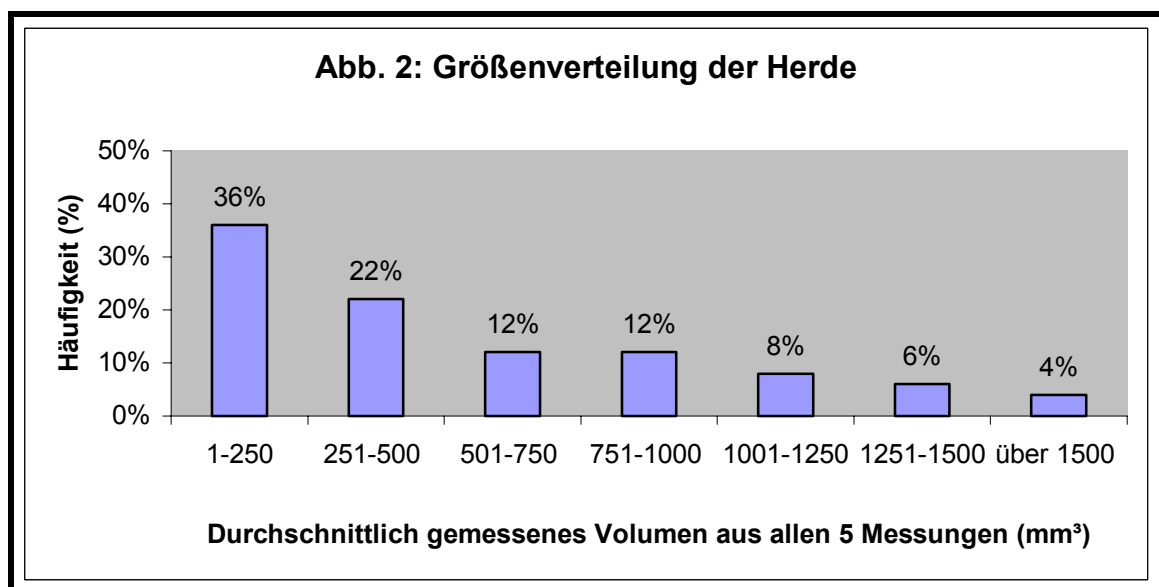


Abb. 2: Übersicht über die Größenverteilung der Herde (n=50 Herde)

Von den untersuchten Herden konnte bei 13 die Diagnose eines malignen Tumors gestellt werden, weitere 20 wurden als benigne Läsionen identifiziert. Somit konnte bei insgesamt 33 Läsionen (66%) die Dignität gesichert werden. 17 der 50 Herde galten als noch nicht definitiv zugeordnet.

Folgende Kriterien wurden angewendet, um einen Prozess als sicher maligne oder benigne zu diagnostizieren: Die Diagnose eines bösartigen Tumors wurde jeweils nach einer histopathologischen Untersuchung einer Probe gestellt, die vom Patienten bei einer transthorakalen oder bronchoskopischen Biopsie oder einer Thorakoskopie bzw. Thorakotomie mit Resektion des verdächtigen Herdes entnommen wurde.

Herde wurden als benigne eingestuft, wenn diese Diagnose ebenfalls durch eine histopathologische Untersuchung eines Bioplates oder Resektates gestellt werden konnte, oder wenn in einem Beobachtungszeitraum von mindestens 2 Jahren auf computertomographischen Kontrollaufnahmen kein Wachstum des Prozesses oder sogar eine Regression desselben nachgewiesen wurde.

Bei einer 43jährigen Patientin wurde in diesem Punkt eine Ausnahme gemacht. Sie hatte einen unklaren solitären pulmonalen Rundherd. Aufgrund eines daneben bestehenden malignen Melanoms entstanden im Beobachtungszeitraum abdominelle sowie multiple pulmonale Metastasen, die eine hohe Wachstumsgeschwindigkeit aufwiesen. Im Gegensatz dazu veränderte sich die Größe des zunächst entdeckten Herdes nicht. Aus diesem Grunde konnte dabei nicht von einer Metastase des Melanoms ausgegangen werden, sondern stattdessen von einem Granulom oder einer ähnlichen benignen Läsion. Die Wahrscheinlichkeit eines neben dem Melanom noch bestehenden primären malignen Lungenprozesses wurde aufgrund des jungen Alters der Patientin als sehr gering eingestuft. Obwohl also formal ein Beobachtungszeitraum von 2 Jahren nicht eingehalten werden konnte, da die Patientin in der Zwischenzeit an dem Melanom verstorben war, wurde anhand des Verlaufes die Diagnose eines gutartigen Herdes gestellt.

Alle Patienten, deren Herde nicht histopathologisch untersucht oder nicht mindestens 2 Jahre beobachtet wurden ohne Wachstum nachzuweisen, wurden hinsichtlich der Dignität der Herde als nicht sicher klassifiziert eingestuft.

Als Beispiel hierfür wird ein 67jähriger Patient angeführt, bei dem ein pulmonaler Rundherd entdeckt wurde. Anamnestisch war ein langjähriger Nikotinabusus bekannt. Bei Verlaufskontrollen wurde ein deutliches Wachstum des Herdes nachgewiesen, so dass dringender Verdacht auf Malignität bestand. Allerdings lehnte der Patient alle weiterführenden diagnostischen und therapeutischen Maßnahmen ab, so dass trotz eines dringenden Verdachtes auf ein Bronchialkarzinom die Dignität als unklar eingestuft wurde.

Bei den anderen unklaren Fällen handelte es sich um Patienten, die sich noch innerhalb des Beobachtungszeitraumes von 2 Jahren befanden, keine Größenänderung des Herdes aufwiesen und aus diesem Grund noch nicht eindeutig klassifizierbar waren.

Unter den bereits als maligne bekannten 13 Herden befanden sich 6 Primärtumoren der Lunge (zwei Adenokarzinome, ein Plattenepithelkarzinom, ein undifferenziertes Karzinom, ein Alveolarzellkarzinom und ein nicht näher spezifiziertes nicht-kleinzelliges Bronchialkarzinom (NSCLC)). Weiterhin wurden 7 zum Teil multiple Metastasen diagnostiziert, die aus Primärtumoren anderer Körperregionen stammen (Metastasen eines Adenokarzinoms, eines Pankreaskarzinoms, eines malignen fibrösen Histiocytems, eines adenoidzystischen Karzinoms sowie die eines Ewing-Sarkoms), außerdem bei einem Patienten zwei pulmonale Manifestationen eines Morbus Hodgkin.

Bei den 20 als benigne eingestuften Herden, wurden 8 als entzündlich klassifiziert, weiterhin konnte bei der Resektion des Herdes je eine Narbe, ein Hamartom sowie ein anthrakotisch veränderter pulmonaler Lymphknoten nachgewiesen werden. Bei 9 Patienten konnte keine genauere Identifikation der Läsion vorgenommen werden, da keine histopathologische Untersuchung stattfand. Die Benignität wurde anhand des fehlenden Wachstums oder der Größenregredienz innerhalb des Beobachtungszeitraumes sicher festgestellt.

Die Doppelmessungen zur Bestimmung der Reproduzierbarkeit beider Verfahren wurden an einem anderen Patientenkollektiv durchgeführt. Es



handelte sich um 10 Patienten, darunter 2 Frauen und 8 Männer, die bekannte multiple pulmonale Metastasen solider Primärtumoren hatten.

Die Untersuchungen erfolgten im Rahmen einer von der zuständigen Ethikkommission genehmigten Studie zum Vergleich eines Niedrigdosis-CT-Protokolls mit dem Standard-CT zur Detektion von Lungenrundherden. Die Patienten waren zwischen 35 und 75 Jahren alt mit einem Altersdurchschnitt von 49,9 Jahren.

Unter den Primärtumoren anderer Organe, die bei diesen Patienten Lungenmetastasen entwickelt hatten, befanden sich 2 Mammakarzinome, 2 Nierenkarzinome, 2 Hodentumoren und ein Harnblasenkarzinom. Bei einem Patienten war der Ursprung der Metastasen nicht geklärt, bei 2 Patienten befand sich auch der Primärtumor in der Lunge.

Insgesamt wurden bei den 10 Patienten 75 Herde vermessen, die mit dem 2D-Verfahren ausgewertet wurde. 90 Herde wurden mit der 3D-Methode vermessen und ausgewertet. Alle Herde hatten einen Durchmesser von 5 mm oder mehr. Die Herdgrößen variierten zwischen 5,6 und 30,1 mm mit einem Mittelwert von 13,2 mm.

### **3.2. Untersuchungsprotokoll**

Die 47 Patienten wurden nach folgendem Protokoll untersucht: Zunächst wurden sie über die wissenschaftliche Verwendung ihrer durch die Untersuchung gewonnenen Daten, den Zweck der Untersuchung sowie deren Risiken aufgeklärt und stimmten dieser zu.

Vor der Untersuchung wurden sie über das Vorgehen bei dieser informiert. Es wurde ihnen die Atemtechnik erklärt, die während der Scans notwendig ist.

Weiterhin wurden sie nach relevanten Vorerkrankungen und Allergien befragt Neben Unverträglichkeitsreaktionen, die eventuell bereits bei vorherigen Untersuchungen mit jodhaltigem Röntgenkontrastmittel aufgetreten sind wurde insbesondere das Vorliegen von Nieren- und Schilddrüsenerkrankungen erfragt,

da das Kontrastmittel renal ausgeschieden wird und aufgrund seines Jodgehaltes eine Hyperthyreose induzieren könnte.

Die Applikation des Kontrastmittels wurde vorbereitet, indem ein intravenöser Zugang in eine antecubitale Vene gelegt wurde.

Anschließend wurden die Patienten in Rückenlage auf dem Untersuchungstisch des jeweiligen Computertomographen gelagert. Zur Untersuchung wurden Geräte der Firmen Siemens (Somatom VZ, Sensation 16) und Philips (AVE) eingesetzt. Die relevanten Messparameter waren für alle Messungen gleich:

Die Messungen wurden auf den zu untersuchenden Herd fokussiert. Die Scanstrecke betrug 40-70 mm, die Röhrenspannung 120 kV. Der effektive Röhrenstrom wurde auf 100 mAs eingestellt. Die Schichtkollimation war auf 4x1 mm (Siemens Somatom VZ), 16x0,75 mm (Siemens Sensation 16) bzw. auf 1 mm (Philips AVE) gewählt worden bei einem Pitch-Faktor von 1,75. Als Rekonstruktionskernel fand der jeweilige Standard-Mediastinalkernel Verwendung (Siemens B30f, Philips Kernel 4). Die Schichtdicke eines Bildes wurde eingestellt auf 1,25 mm (Somatom VZ) bzw. 1 mm (AVE und Sensation 16), das Rekonstruktionsintervall betrug 0,8 mm.

So wurde zunächst eine Aufnahme ohne Gabe von Kontrastmittel (eine sogenannte Nativaufnahme) des Herdes angefertigt. Anschließend wurden dem Patienten über eine Doppelkolbenspritze 110 ml eines iodhaltigen, wasserlöslichen Röntgenkontrastmittels (Solutrast 300, Schering AG, Berlin, Deutschland), danach noch 30 ml Kochsalzlösung als Bolus intravenös verabreicht. Dabei wurde ein Flow von 3 ml/s gewählt.

Die Zeit nach Beginn der Kontrastmittelgabe wurde registriert und es wurden weitere Aufnahmen angefertigt. Diese erfolgten 60, 120, 180 und 240 Sekunden nach Beginn der Kontrastmittelverabreichung. Dabei wurden die gleichen Parameter angewendet wie bei der Nativaufnahme.

Bei einem Patienten erfolgte die fünfte Aufnahme irrtümlich nicht 240, sondern 300 Sekunden nach Beginn der Kontrastmittelgabe. Trotz Abweichens vom Protokoll in diesem Punkt wurden die Daten des Patienten für diese Studie dennoch ausgewertet und genutzt, da keine relevanter Einfluss auf die Fragestellung dieser Studie zu erwarten war.

Alle Scans wurden, wie für Untersuchungen des Thorax üblich, in suspendierter tiefer Inspiration akquiriert.

Die Doppeluntersuchungen erfolgten nach folgendem Protokoll: Die Untersuchungen wurden mit dem Siemens-Gerät Somatom VZ (Somatom VolumeZoom, Siemens Medical Solutions, Forchheim, Deutschland) durchgeführt. Innerhalb weniger Minuten wurden bei den 10 zu untersuchenden Patienten zwei Niedrigdosis-Scans des Thorax vorgenommen, die identische Scanparameter hatten: Röhrenspannung 120 kVp, effektiver Röhrenstrom 20 mAs, Kollimation 4x1 mm, Schichtdicke 1,25 mm, Rekonstruktionsintervall 0,8 mm, Standard-Lungenkernel (B50f). Bei beiden Untersuchungen erfolgte keine Applikation von Kontrastmittel.

Die durch die Untersuchungen gewonnenen Daten wurden auf eine Workstation geladen und konnten nach den im Folgenden beschriebenen Methoden ausgewertet werden.

### **3.3. Auswertung der Daten**

Die einzelnen Scans wurden jeweils einmal nach der Methode rekonstruiert und vermessen, die Swensen (98) vorgeschlagen hat. Anschließend wurden die Herde segmentiert, um die daraus berechneten dreidimensionalen Bilder ebenfalls auszuwerten.

Analog wurden die Doppelmessungen nach der Swensen-Methode und der 3D-Methode ausgewertet und anschließend miteinander verglichen.

### 3.3.1. Auswertung nach Swensen

Die Rekonstruktion der Bilder erfolgte mittels einer Graphik-Workstation (Wizard™, Siemens Medical Solutions, Forchheim, Deutschland) und einer MPR-Software (MultiPlanar Reformatting-Software).

Nach dem Protokoll von Swensen wurden so alle vorliegenden Bilder einheitlich rekonstruiert, und zwar auf eine Schichtdicke von 3 mm und jeweils 1 mm Abstand zwischen zwei Schichten. Dabei wurden jeweils axiale Schichten rekonstruiert

Nur bei einem Patienten konnte dies nicht erfolgen. Die Scan-Parameter bei seiner Messung waren auf eine Schichtdicke von 4 mm eingestellt worden, so dass eine Rekonstruktion auf 3 mm nicht möglich war. Diese Bilder wurden bei der späteren Dichtebestimmung mit 4 mm Schichtdicke vermessen.

Die Dichtemessung der einzelnen Herde erfolgte anhand der Bilder, die aus der obigen Rekonstruktion gewonnen wurden.

Es wurde die Schicht gewählt, auf der der zu untersuchende Herd in seiner größten Ausdehnung zu sehen war. Um vergleichbare Dichtemessungen zwischen den einzelnen Scans vor Kontrastmittelgabe und jeweils 60, 120, 180 und 240 Sekunden nach Beginn der Kontrastmittelgabe vornehmen zu können, wurde darauf geachtet, aus den vorliegenden Schichten möglichst gleiche Anschnitte des Herdes auszuwählen. Als dies erfolgt war, wurde der Herd in die Mitte des Bildschirms zentriert und maximal vergrößert, um besser den auszumessenden Bereich innerhalb des Herdes definieren zu können.

Dieser Bereich, die sogenannte Region Of Interest (ROI), konnte mittels einer ovalen Markierung festgelegt werden, in deren Grenzen die Messwerte bestimmt wurden. Nach dem Swensen-Protokoll sollte die Größe der Region möglichst zentral innerhalb des zu untersuchenden Herdes liegen. Ihre Größe sollte etwa 70% des größten Durchmessers des Herdes betragen. Dieses Vorgehen war nicht immer in dieser Form umzusetzen, da es Herde gab, die nicht rund oder oval geformt waren, sondern länglich bis sichelförmig.

Weiterhin war es wichtig, kein Partialvolumen zu erfassen, um die Messergebnisse der Dichte nicht zu verfälschen. Unter Partialvolumen versteht man einen Bereich, der nicht nur die Struktur, deren Dichte man messen will (beispielsweise einen Tumor) enthält, sondern zusätzlich noch umgebendes Gewebe. Da stark luftgefülltes Lungengewebe extrem geringe Dichtewerte aufweist, würde die Messung der mittleren Dichte dadurch sehr beeinflusst werden. Aus diesem Grund musste die Messregion teilweise in homogenere Randbereiche des Herdes gelegt werden bzw. es konnte eine Größe von 70% des maximalen Herddurchmessers nicht eingehalten werden.

Bei den fünf zu vergleichenden Scans wurden die Messregionen außerdem in möglichst identischer Weise festgelegt, so dass durch gleiche Lokalisation der Regionen eine Vergleichbarkeit der Dichtemessungen gegeben ist.

Bei der anschließenden Dichtemessung wurden bei allen 5 Aufnahmen die durchschnittliche Dichte, die in Hounsfield-Einheiten (HU) angegeben wurde, der maximale und der minimale Dichtewert (ebenfalls in HU) sowie die Standardabweichung erfasst, die alle vom oben genannten System in der Region of Interest berechnet werden konnten.

Bei den jeweiligen Messungen bestand keine Kenntnis darüber, ob es sich bei dem jeweiligen Herd um einen bereits diagnostizierten benignen oder malignen Prozess handelte.

Analog erfolgte die Auswertung der Daten, die aus den Doppeluntersuchungen gewonnen wurden. Hier wurde bei beiden Untersuchungen die Dichtemessungen wie oben beschrieben durchgeführt.

### 3.3.2. Auswertung nach dem 3D-Verfahren

Die Auswertung nach dem 3D-Verfahren erfolgte ebenfalls auf der oben genannten Wizard<sup>TM</sup>-Workstation, die zusätzlich mit dem Prototypen eines Programmes (syngo LungCare CT<sup>TM</sup>, Siemens AG, Forchheim, Deutschland) ausgestattet war, welcher neben der Segmentierung auch die Bestimmung der mittleren Dichte der Herde ermöglichte. Der Algorithmus der Segmentierung wurde ausführlich von Fan et al. (19) beschrieben. Er erfolgt in mehreren Schritten:

Nachdem man den zu untersuchenden Herd identifiziert hat, wird dieser mit einem Mausklick in dessen zentralen Anteil markiert. Von diesem Punkt ausgehend führt das Programm zunächst eine grobe Segmentierung des Herdes durch. Mit einem Schwellenwert für die Dichte lässt es einen Bereich um den markierten Punkt wachsen, bis am Rande des Prozesses die Dichte wieder abfällt. Diesen Vorgang bezeichnet man als region growing.

Da in der so entstehenden groben Darstellung des Herdes auch noch Brustwand enthalten sein kann (wie es beispielsweise bei pleuraassoziierten Prozessen der Fall ist) werden diese Strukturen anschließend durch spezielle Berechnungsverfahren ausgeschlossen.

Anschließend erfolgt eine Schätzung des Herdes hinsichtlich seiner Position, Größe und Form. Dann ermittelt das Programm das Zentrum bzw. den Kern des zu untersuchenden Herdes und kann anhand dieser Daten die Größe und Ausrichtung des Kerns bestimmen.

Schließlich wird in einem nächsten Schritt ein kugelförmiges 3D-Template am Kern angelegt, welches kleiner als der berechnete Kern, aber in gleicher Weise ausgerichtet ist. Dieses Template wird nun so modifiziert, dass es den Herd in bestmöglicher Weise repräsentiert.

Dieses geschieht wie folgt: Das ursprüngliche Template vergrößert sich schrittweise. Bei jedem Schritt erfolgt ein Vergleich mit dem Ergebnis aus dem

im ersten Schritt erfolgten region growing. Es wird eine Cross-Korrelation beider Größen berechnet. An dem Punkt, an dem die Korrelation zwischen beiden ihr Maximum erreicht, wird das Template um noch einen weiteren Schritt vergrößert. Dadurch wird der Punkt der visuell besten Übereinstimmung zwischen dem Template und dem Ergebnis des region growing erreicht, und man erhält das optimale Template.

Anschließend werden die Ergebnisse des region growing und des optimalen Templates miteinander im Sinne einer UND-Verknüpfung kombiniert.

Das bedeutet, dass nur Punkte einbezogen werden, die sowohl innerhalb des Templates als auch innerhalb des Bereiches des region growings liegen. Dadurch fallen sowohl viele Oberflächenstrukturen des Herdes, als auch umgebende Strukturen wie assoziierte Gefäße oder Brustwand heraus.

Da aber bestimmte Oberflächenstrukturen, wie zum Beispiel Tumorausläufer bei der Beurteilung, ob ein Herd benigne oder maligne ist, wichtig sein können, erfolgt noch eine weitere Modifizierung.

Alle umliegenden Strukturen, die bei der Verknüpfung herausgefallen sind und eine bestimmte Distanz zum segmentiertem Herd haben, werden nach bestimmten Kriterien analysiert und dann unter Umständen wieder als Teil des Herdes definiert. Daraus erhält man dann eine realitätsnähere Darstellung des Tumors mit seinen Ausläufern und Oberflächenstrukturen, der entweder einzeln liegt oder an der Brustwand oder Gefäßen anhängig ist.

Die Segmentierung unterteilt den Herd in einzelne dreidimensionale Würfel, die man als Voxel bezeichnet.

Neben dem Volumen des Herdes wird seine Ausdehnung in X-, Y-, und Z-Richtung, seine maximalen und minimalen Durchmesser, sowie durch Mittelung der Dichte der einzelnen Voxel die durchschnittliche Dichte des Herdes angegeben.

Erfasst wurden bei diesen Messungen die durchschnittliche Dichte (in HU), die Standardabweichung der Dichtewerte, das Herdvolumen (in  $\text{mm}^3$ ) sowie seinen maximalen und minimalen Durchmesser.

### 3.3.3. Statistische Auswertung der Daten

Die statistische Auswertung der Daten soll drei Ziele verfolgen. Zum ersten soll untersucht werden, ob oder in welchem Masse eine Übereinstimmung der Messergebnisse zwischen dem 2D-Verfahren nach Swensen und dem neuen 3D-Ansatz besteht. In einem zweiten Schritt geht es darum, Aussagen hinsichtlich der Testgütekriterien wie Sensitivität, Spezifität sowie negative und positive prädiktive Werte zu ermitteln und mit diesen Daten eine ROC-Analyse (38) durchzuführen. In einem dritten Schritt sollen die Ergebnisse der Doppelmessungen Aufschluss darüber geben, welches der beiden Verfahren eine bessere Reproduzierbarkeit der Messergebnisse aufweist.

Die Auswertung erfolgte mit den Statistik-Programmen SPSS™ (Version 11.0 für Windows) und Microsoft Excel 2002™ (Microsoft Office XP für Windows)

#### 3.3.3.1. *Übereinstimmung (Agreement) beider Messmethoden*

Hinsichtlich der Übereinstimmung der Messergebnisse stellt sich zunächst die Frage, welche Daten für diese Analyse herangezogen werden müssen. So liegt die Problematik darin, dass beide Verfahren ganz unterschiedliche Bereiche der Herde ausmessen. Während nach dem Swensen-Verfahren nur eine Schicht des Herdes vermessen wird, und daraus nur eine bestimmte Region of Interest, wird beim 3D-Verfahren der komplette Herd in seiner durchschnittlichen Dichte erfasst. Das diese Messungen nicht übereinstimmen können liegt auf der Hand.

Die Fragestellung dieser Studie bezieht sich allerdings auch nicht auf die absoluten Dichtewerte, sondern vielmehr auf die Anreicherungen bzw. Dichtedifferenzen, die sich nach der Gabe von Kontrastmittel ergeben.

Wenn also beide Verfahren übereinstimmende Ergebnisse zeigen sollen, so muss dafür die Differenz der Dichte nach Kontrastmittelgabe ( $D_n$ ) und vor Kontrastmittelgabe ( $D_v$ ) zugrunde gelegt werden.



Es ergeben sich so jeweils vier Werte, die die Dichtedifferenz 60, 120, 180 und 240 Sekunden nach Kontrastmittelgabe im Vergleich zur Nativaufnahme zeigen:

|               |   |                            |                   |
|---------------|---|----------------------------|-------------------|
| Dv – Dn (60)  | = | Kontrastmittelanreicherung | nach 60 Sekunden  |
| Dv – Dn (120) | = |                            | nach 120 Sekunden |
| Dv – Dn (180) | = |                            | nach 180 Sekunden |
| Dv – Dn (240) | = |                            | nach 240 Sekunden |

Für die statistische Auswertung erhält man so bei 50 Herden 200 Werte jeweils für das 2D- und 3D-Verfahren, die man miteinander vergleichen kann.

Für die Auswertung wurde ein Verfahren angewendet, dass 1986 von Bland und Altman (7) vorgeschlagen wurde. Durch einfache graphische Darstellung der Daten und wenige Berechnungen, können beide Methoden hinsichtlich ihrer Übereinstimmung untersucht werden.

In einem ersten Schritt werden die Daten in einem einfachen Diagramm dargestellt. Dabei repräsentiert die X-Achse die Dichtedifferenz des 2D-Verfahrens, die Y-Achse den des 3D-Verfahrens.

Man erhält so für jedes Wertepaar (2D/3D) einen Punkt im Koordinatensystem und kann sich nun die Verteilung der Punkte anschauen. Als Maßstab kann die Winkelhalbierende eingezeichnet werden, die die völlige Übereinstimmung beider Verfahren repräsentiert ( $X=Y$ ).

Abb. 3 zeigt beispielhaft ein derartiges Diagramm, dass zur Darstellung des Agreements zweier Verfahren eingesetzt werden kann.

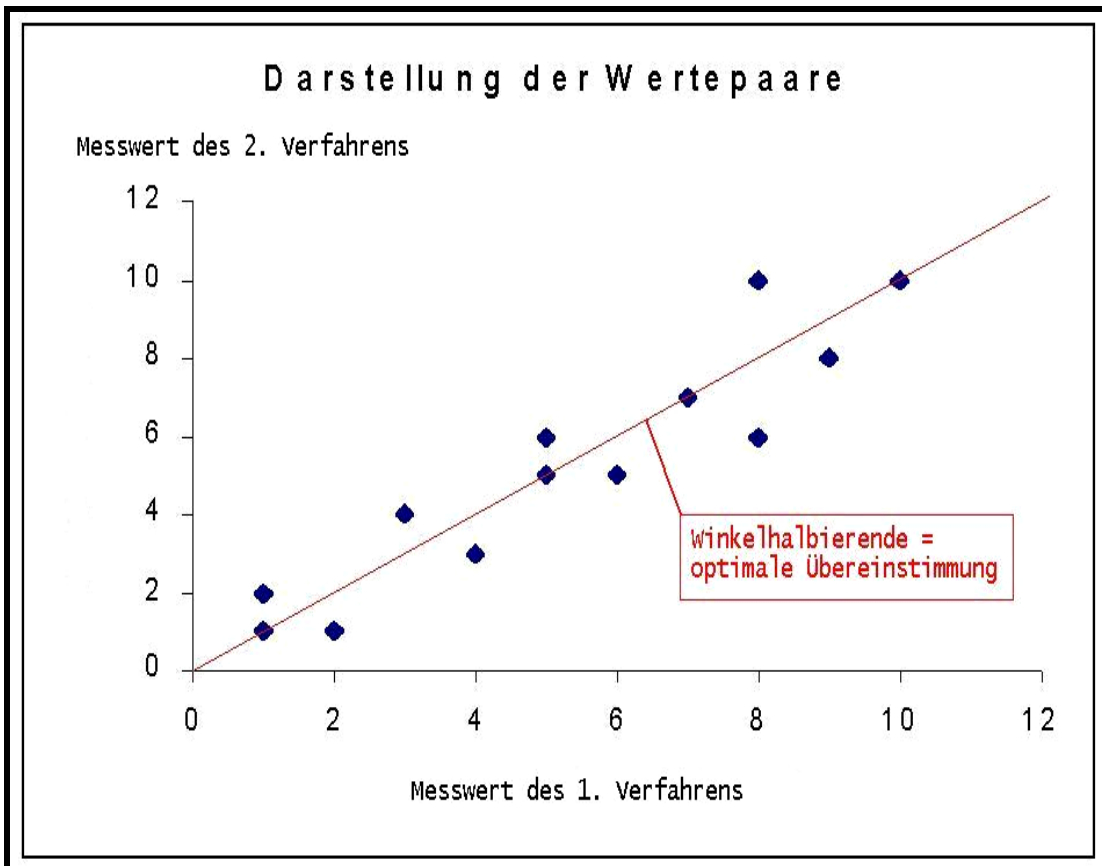


Abb. 3: Gegenüberstellung der Messwerte beider Verfahren (nach Bland und Altman, 1986 (7))

Eine zweite Graphik soll die Unterschiede beider Verfahren noch deutlicher machen. Gemäß Abb. 4 wird auf der X-Achse jeweils der Mittelwert eines Wertepaares aufgetragen werden, während auf der Y-Achse die zugehörige Differenz beider Werte erfasst wird. Durch drei horizontale Linien werden die durchschnittliche Abweichung (Mittelwert)  $M$  beider Verfahren markiert, sowie ein Bereich, der durch die doppelte Standardabweichung ( $SD$ ) nach oben und unten begrenzt wird. Dieser Bereich hat also als Obergrenze  $M + 2 SD$  und als Untergrenze  $M - 2 SD$ .

Für den Fall, dass die Differenzen beider Verfahren einer Normalverteilung nach Gauss folgen, würden in diesem Bereich 95% aller Werte liegen.

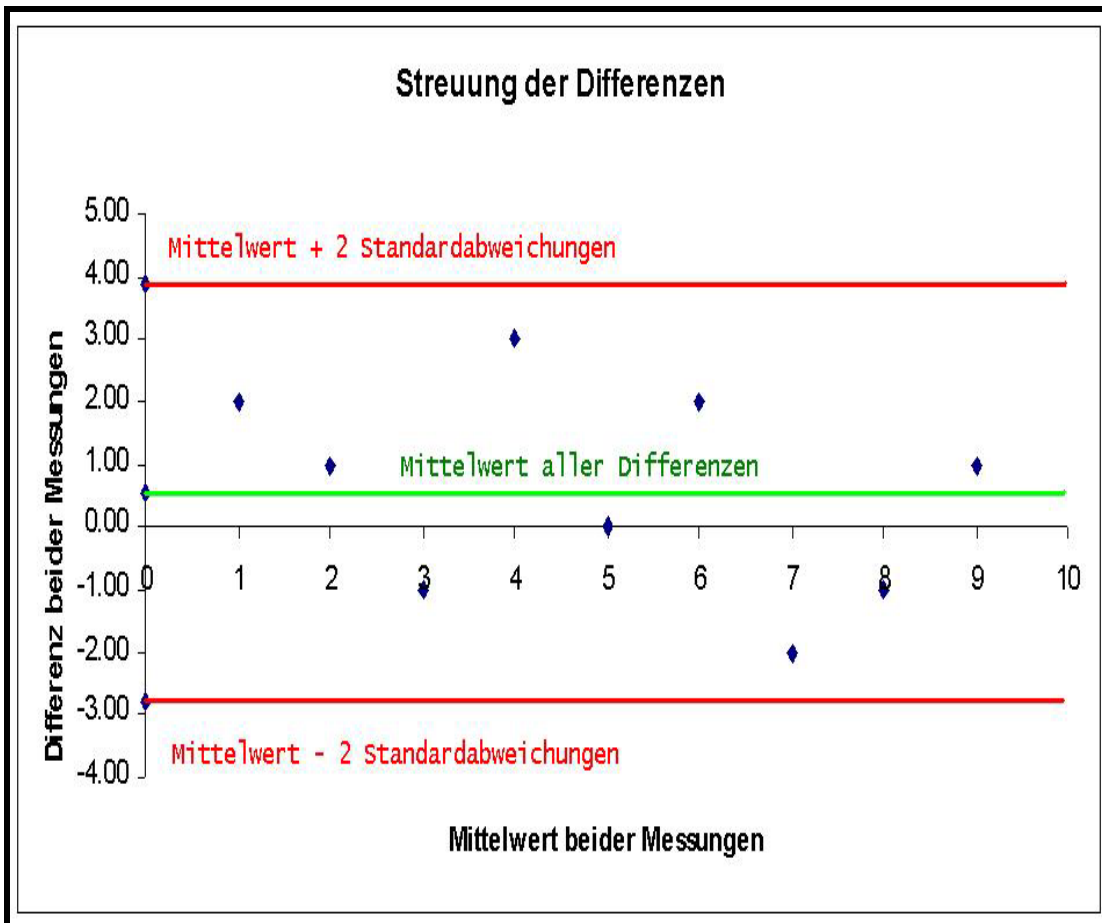


Abb. 4: Darstellung der Differenzen beider Verfahren (nach Bland und Altman 1986 (7))

Aus dieser Darstellung lassen sich nun wichtige Aussagen ableiten, hinsichtlich der Größe des Streubereiches sowie des Zusammenhanges zwischen der absoluten Größe der Dichtedifferenz (X-Achse) und der Differenz beider Verfahren (Y-Achse).

### 3.3.3.2. Testgütekriterien beider Verfahren

Die Testgütekriterien lassen sich durch einfache Berechnungen ermitteln. Zugrunde gelegt wurden die Daten der Herde, die sicher maligne oder benigne diagnostiziert waren, was auf insgesamt 33 der 50 Herde zutraf. Als Schwellenwert für eine signifikante Anreicherung wurde der von Swensen empirisch ermittelten Wert von 15 HU gewählt. Herde, die bei einer der 4

Messungen nach Kontrastmittelgabe 15 HU oder mehr anreicherten, wurden als positiv getestet eingestuft. Man berechnet nun:

$$\text{Sensitivität} = \frac{\text{Zahl der positiv getesteten malignen Herde (> 15 HU)}}{\text{alle maligne Herde}}$$

$$\text{Spezifität} = \frac{\text{Zahl der negativ getesteten benignen Herde (< 15 HU)}}{\text{alle benigne Herde}}$$

$$\text{Positiv prädiktiver Wert (PPV)} = \frac{\text{Zahl der malignen Herde}}{\text{alle positiv getesteten Herde (> 15 HU)}}$$

$$\text{Negativ prädiktiver Wert (NPV)} = \frac{\text{Zahl der benignen Herde}}{\text{alle negativ getesteten Herde (< 15 HU)}}$$

$$\text{Akkuranz} = \frac{\text{Zahl der mit dem Verfahren richtig diagnostizierten Herde}}{\text{alle Herde}}$$

Aus den Angaben zu Sensitivität und Spezifität lässt sich nun eine ROC-Analyse erstellen, die Auskunft über die diagnostische Qualität des jeweiligen Verfahrens gibt (38). Eine ROC-Analyse wird durchgeführt, indem man für jeden Schwellenwert, der ja Beurteilungsgrundlage dafür ist, ob ein Herd als maligne oder benigne eingestuft wird, jeweils einen zugehörigen Wert für die Sensitivität und Spezifität berechnet, die damit erreichbar sind.

Nun bildet man Wertepaare für jeden einzelnen Schwellenwert bestehend aus Sensitivität und 1-Spezifität und trägt die Punkte in ein Koordinatensystem ein. Dabei wird auf der x-Achse der Wert für 1-Spezifität, auf der y-Achse der Wert für die Sensitivität aufgetragen. Man erhält eine Kurve die für jeden Schwellenwert die jeweiligen Testgütekriterien angibt.

Die Fläche unter der Kurve (Area Under the Curve, AUC) ist ein Maß für die Qualität des Verfahrens und beträgt im Idealfall 1, was bedeuten würde, dass für jeden Schwellenwert eine Sensitivität von 1 und eine Spezifität von 1 (1-Spezifität=0) bestehen würde.

Vergleichende graphische Darstellung nach dieser Methode sowie Berechnung der AUC-Werte erfolgt nun für das 2D und das 3D-Verfahren. Dabei wird für die Ermittlung der Testgütekriterien die jeweils maximale Anreicherung, die sich aus allen 4 Messungen nach Kontrastmittelgabe ergibt, herangezogen.

Nachdem so die beiden Verfahren verglichen wurden, wird nun untersucht, ob sich die jeweils maximalen Anreicherungen, wie Swensen vorgeschlagen hatte, am besten für die Berechnung eignen, oder ob noch bessere Resultate mit anderen Methoden zu erreichen sind, nämlich indem man statt der maximalen Anreicherung andere Werte für die Berechnung zugrunde legt.

Dazu werden anhand der 2D-Daten weitere ROC-Analysen durchgeführt, deren Werte für Sensitivität und Spezifität allerdings nicht mit der maximalen Anreicherung berechnet wird, sondern mit der durchschnittlichen Anreicherung nach Kontrastmittelgabe. Anschließend werden für die einzelnen Messwerte 60, 120, 180 und 240 Sekunden nach Kontrastmittelgabe ROC-Analysen durchgeführt und zum Abschluss mit dem maximalen Anreicherungswert aus den ersten drei Messungen.

Daraus werden die AUC-Werte ermittelt und die Frage beantwortet, mit welchen Messungen, Maxima oder Mittelwerten die besten Resultate erzielt werden können.

#### *3.3.3.3. Statistische Auswertung der Doppeluntersuchungen*

Alle Herde wurden jeweils zweimal nacheinander mit den gleichen Parametern gescannt und dann nach der 2D-Methode und der 3D-Methode ausgewertet.

Nun wurde untersucht, welche der beiden Methoden eine bessere Reproduzierbarkeit der Ergebnisse aufweist. Dazu werden zunächst für beide

Methoden die Differenzen gebildet, die sich aus den beiden Messungen ergeben. Daraus würde sich dann der absolute Messfehler berechnen, der für beide Verfahren festgestellt werden könnte. Bei optimaler Reproduzierbarkeit, also gleichen Messergebnissen bei beiden Messungen würde sich eine Differenz von 0 ergeben. Das bedeutet, ein Messfehler wäre nicht feststellbar. Ergänzend wird die Standardabweichung des Messfehlers angegeben, die Auskunft über dessen Streuung bei den einzelnen Messungen gibt.

Zur graphischen Veranschaulichung werden die Daten noch wie oben beschrieben nach der Methode von Bland und Altman ausgewertet. Diesmal werden aber nicht die Messergebnisse der 2D- und 3D-Methode vergleichend gegenübergestellt, sondern die Ergebnisse der beiden Scans der Doppeluntersuchung für jedes Verfahren separat.

Schließlich soll noch eine Aussage über die Rate der falsch positiven und falsch negativen Testergebnisse gemacht werden, die sich aus der Doppeluntersuchung ergeben würden. Da diese Untersuchungen ohne Gabe von Kontrastmittel durchgeführt wurden, kann bei keinem Herd eine Anreicherung stattgefunden haben. Somit würden alle Herde, deren Dichte bei der zweiten Messung um mehr als 15 HU (sofern man den Schwellenwert von Swensen zugrunde legt), zugenommen hat, als falsch positive Messungen ansehen können. Entsprechendes gilt für alle anderen Schwellenwerte.

Die Ermittlung der Rate der falsch negativen Befunde erfordert eine theoretische Überlegung. Da keine echte Anreicherung stattgefunden haben kann, wird ein Enhancement simuliert. Dazu nimmt man die Verteilung der Dichtewerte der vermessenen Herde und addiert nun jeweils den Betrag hinzu, den man als Anreicherung simulieren möchte. Dies beruht auf der Überlegung, dass bei einer Anreicherung um diesen Betrag eine ähnliche Verteilung der Dichtewerte vorhanden wäre, nur eben um das simulierte Enhancement nach oben verschoben. Diese Methode wurde in einer Arbeit von Wormanns et al. (104) beschrieben.

So kann man für jede beliebige Anreicherung die Verteilung der Dichtewerte simulieren und daraus die Rate der falsch negativen Befunde berechnen. Diese ergibt sich aus den Herden, die um einen geringeren Betrag als den jeweiligen zugrunde liegenden Schwellenwert anreichern würden.

Aus den Raten für die falsch positiven und den falsch negativen Ergebnisse bei verschiedenen Schwellenwerten für die Feststellung von signifikantem Enhancement (beispielsweise bei Swensen: 15 HU) kann nun wie oben beschrieben eine ROC-Analyse durchgeführt werden. Vergleichend können nun die ROC-Analysen für beide Methoden gegenübergestellt werden.

## 4. Ergebnisse

### 4.1. Übereinstimmung (Agreement) beider Methoden

Nach Auswertung der Messwerte ergaben sich folgende Ergebnisse.

Aus Abb. 5 und 6 können die graphischen Darstellungen nach Bland und Altman (7) entnommen werden.

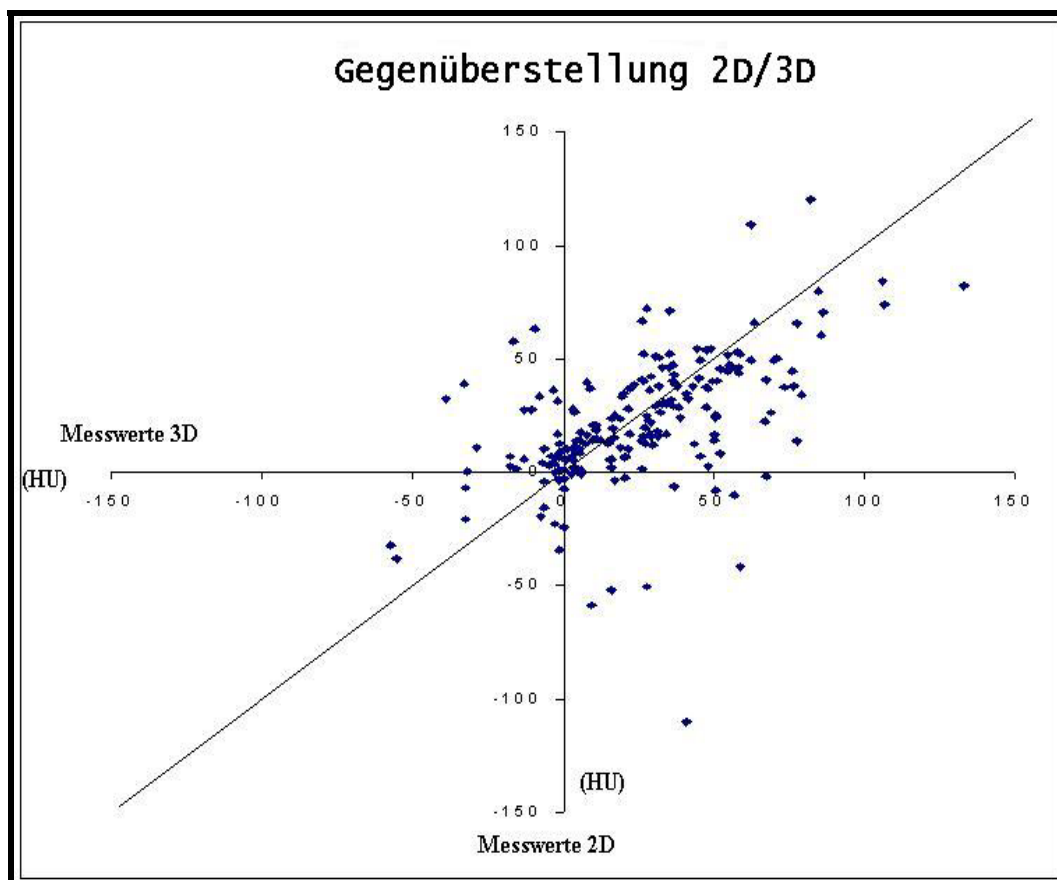


Abb. 5 Grad der Übereinstimmung der Messwerte beider Verfahren für die Herddichte

Abb. 5 zeigt die Anreicherungen, die nach der 2D- und der 3D-Methode bestimmt wurden. Würden beide Verfahren in ihren Ergebnissen übereinstimmen, lägen alle Punkte auf der eingezeichneten Winkelhalbierenden. Man erkennt, dass sich zwar die Punkte um die Winkelhalbierende scharen, aber auch, dass eine ausgeprägte Streuung besteht, die durch zum Teil erhebliche Diskrepanzen der Messwerte beider



Verfahren zustande kommt. Die Verteilung der Wertepaare lässt erkennen, dass mehr Wertepaare unterhalb der Winkelhalbierenden liegen. Auch sind die Abweichungen nach unten und rechts größer als nach oben und links. Daran kann man erkennen, dass im Mittel mit dem 2D-Verfahren (X-Achse) höhere Werte gemessen wurden, als mit der 3D-Methode (Y-Achse).

Nach diesem ersten Überblick über die Übereinstimmung beider Methoden, können konkretere Informationen aus Abb. 6 entnommen werden.

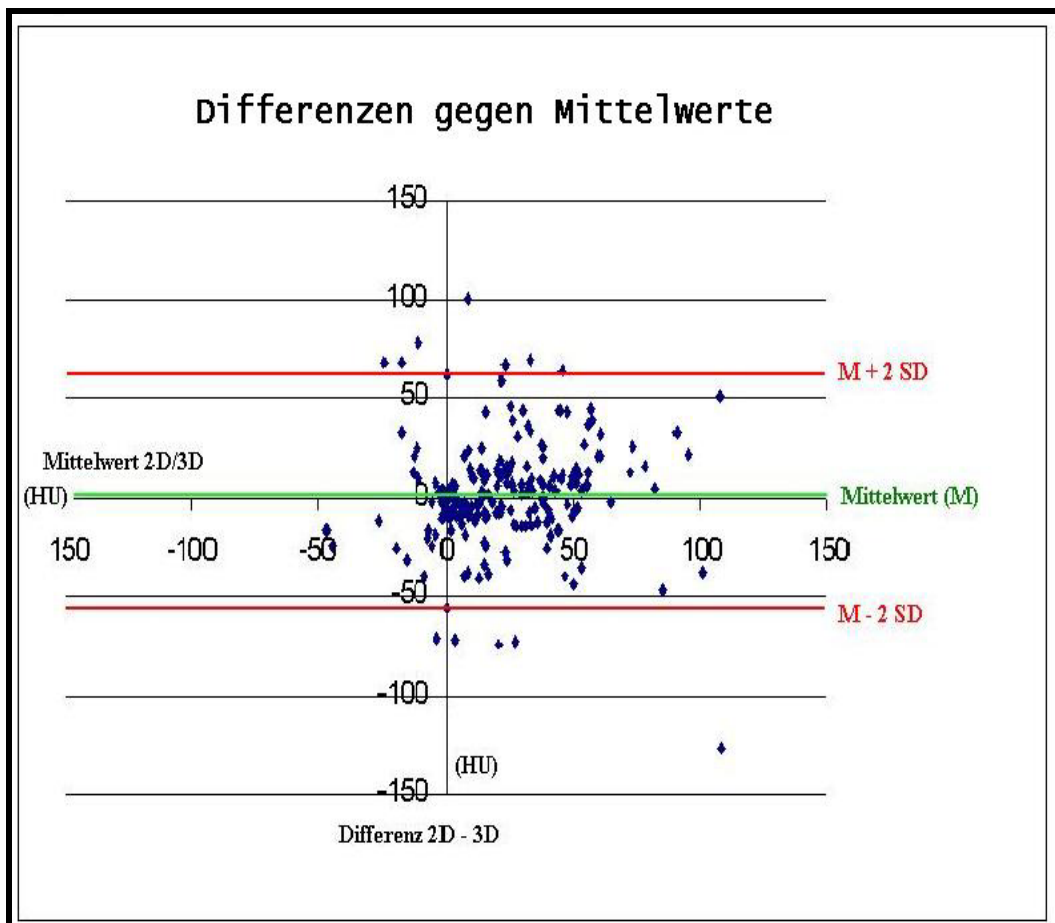


Abb. 6 Darstellung der Differenzen beider Verfahren (M ist der Mittelwert aus der mit 2D- und 3D-Verfahren gemessenen Dichte)

Die Standardabweichung der einzelnen Differenzen vom Mittelwert beträgt 29,6. Bestimmt man nun einen Bereich, der durch die zweifache Standardabweichung um den Mittelwert nach oben und unten begrenzt wird, so liegt dessen untere Grenze bei -56,6 HU ( $0 - 2 \times 29,6$ ) und die obere Grenze

bei 61,8 HU ( $2,6 + 2 \times 29,6$ ). In diesem Bereich liegen 183 der insgesamt 196 Differenzen, was einem Anteil von 93,4% entspricht.

Aus den Differenzen, die man aus den jeweilig berechneten Anreicherungen beider Verfahren ermitteln konnte, lässt sich ein Mittelwert von 2,6 HU bestimmen. Da die Differenz  $D$  aus  $D = \text{Anreicherung } 2D - \text{Anreicherung } 3D$  berechnet wurde, bedeutet das, dass das 2D-Verfahren im Mittel 2,6 HU mehr Anreicherung gemessen hat, als das 3D-Verfahren. Dieses Ergebnis korreliert mit der Darstellung in Abb. 5. Aufgrund der hohen Standardabweichung und des daraus resultierenden großen Konfidenzintervalles kann man von diesem nur relativ gering von 0 abweichenden Wert allerdings nicht auf einen signifikanten Verfahrensunterschied schließen.

In der Graphik ist weiterhin eine gleichmäßige Verteilung der Wertepaare um die Linie des Mittelwertes zu erkennen. Man erkennt keine Abhängigkeit der Differenz beider Messungen zur jeweiligen Höhe der Messung, die auf der X-Achse durch den Mittelwert beider Messungen repräsentiert wird.

#### **4.2. Testgüteparameter**

Als nächstes wurde untersucht, welches Verfahren besser zwischen benignen und malignen Läsionen unterscheiden konnte. Von den insgesamt 50 Herden konnte im Verlauf der Studie bei 33 Herden durch histopathologische Untersuchung oder klinische Verlaufskontrolle die Dignität festgestellt werden.

Aus diesen Daten ließen sich für beide Verfahren die 4-Felder-Tafeln ermitteln, die in Tab. 1 und 2 dargestellt sind. Als Schwellenwert für die Beurteilung, ob ein Herd als benigne oder maligne klassifiziert wird, wurde der von Swensen empirisch ermittelte Wert von 15 HU gewählt.

Mit der 2D-Methode wurden alle 13 malignen Tumoren erkannt, was einer Sensitivität von 100% entspricht.

Von den 20 benignen Läsionen reicherten 4 weniger als 15 HU an. 16 wurden demnach aufgrund ihrer Anreicherung als maligne eingestuft. Daraus ergibt sich eine Spezifität von 20,0%.

Von den insgesamt 29 Herden, die mehr als 15 HU anreicherten, waren 13 tatsächlich maligne (positiv prädiktiver Wert = 44,8%).

Alle Herde, die weniger als 15 HU anreicherten, stellten sich auch in der Histopathologie bzw. nach klinischen Verlaufskontrollen als gutartig heraus (negativ prädiktiver Wert = 100%).

Insgesamt konnten mit dem 2D-Verfahren 17 der 33 richtig einer Dignität zugeordnet werden. Dies entspricht einer Akkuranz von 51,5%.

| <b>2D</b>                      | <b>benigne</b>   | <b>maligne</b>   | <b>Summe</b>      |
|--------------------------------|------------------|------------------|-------------------|
| <b>Anreicherung &gt; 15 HU</b> | <b>16 (80%)</b>  | <b>13 (100%)</b> | <b>29 (87,8%)</b> |
| <b>Anreicherung &lt; 15 HU</b> | <b>4 (20%)</b>   | <b>0 (0%)</b>    | <b>4 (12,2%)</b>  |
| <b>Summe</b>                   | <b>20 (100%)</b> | <b>13 (100%)</b> | <b>33 (100%)</b>  |

Tab. 1: Vier-Felder-Tafel für das 2D-Verfahren

Von den 13 malignen Herden reicherten bei Auswertung mit der 3D-Methode 12 mehr als 15 HU an. Bei einem Herd konnte bei keiner der vier Messungen nach Kontrastmittelgabe eine Anreicherung über den Schwellenwert hinaus festgestellt werden. Die Sensitivität liegt daher bei 92,3%.

7 der insgesamt 20 benignen Läsionen zeigten keine Anreicherung über den Schwellenwert, was einer Spezifität von 35,0% entspricht.

Von den 25 mehr als 15 HU anreichernden Herden waren 12 maligne (positiv prädiktiver Wert = 48,0%), 7 der insgesamt 8 Herde, die nicht signifikant anreicherten, konnten als benigne diagnostiziert werden (negativ prädiktiver Wert = 87,5%).

Somit konnten durch die Auswertung mit der 3D-Methode insgesamt 19 der 33 Läsionen hinsichtlich ihrer Dignität korrekt diagnostiziert werden. Die Akkuranz liegt somit bei 57,6%.

| <b>3D</b>                      | <b>benigne</b> | <b>maligne</b> | <b>Summe</b> |
|--------------------------------|----------------|----------------|--------------|
| <b>Anreicherung &gt; 15 HU</b> | 13 (65%)       | 12 (92,3%)     | 25 (75,8%)   |
| <b>Anreicherung &lt; 15 HU</b> | 7 (35%)        | 1 (7,7%)       | 8 (24,2%)    |
| <b>Summe</b>                   | 20 (100%)      | 13 (100%)      | 33 (100%)    |

Tab. 2: Vier-Felder-Tafel für das 3D-Verfahren

Eine Gegenüberstellung der Testparameter beider Verfahren kann Tab. 3 entnommen werden.

| <b>Testparameter</b> | <b>2D</b> | <b>3D</b> |
|----------------------|-----------|-----------|
| Sensitivität:        | 100.0%    | 92.3%     |
| Spezifität:          | 20.0%     | 35.0%     |
| Akkuranz:            | 51.5%     | 57.6%     |
| PPV:                 | 44.8%     | 48.0%     |
| NPV:                 | 100.0%    | 87.5%     |

Tab. 3: Testgütekriterien

Bestimmt man nun die Sensitivität und Spezifität nicht nur für den von Swensen vorgeschlagenen Schwellenwert von 15 HU, sondern für alle weiteren denkbaren Schwellenwerte, erhält man eine ROC-Kurve, die Auskunft über die

Qualität des Verfahrens unabhängig von einem bestimmten Schwellenwert gibt.

Die Durchführung der ROC-Analysen zeigte für das 3D-Verfahren ein besseres Resultat, als für das 2D-Verfahren. Vergleicht man die AUC-Werte für die 2D- und die 3D-Messungen, so schneidet das 2D-Verfahren mit 0,523 (Konfidenzintervall: 0,309- 0,737) schlechter ab, als das 3D-Verfahren mit 0,708 (Konfidenzintervall: 0,519 – 0,897).

Die Darstellungen der durchgeführten ROC-Analysen sind den folgenden Abbildungen (7, 8) zu entnehmen.

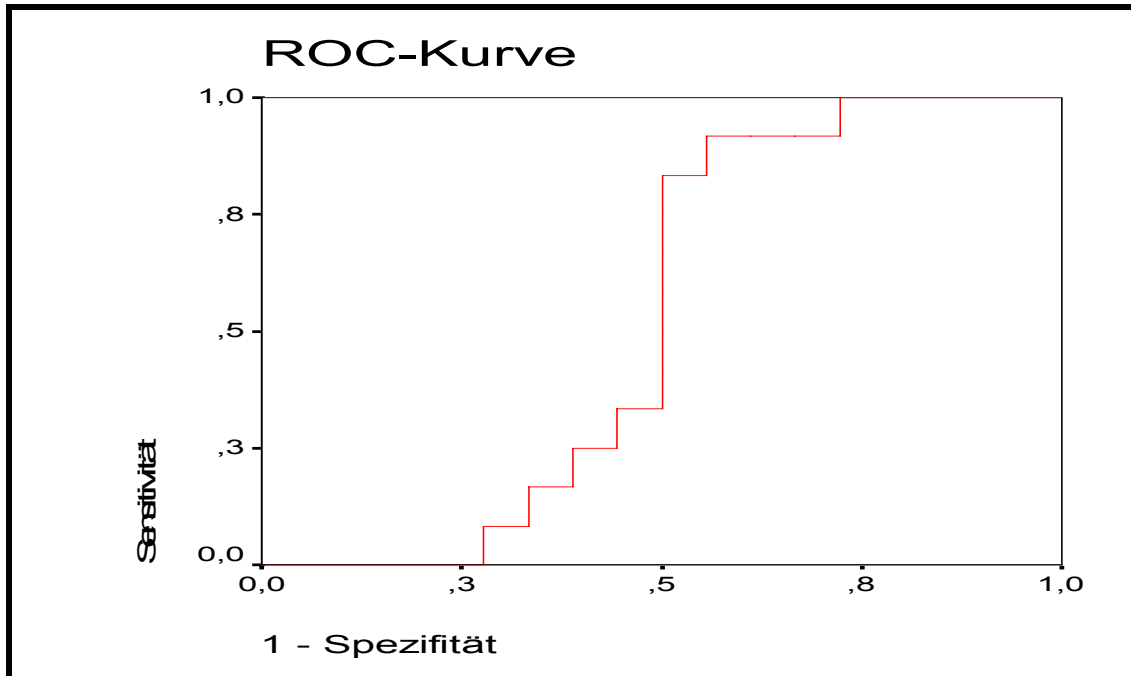


Abb. 7: ROC-Kurve 2D (AUC 0,523)

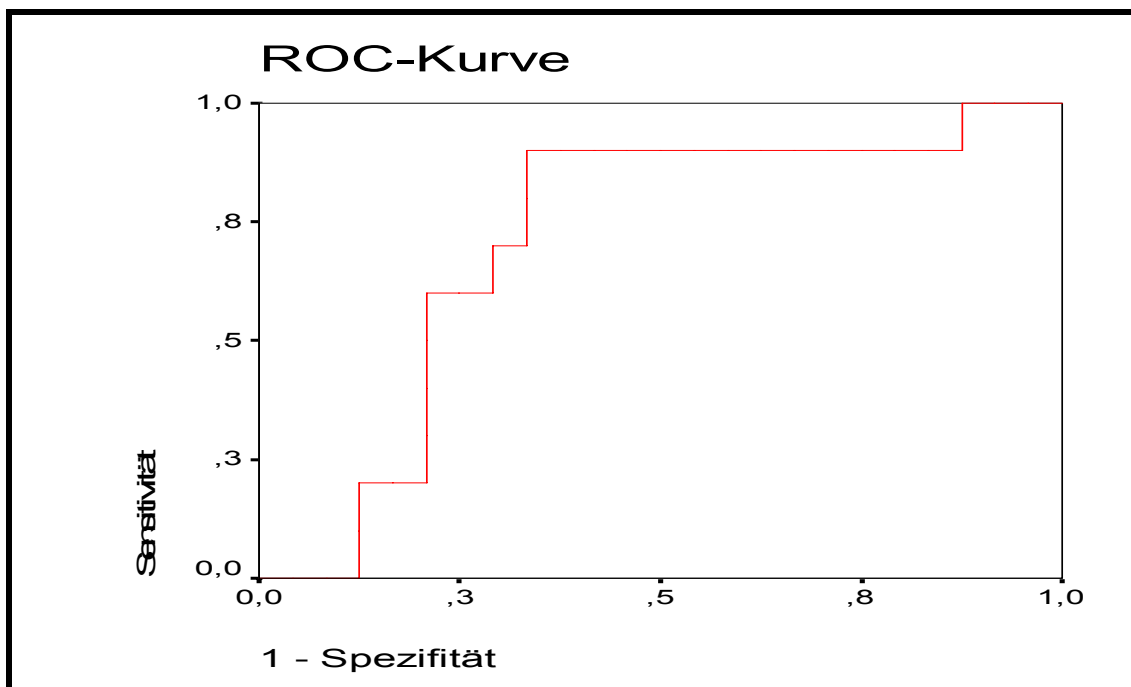


Abb. 8: ROC-Kurve 3D (AUC 0,708)

Deutlichere Unterschiede zeigen sich, wenn man zur Berechnung der Sensitivität und Spezifität nicht die maximalen Anreicherungen aus allen

Messungen, sondern andere Werte zugrunde legt. Die AUC-Werte für alle durchgeführten ROC-Analysen, die beispielhaft mit den 2D-Daten durchgeführt wurden, können Tab. 4 entnommen werden.

| ROC-Analysen |       |
|--------------|-------|
| Grundlage    | AUC   |
| Mittelwert   | 0.556 |
| Maximum      | 0.523 |
| 60           | 0.699 |
| 120          | 0.495 |
| 180          | 0.558 |
| 240          | 0.463 |
| 60+120+180   | 0.576 |

Tab. 4: AUC-Werte für verschiedene Berechnungsgrundlagen

Hier ist zu erkennen, dass die besten AUC-Werte mit der Nutzung allein des 60 Sekunden-Wertes erzielt wurden. Nimmt man stattdessen das Maximum oder den Mittelwert aller Anreicherungen, ist ein deutlicher Abfall der AUC zu erkennen.

Auch die Nutzung der jeweiligen Einzelmessungen 120, 180 und 240 Sekunden nach Kontrastmittelgabe sowie ein Maximalwert aus den ersten 3 Messungen führt zu ähnlich schlechten Ergebnissen.

### 4.3. Kontrastmitteldynamik

Die Analyse, bei welcher die jeweiligen Herde ihre maximale Anreicherung zeigen, erbrachte folgende Ergebnisse: Wenn man alle 51 untersuchten Herde betrachtet, erhält man ein sehr homogenes Bild der Maxima: 25% der Herde erreichten bei der 60-Sekunden-, 25% bei der 120-Sekunden-Messung ihr Dichtemaximum, bei der 180-Sekunden-Messung waren es 27%, bei der 240-Sekunden-Messung 22%.

Wenn man nun jeweils getrennt die gutartigen und die bösartigen Herde untersucht, ergibt sich allerdings ein recht deutlicher Unterschied:

Bei den malignen Herden erreichten 38% ihr Maximum in den ersten 60 Sekunden. Über 60% der Herde benötigten mindestens 3 Minuten für das Erreichen ihres Dichtemaximums. Es gab also 2 Gruppen von Herden, eine sehr schnell anreichernde Gruppe, und eine, die erst nach einer längeren

Verzögerung ihr Enhancement zeigt. Bei den gutartigen Herden war die Verteilung der Anreicherungsmaxima etwas ausgeglichener.

Im Einzelnen ergaben sich die Maxima, die in Abb. 9 dargestellt sind:

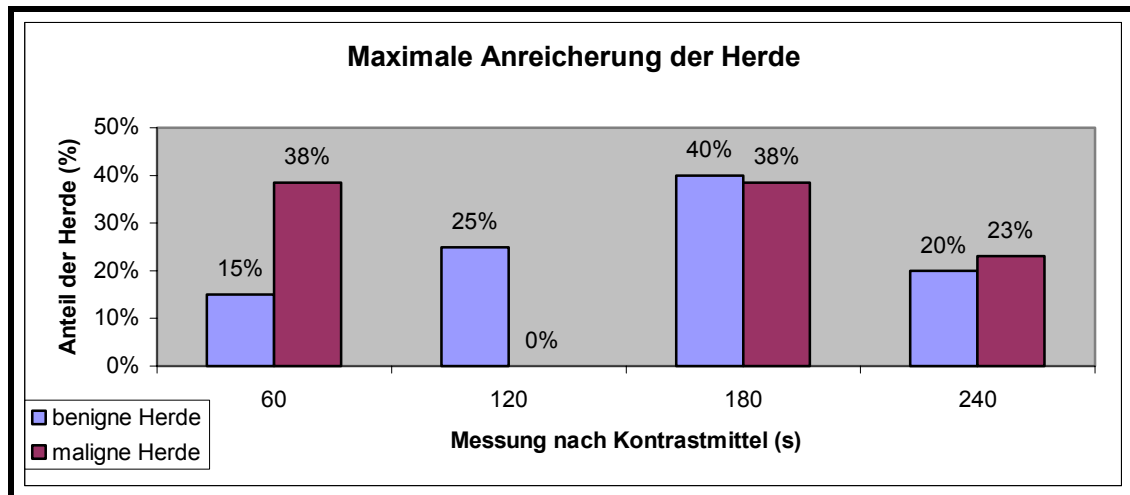


Abb. 9: Anreicherungsmaxima der benignen und malignen Herde (n=33).

Bei den noch nicht sicher diagnostizierten Herden ergab sich ein umgekehrtes Bild: 72% der Herde brauchten 2 Minuten oder weniger, um ihre Maximaldichte zu erreichen (28% bei 60 Sekunden, 44% bei 120 Sekunden). 6% erreichten bei der 180- und 22% bei der 240-Sekundenmessung ihr Maximum.

Da bei dieser Analyse alle untersuchten Herde betrachtet wurden, sind natürlich auch diejenigen noch enthalten, die nicht signifikant (d.h. über den Schwellenwert von 15 HU) angereichert haben und somit nur durch geringe Zufallsschwankungen bei einer der 4 Messungen nach Kontrastmittelgabe ein Dichtemaximum erreichen.

Des Weiteren sind anreichernde Herde enthalten, deren Dichtemaximum offensichtlich durch einen Messfehler zustande gekommen ist, bei denen es beispielsweise nach einer tendenziellen Abnahme der Dichte zu einer weiteren plötzlichen unplausiblen Dichtezunahme gekommen ist.

Da diese Herde das Ergebnis verfälschen können, wurden sie in einem weiteren Auswertungsschritt aus der Statistik entfernt. Übrig blieben 37 Herde, die alle eine deutliche Dichteerhöhung nach Kontrastmittelgabe zeigten und bei denen ein Messfehler nicht offensichtlich vorhanden war.

Bei der Auswertung dieser Herde zeigte sich folgendes Bild: Der Anteil der Herde, die nur 2 Minuten für ihr Dichtemaximum brauchten war auch bei den benignen Herden (50%) höher, als bei den malignen (45%). Es zeigt sich aber auch, dass die 240-Sekunden-Messung an Bedeutung verliert, denn 86% der gutartigen und 91% der bösartigen Herde brauchen nur 3 Minuten oder weniger, um ihren maximalen Dichtewert zu erreichen.

Die graphische Darstellung der bereinigten Anreicherungsmaxima erfolgt in Abb. 10.

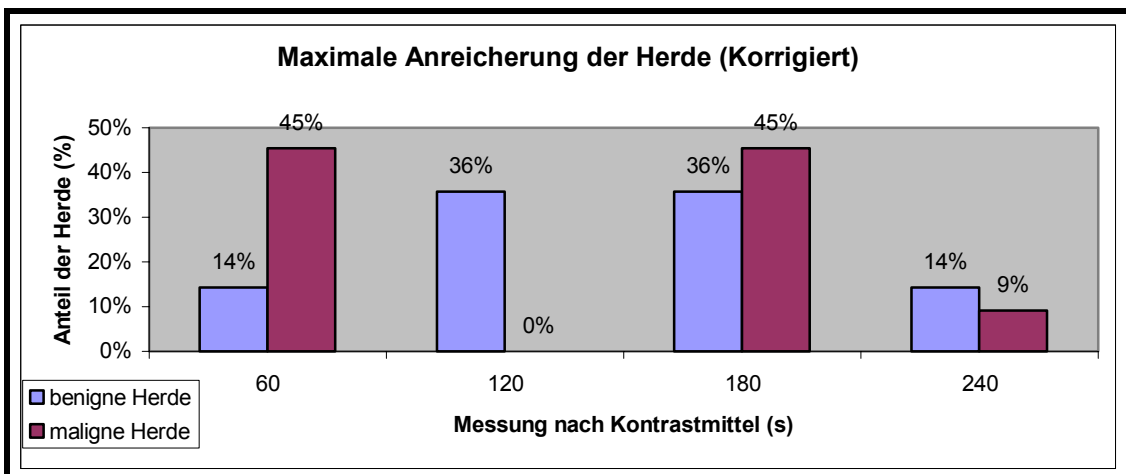


Abb. 10: Anreicherungsmaxima der benignen und malignen Herde nach Abzug der nicht-anreichernden Herde sowie der offensichtlichen Messfehler



#### 4.4. Doppelmessungen

Die Auswertung der Doppelmessungen zeigt erhebliche Unterschiede im Vergleich der 2D- und 3D-Methode.

Die Standardabweichung des Messfehlers, die die Größe des Streuungsbereiches der Werte repräsentiert betrug 42,0 für die 2D-Methode und 14,4 für die 3D-Methode. Der absolute Messfehler, der sich aus den durchschnittlichen Differenzen beider Messungen berechnet, betrug für das 2D-Verfahren 26,4 HU, für das 3D-Verfahren 9,9 HU.

An den graphischen Darstellungen der Doppelmessungen nach Bland und Altman (Abb. 11 und 12) ist zu erkennen, dass die Streuung der Koordinatenpunkte um die Winkelhalbierende bei dem mit der 2D-Methode ermittelten Messwerten erheblich größer ist, als bei denen, die mit der 3D-Methode vermessen wurden. Dies zeigt schon optisch an, dass die Differenzen der Doppelmessungen, und somit der Messfehler, des 2D-Verfahrens über denen des 3D-Verfahrens liegen.

Die Zahl der Ausreißer, also Messdifferenzen, die erheblich von dem üblich zu beobachtenden Ausmaß des Messfehlers abweichen, ist ebenfalls bei der 2D-Auswertung wesentlich höher, als bei der Auswertung mit der 3D-Methode. Bei der 3D-Methode haben nur sehr wenige Koordinatenpunkte einen auffällig weiten Abstand zur Winkelhalbierenden.

Weiterhin fällt auf, dass bei den 3D-ausgewerteten Doppelmessungen 89 von 90 Koordinatenpunkten (dies entspricht 98,9%) im 1. Quadranten des Koordinatensystems liegen, also beide Messungen positive Dichtewerte ergeben haben. Dagegen liegen nur 48 der dargestellten 68 2D-Koordinatenpunkte (dies entspricht 70,6%) im 1. Quadranten. Bei fast 30% lagen einer oder beide Messwerte im negativen Dichtebereich.

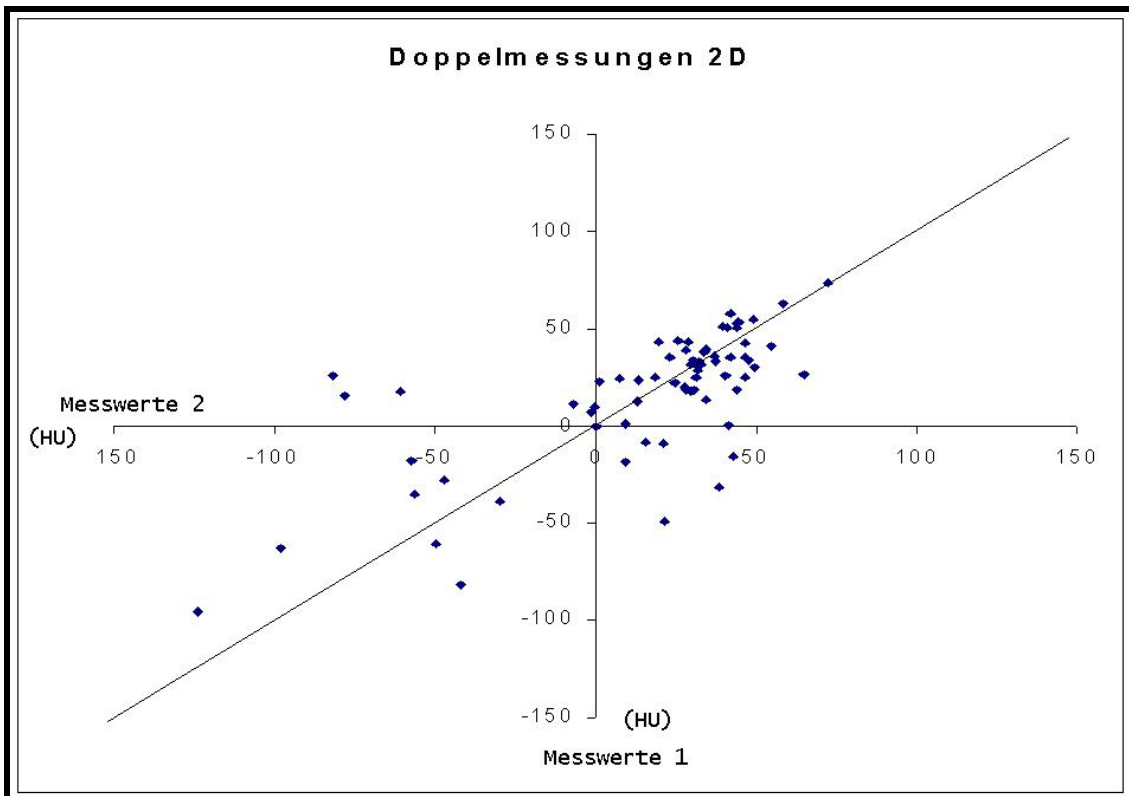


Abb. 11: Grad der Übereinstimmung der Messwerte des 2D-Verfahrens (abgebildet sind 68 der 75 vermessenen Herde, 7 lagen außerhalb des dargestellten Bereiches)

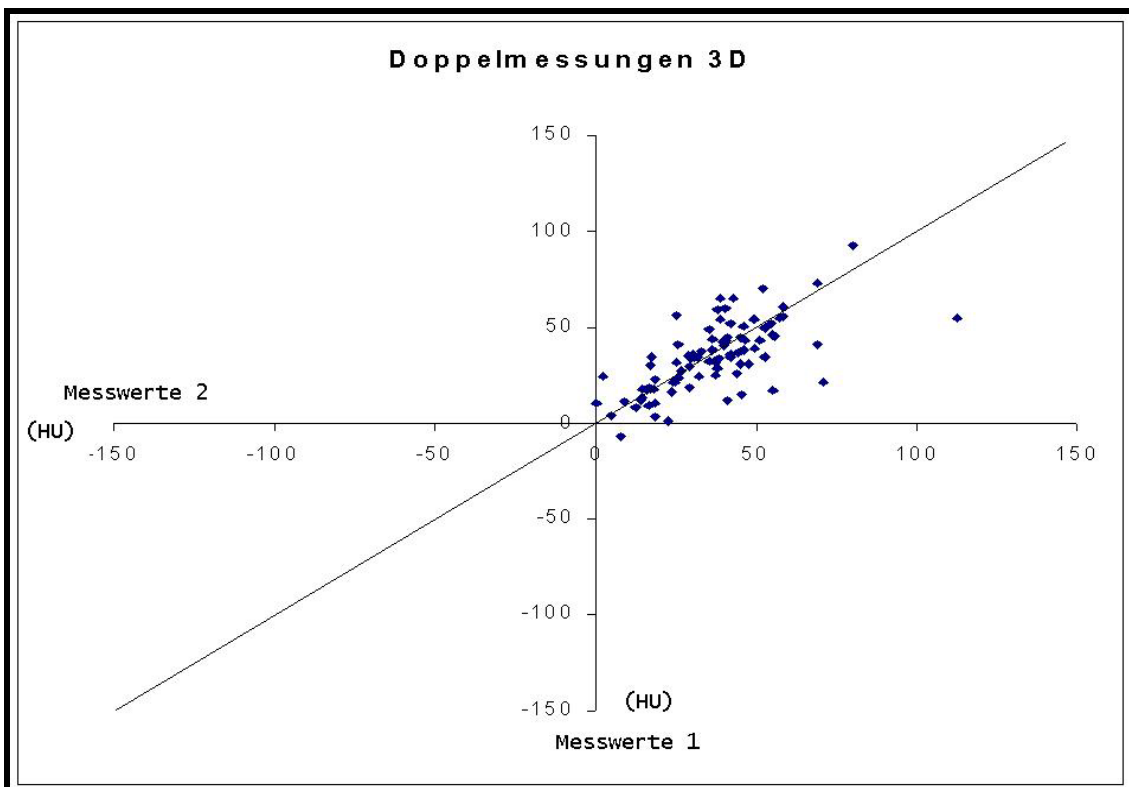


Abb. 12: Grad der Übereinstimmung der Messwerte des 3D-Verfahrens

Quantifiziert werden können die erkennbaren Streuungsunterschiede mit der zweiten Darstellung nach Bland und Altman (Abb. 13 und 14).

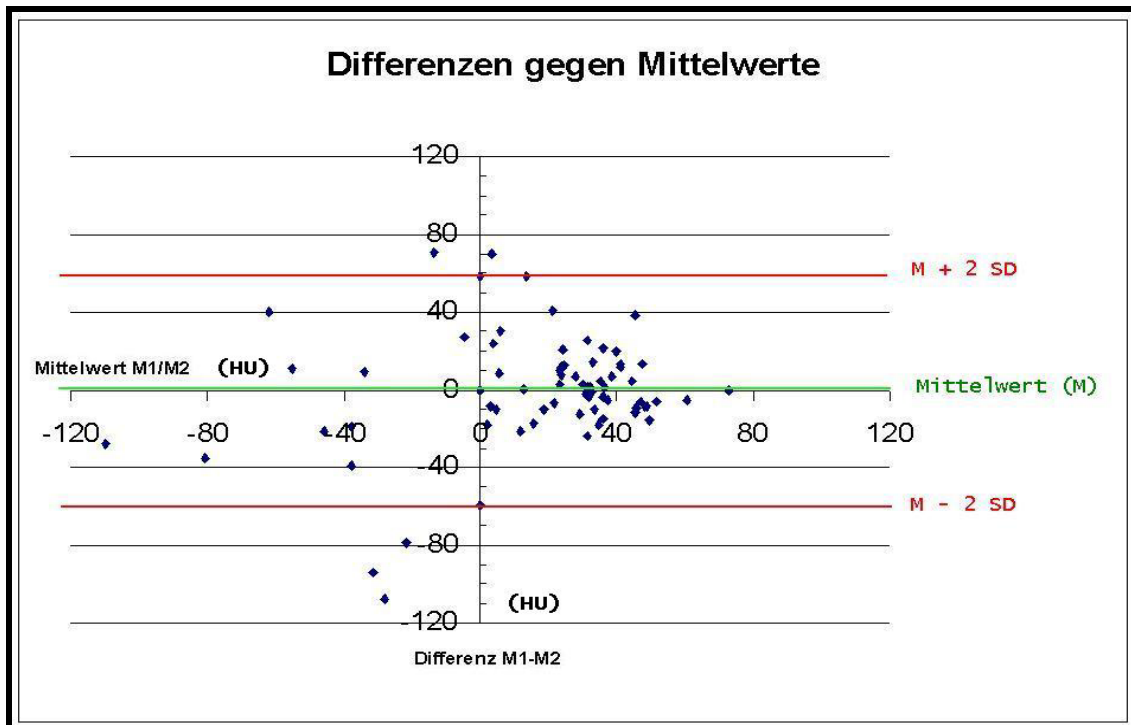


Abb. 13: Darstellung der Differenzen des 2D-Verfahrens (M ist der Mittelwert der Dichte aus beiden Messungen)

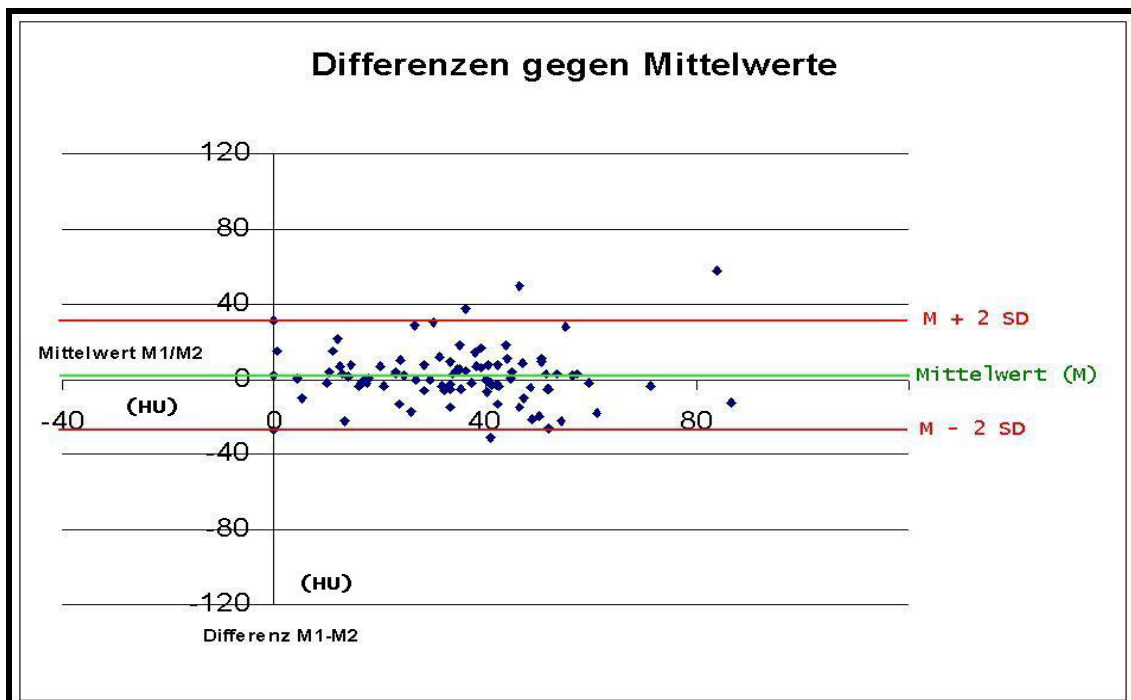


Abb. 14: Darstellung der Differenzen des 3D-Verfahrens (M ist der Mittelwert der Dichte aus beiden Messungen)

Bei dieser zweiten Darstellung fällt die größere Streuung der 2D-Koordinatenpunkte auf. Der Bereich, der durch die zweifache Standardabweichung oberhalb und unterhalb des Mittelwertes aller Differenzen begrenzt wird, ist wesentlich breiter, als bei den 3D-Messungen.

Der Mittelwert aller mit dem 2D-Verfahren bestimmten Differenzen beträgt -0,6 HU. Die Standardabweichung beträgt dabei 29,5. Dementsprechend hat der Bereich der doppelten Standardabweichung um den Mittelwert eine Untergrenze von -59,6 HU und eine Obergrenze von 58,4 HU. In diesem Bereich liegen 62 der dargestellten 68 Koordinatenpunkte (91,2 %).

Der Mittelwert der mit der 3D-Methode ermittelten Differenzen beträgt 2,0 HU. Die Standardabweichung beläuft sich auf 14,4 und ist somit weniger als halb so groß, als bei der 2D-Methode, was für die geringere Streuung der 3D-Werte spricht. Die Bereichsgrenzen liegen hier bei -26,8 HU nach unten und 30,8 HU nach oben. In diesem Bereich liegen 86 der 90 Koordinatenpunkte, was einer Quote von 95,6% entspricht.

| Simuliert | AUC-Wert 2D | AUC-Wert 3D |
|-----------|-------------|-------------|
| 10 HU     | 0.609       | 0.723       |
| 15 HU     | 0.659       | 0.803       |
| 20 HU     | 0.706       | 0.859       |
| 30 HU     | 0.773       | 0.932       |
| 40 HU     | 0.819       | 0.968       |
| 50 HU     | 0.850       | 0.982       |

Tab. 5: ROC-Analysen für die Simulation von verschiedenen Anreicherungen:  
AUC-Werte für 2D- und 3D-Methode

Als letztes wurde die Anzahl der falsch positiven und falsch negativen Testresultate aus den Ergebnissen der Doppelmessungen bestimmt. Die Rate der falsch positiven Testresultate betrug mit der 2D-Methode ausgewertet 20,0%, mit der 3D-Methode ausgewertet 11,1%.

Nach der Enhancement-Simulation und der daraus theoretisch ermittelten Rate der falsch negativen Testresultate wurden die entsprechenden ROC-Kurven ermittelt.

Wenn man den von Swensen vorgeschlagenen Schwellenwert von 15 HU simuliert, ergibt sich eine Area under the curve (AUC) von  $A_z = 0,659$  für das 2D-Verfahren, während mit dem 3D-Verfahren ein AUC-Wert von  $A_z = 0,803$  zu erzielen ist.

Tab. 5 gibt einen Überblick über die zu erzielenden AUC-Werte bei unterschiedlichen Schwellenwerten.

## **5. Diskussion**

### **5.1. Diskussion der Ergebnisse**

#### 5.1.1 Übereinstimmung (Agreement) beider Verfahren

Die erste Aufgabenstellung dieser Arbeit war es zu untersuchen, ob die zu vergleichenden Verfahren annähernd übereinstimmende Messwerte liefern. Diese Frage war insofern wichtig, als daraus abgeleitet werden kann, ob beide Methoden im klinischen Gebrauch austauschbar sind und nebeneinander eingesetzt werden können.

Die graphische Aufarbeitung nach Bland und Altman zeigt, dass von keiner zufrieden stellenden Übereinstimmung gesprochen werden kann. Bereits die erste Graphik (Abb. 5) zeigt zum Teil erhebliche Unterschiede beider Verfahren. Diese Unterschiede können durch die zweite Darstellung (Abb. 6) quantifiziert werden. Die Standardabweichung der Differenzen vom Mittelwert beträgt bereits 29,6. Bedenkt man, dass der von Swensen vorgeschlagene Schwellenwert in einem Bereich von nur 15 HU liegt, dann ist diese Streuung der Differenzen der beiden Messmethoden in einem inakzeptabel hohen Bereich. Bestätigt wird dieses Urteil, wenn man einen Bereich berechnet, in dem 95% aller Wertepaare zu finden ist. Dieser Bereich würde eine Spannweite von über 118 HU umfassen. Aus diesen Zahlen ist zu ersehen, dass von Übereinstimmung der Messwerte nicht gesprochen werden kann und beide Verfahren nicht austauschbar sein können.

Die erstellten Graphiken zeigen allerdings noch einen interessanten Nebenaspekt. In Abb. 5 liegt die Mehrheit der Wertepaare unter der Winkelhalbierenden, was bedeutet, dass mit der 2D-Methode im Mittel höhere Werte für die Anreicherung gemessen wurden, als mit der 3D-Methode. Abb. 6 kann dies quantifizieren. Der Mittelwert der Differenzen  $D = 2D - 3D$  liegt bei 2,6 HU.

Als nächstes bestand die Fragestellung, ob das Ausmaß der Differenz beider Messungen, von der Höhe des jeweiligen Messwertes abhängt. Mit anderen Worten sollte geklärt werden, ob bei sehr großen oder sehr kleinen Messwerten die beiden Methoden dann mehr oder weniger übereinstimmen, oder ob dort kein Zusammenhang besteht. Würde ein solcher Zusammenhang bestehen, würde in Abb. 6 eine Anordnung der Wertepaare zu erkennen sein, die mit zunehmender Höhe des Messwertes (X-Achse) entweder größere oder kleinere Übereinstimmung (Y-Achse) beider Methoden zeigen würde. Eine solche Anordnung ist hier nicht zu erkennen, so dass kein Zusammenhang zur absoluten Messwerthöhe nachgewiesen werden kann.

Auch zu der Frage, welches Verfahren besser für die klinische Anwendung geeignet ist, haben sich einige Aspekte ergeben.

Die Problematik, die im Gegensatz zum 3D-Verfahren bei der 2D-Methode auftritt, war schon während der Auswertung der Daten zu bemerken.

Nach Auswahl eines adäquaten Schnittbildes, welches den Herd in möglichst großer Ausdehnung darstellen sollte, musste die region of interest (ROI) definiert werden. Dabei gab es erhebliche Unterschiede bei den einzelnen Herden. Während es bei großen Herden, die dazu noch relativ rund oder oval waren, kein Problem darstellte die ROI zu definieren, machten andere Herde mehr Schwierigkeiten. Bei sehr kleinen Herden stellte es sich als schwierig heraus, einen Bereich zu finden, der kein Partialvolumen enthielt, um nicht Bereiche in die Dichtemessung mit einzuschließen, die nicht zum Tumor gehörten. Gleiches galt für Prozesse, die eher schmal waren und mit nur kleinem Zentrum und dafür zum Teil langen Ausläufern das Lungenparenchym durchzogen. Weitere Schwierigkeiten traten bei Herden auf, deren Schnittfläche nicht homogen war. Insbesondere wenn das Zentrum des Herdes Verkalkungen oder Nekrosen enthielt, erwies es sich als schwierig, eine optimale Lage für die ROI zu finden.

Die Probleme spiegelten sich in der Dichtemessung wieder. Es war zu erkennen, dass die Schwankungen der Messwerte bei den gut

auszumessenden Herden wesentlich geringer ausfielen, als bei den problematischen Herden. Dort war zu beobachten, dass schon kleine Verschiebungen, Vergrößerungen oder Verkleinerungen der ROI enorme Differenzen der mittleren Dichte verursachten.

Der Messfehler dürfte mit dem 2D-Verfahren bei diesen Herden um ein Vielfaches höher liegen als bei den ersteren.

### 5.1.2 Ergebnisse der Doppelmessungen

Dieser Eindruck bestätigt sich, wenn man die Ergebnisse der Doppelmessungen betrachtet: Die Standardabweichung des Messfehlers war mit 42 HU bei der 2D-Methode etwa dreimal so hoch wie bei der 3D-Auswertung mit 14,4. Auch der absolute Messfehler zeigt mit 26,4 HU (2D) und 9,9 HU (3D) ein ähnliches Verhältnis.

Die Darstellungen nach Bland und Altman (Abb. 11-14) demonstrieren auf graphische Weise die größere Schwankungsbreite der 2D-Messwerte.

Insbesondere bei den kleinen Herden dürfte die Ursache für die Ungenauigkeiten des 2D-Verfahrens zu finden sein.

Da es darum geht, besonders kleine Herde, die schlecht zu biopsieren sind und sich, falls sie bösartig sind, noch in einem frühen Stadium befinden, radiologisch zu bestimmen, ist es für das Verfahren natürlich besonders nachteilig, wenn gerade diese Herde von einem intolerablen Messfehler bei deren Ausmessung betroffen sind.



### 5.1.3. Testgüteparameter

#### 5.1.3.1. Sensitivität

Wie bereits in der Einleitung erwähnt, kommt es bei diesen Verfahren hauptsächlich darauf an, Herde mit hoher Sicherheit als benigne zu erkennen, um diese aus dem weiteren Algorithmus von Biopsie und Resektion herauszunehmen und sie weiter zu beobachten.

Dabei darf allerdings kein maligner Tumor übersehen werden. Aus diesem Grund waren die Sensitivität und somit auch die Rate der falsch negativen Befunde besonders entscheidend für die Beurteilung der Verfahren.

Danach hätte sich hier das 2D-Verfahren besser bewährt, da kein maligner Herd übersehen wurde. Man muss natürlich dabei berücksichtigen, dass hier aufgrund der geringen Zahl sicher diagnostizierter Herde eine Aussage, dass das 2D-Verfahren besser sei als die 3D-Methode, sicherlich unzulässig ist. Es war reiner Zufall, dass unter den wenigen malignen Herden gerade hier einer dabei war, bei dem aus ganz bestimmten Gründen (siehe Kapitel 5.2.1) keine Anreicherung gemessen wurde.

Eine Studie mit größeren Patientenzahlen würde verlässlicher über die Sensitivität beider Verfahren Auskunft geben.

#### 5.1.3.2. Spezifität

Neben der Vermeidung falsch negativer Diagnosen, ist natürlich ebenfalls entscheidend, wie viele gutartige Herde als solche erkannt werden. Wenn nur tatsächlich 20% aller benignen Läsionen (wie hier mit der 2D-Methode ermittelt) aussortiert werden könnten, so wäre das angesichts der großen Zahl von gutartigen Herden die insgesamt unnötig biopsiert werden, zwar immer noch eine diagnostische Hilfe, doch zeigt sich hier, dass die Spezifität noch weitaus höher liegen könnte. In dieser Untersuchung konnten mehr als ein Drittel der benignen Läsionen mit dem 3D-Verfahren als solche identifiziert werden.

Der bei den Doppelmessungen ermittelte geringere Messfehler der 3D-Methode erklärt die bessere Spezifität, die sich hier gezeigt hat. Gutartige Herde, die kein

Kontrastmittel anreicherten, wurden mit dem 3D-Verfahren besser erkannt (Spezifität 35,0%), weil aufgrund der geringeren Schwankungsbreite die Wahrscheinlichkeit höher ist, dass hier auch alle 4 Messungen nach Kontrastmittelgabe nicht fälschlicherweise über den Schwellenwert von 15 HU hinaus ausschlagen und zu einer falsch positiven Diagnose führen. Aufgrund des höheren Messfehlers ist dies bei der 2D-Methode häufiger der Fall (Spezifität 20,0%). Bei der Erkennung gutartiger Herde besteht insbesondere mit dem 2D-Verfahren also noch Verbesserungsbedarf.

Doch dürfte auch klar sein, dass sich die Spezifität nicht beliebig steigern lässt. Keinesfalls können Werte für die Spezifität nahe der 100%-Marke erreicht werden, wie bei der Sensitivität anzustreben ist.

Der Grund dafür liegt darin, dass wie bereits erwähnt auch viele gutartige Prozesse vermehrt perfundiert werden und somit ebenso Kontrastmittel anreichern. In einer Studie von Ohno et al. (74) wurde festgestellt, dass eine aktive Entzündung sogar höhere Anreicherungswerte erreichen kann, als ein maligner Tumor. Selbst bei einer optimalen Messmethode mit nur minimalem Messfehler würde daher nie eine annähernd 100prozentige Spezifität erreicht werden können.

Aus diesem Grund bleibt zu diskutieren, wie man mit Fällen von festgestellter Kontrastmittelanreicherung umgeht. Man könnte grundsätzlich alle anreichernden Herde einer Biopsie unterziehen. Eine Alternative wäre sicherlich, zunächst eine Diagnostik einzuleiten, die das Ziel hat, eine floride Entzündung zu erkennen. Bestimmte Laboruntersuchungen (Blutbild, Akute-Phase-Proteine) wären dafür geeignet. Bei Verdacht auf einen entzündlichen Prozess könnte dann zunächst eine antimikrobielle Therapie durchgeführt werden um nach einiger Zeit die Computertomographie mit der gleichen Methode zu wiederholen. Nach Abklingen des akuten Entzündungsprozesses müsste dann dabei ein Rückgang des Enhancements zu verzeichnen sein. Ist dies nicht der Fall, kann dann kurzfristig immer noch eine Biopsie mit histologischer Untersuchung vorgenommen werden. Dies ist ein Vorgehen, was

bereits im Rahmen der ELCAP-Studien (5, 16, 17, 21, 22, 25-27, 45, 65, 66, 89, 95, 96) vorgeschlagen wurde. In diesen Studien wurden mögliche Screeningverfahren für Lungentumoren bei Risikopopulationen (Raucher) untersucht und geeignete Algorithmen bei der Entdeckung pulmonaler Rundherde vorgeschlagen (siehe Kapitel 5.4.).

## **5.2. Probleme bei der Datenauswertung**

Bei dieser Untersuchung hat sich herausgestellt, dass bei der Auswertung mit dem 3D-Verfahren ein maligner Tumor nicht erkannt wurde. Im Gegensatz zum 2D-Verfahren konnte keine signifikante Anreicherung gemessen werden.

Dieser einzige Fall, der ein falsch negatives Ergebnis lieferte, bietet sich zur näheren Betrachtung an, da hier die Grenzen und Fehlermöglichkeiten des 3D-Verfahrens aufgezeigt werden können, aber auch eine Problematik des 2D-Verfahrens verdeutlicht werden kann.

### **5.2.1. Problemanalyse**

Bei dem beschriebenen Fall handelte es sich um einen 77jährigen Patienten, bei dem ein in die Lunge metastasiertes Adenokarzinom diagnostiziert wurde. Nach der Dichtebestimmung mit dem 2D-Verfahren zeigte sich eine deutliche Kontrastmittelanreicherung von 26,3 HU, was mit dem 3D-Verfahren nicht bestätigt werden konnte. Eine Fehleranalyse zeigte die Ursache für das Versagen des Verfahrens. Zu dem Zeitpunkt, als bei dem Patienten die Nativaufnahme gemacht wurde, konnte auf dieser erkannt werden, dass der betreffende Herd, eine Metastase des Adenokarzinoms, mit einer pleuraständigen Atelektase assoziiert war.

In Abb. 15 (s.u.) ist an der anterioren Thoraxwand der Herd zu sehen, der mit der Pleura verbunden zu sein scheint.

Abb. 16 zeigt allerdings, dass die Verbindung zwischen dem Herd und der Brustwand nicht mehr vorhanden ist. Daher kann es sich nicht um einen pleuraständigen Herd handeln, sondern um eine temporäre Atelektase, die sich in der zweiten Messung geöffnet hat. Dieses kann mit den tiefen Atemexkursionen des Patienten erklärt werden, zu denen er bei dieser Untersuchung aufgefordert wurde.

Nach 120 Sekunden (Abb. 17) sind auch die letzten Reste der Atelektase nicht mehr sichtbar und treten bis zum Ende der Untersuchung (Abb. 18) nicht wieder in Erscheinung

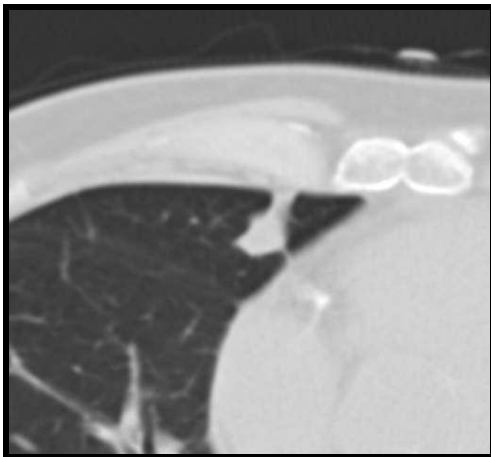


Abb. 15: Nativaufnahme

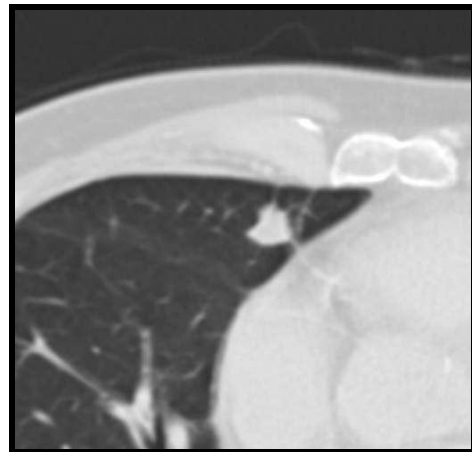


Abb. 16: Aufnahme nach 60 s

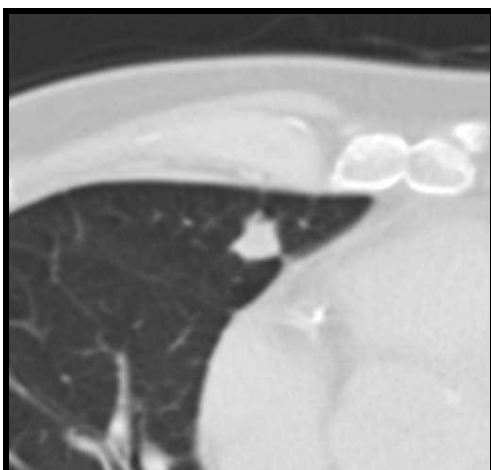


Abb. 17: Aufnahme nach 120 s

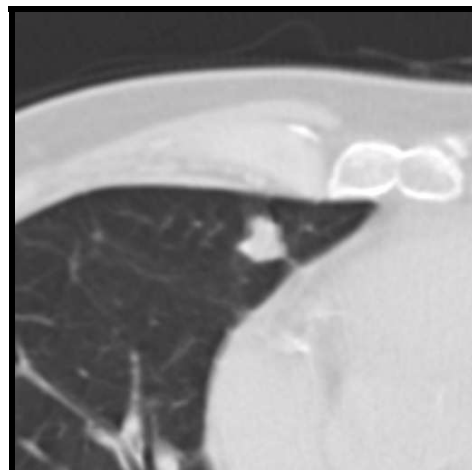


Abb. 18: Aufnahme nach 240 s

Es zeigte sich, dass die Atelektase offensichtlich zu einer falsch hohen Dichtemessung während der Nativaufnahme geführt haben muss, weil das Auswertungsprogramm die Atelektase dem Tumor zugeordnet hat, diese somit mit segmentiert wurde und in die Dichtemessung mit eingeflossen ist. Dadurch reduzierten sich die Differenzen zu den mit Kontrastmittel durchgeführten Aufnahmen, in denen die Atelektase nicht mehr vorhanden war.

Dies führte letztendlich dazu, dass kein relevantes Enhancement berechnet werden konnte.

Als die Messung mit dem 3D-Verfahren manuell so eingestellt wurde, dass nur der innere Kern des Tumors in die Messung mit einbezogen werden sollte, wurde die Nativaufnahme im Vergleich zu den Kontrastmittel-Aufnahmen mit wesentlich geringerer Dichte gemessen, so dass eine berechnete relevante Anreicherung zu einem richtig positiven Ergebnis geführt hätte.

Die gemessenen Dichtewerte können im Überblick der Tab. 6 entnommen werden.

| Messung                  | Dichte 2D (HU) | Dichte 3D vorher <sup>1</sup> (HU) | Dichte 3D nachher <sup>2</sup> (HU) |
|--------------------------|----------------|------------------------------------|-------------------------------------|
| nativ                    | 75.1           | 8.95                               | 24.33                               |
| 60                       | 95.9           | 6.27                               | 40.92                               |
| 120                      | 42.6           | 1.88                               | 22.82                               |
| 180                      | 91.3           | 10.97                              | 13.37                               |
| 240                      | 101.4          | 10.49                              | 30.91                               |
| <b>Max. Anreicherung</b> | <b>26.3</b>    | <b>2.02</b>                        | <b>16.59</b>                        |

Tab. 6: automatisch und manuell bestimmte Dichtewerte des Patienten mit der falsch negativen Diagnose:

<sup>1</sup> vom 3D-Segmentierungsprogramm automatisch bestimmte Dichtewerte

<sup>2</sup> manuelle Messung in den zentralen Anteilen (ca. 650 mm<sup>3</sup>) des Herdes

Dieser Fall zeigt die zwei Grundproblematiken, die bei diesen Verfahren auftreten können.

Das erste Problem gilt nur für das 3D-Verfahren. Dessen Vorteil der Objektivität kann nämlich auch zum Nachteil werden. Während beim 2D-Verfahren die Platzierung der region of interest (ROI) durch den jeweiligen Auswerter erfolgt

und somit schlecht objektivierbar ist, wird beim 3D-Verfahren die auszumessende Region vom Programm berechnet. Das ermöglicht zwar eine höhere Auswertungsobjektivität, kann wie hier zu sehen aber auch zum Problem werden. Mit der 2D-Methode wurde der Prozess korrekt als anreichernd und somit malignitätsverdächtig eingestuft, weil bei der Auswertung der Nativaufnahme die Atelektase als solche erkannt und somit nicht in die ROI mit einbezogen wurde. Da das Segmentierungsprogramm nur mit Dichtewerten rechnet, wurde die Atelektase dem Tumor mit zugeordnet. So entstand der falsch hohe Dichtewert für die Nativaufnahme und im Endeffekt eine falsch negative Diagnose.

Die zweite Problematik betrifft beide Verfahren, kann an diesem Beispiel aber auch verdeutlicht werden. Wie hier gesehen kann ein falsch gemessener Dichtewert in der Nativaufnahme die Diagnose erheblich beeinflussen. Während es vier Messungen nach Kontrastmittelgabe gibt und damit eine falsch niedrige oder hohe Messung durch die anderen drei kompensiert werden könnte, gibt es nur eine native Aufnahme, deren falsche Messung nicht mehr ausgeglichen wird.

Dabei ist ein zu niedrig gemessener Wert weniger problematisch als ein falsch hoher Wert. Ein zu geringer nativer Dichtewert führt zu einer Zunahme der Dichtedifferenz in den nachfolgenden Messungen und somit bei benignen Herden eventuell zu einer falsch positiven Diagnose. Dies würde aufgrund eines falschen Malignitätsverdachtes zu einer Biopsie führen, die allerdings ohne Einsatz des radiologischen Diagnoseverfahrens ohnehin erfolgt wäre.

Ein zu hoher nativer Dichtewert wie in diesem Beispiel hat bei einem malignen Prozess allerdings ungleich größere Konsequenzen. Da die Differenzen zu den Messungen mit Kontrastmittel zu gering werden, wird eine falsch negative Diagnose gestellt. Verlässt man sich auf dieses Verfahren und führt statt einer Biopsie nur weitere Verlaufskontrollen durch, so geht wertvolle Zeit verloren. Ein lokal wachsendes Karzinom, welches durch eine Operation noch potentiell heilbar wäre, könnte im schlimmsten Fall in der Zwischenzeit metastasieren, was zu einer Prognoseverschlechterung für den Patienten führen kann. Ein

Zusammenhang zwischen Tumorgröße und Überleben konnte jedoch zumindest für das Stadium I bislang nicht nachgewiesen werden (75).

### 5.2.2. Lösungsansätze

Es ist nun zu diskutieren, wie man diesen Problemen begegnet.

Das erste Problem kann nur gelöst werden, wenn die Berechnungen des 3D-Segmentierungsprogrammes vom Auswerter nicht kritiklos hingenommen werden. Jede Messung muss daraufhin kontrolliert werden, welche Bereiche vom Programm in die Dichtemessung mit einbezogen wurden. Bezirke, wie in diesem Fall die Atelektase, die offensichtlich nicht zum Tumor gehören, müssen dann manuell ausgeschlossen werden, um fehlerhafte Dichtemessungen und somit falsche Diagnosen zu vermeiden. Die verwendete Software bietet für eine manuelle Nachbearbeitung die Option, cut planes zu setzen und die Größe des Templates zu beeinflussen. Dabei können durch eine manuell eingeführte Linie (cut plane), Bereiche abgegrenzt werden, die nicht mehr in die Segmentierung einbezogen werden sollen.

Zur Kontrolle der Messungen würden sich Enhancement Maps anbieten, die in einer Arbeit von Wormanns und Diederich (105) beschrieben werden.

Bei diesem Verfahren erfolgt eine Subtraktion der Nativaufnahme von der zu untersuchenden Aufnahme nach Kontrastmittelgabe. So kann das Enhancement visuell dargestellt werden.

Beispielsweise aufgrund unterschiedlicher Atemzugtiefen während beider Aufnahmen können dabei zwar Subtraktionsartefakte entstehen, die allerdings durch moderne Workstations weitgehend korrigiert werden können, so dass akzeptable Ergebnisse erzielt werden.

Abb. 19 zeigt ein Beispiel für eine derartige graphische Darstellung. Abgebildet ist ein bronchoalveoläres Karzinom mit einem Durchmesser von 5 mm.

Abb. 19a zeigt das Ergebnis einer Subtraktion der Nativaufnahme von einer Aufnahme, die 60 Sekunden nach Kontrastmittelgabe angefertigt wurde. Das

Enhancement wird visuell sichtbar. Weiterhin ist noch ein Subtraktionsartefakt zu sehen (Pfeil), welches durch die Workstation in Abb. 19b weitgehend ausgeglichen werden konnte.

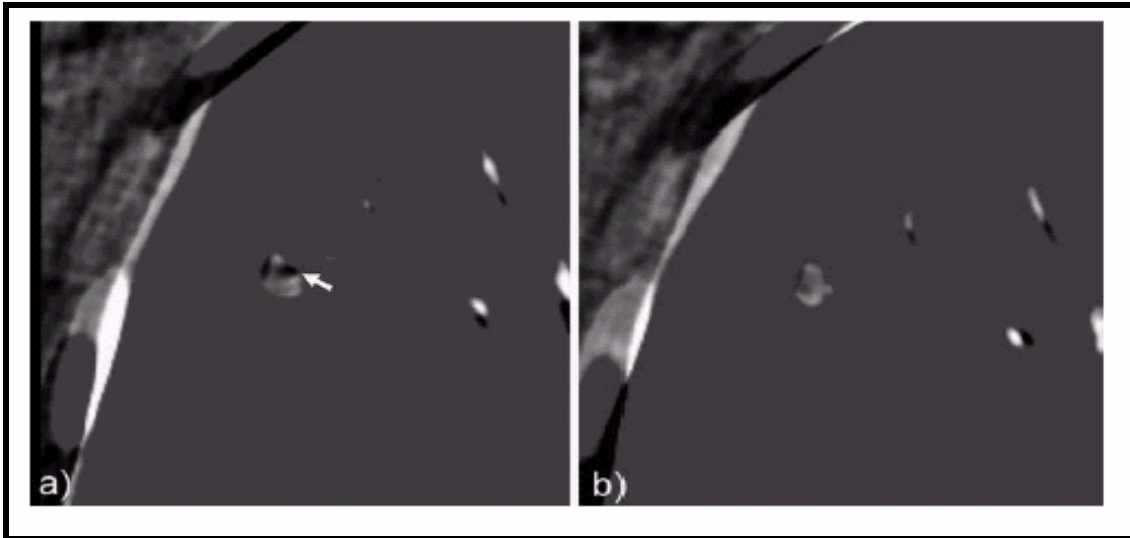


Abb. 19) Enhancement Map aus Wormanns u. Diederich, Eur Radiol 2004 (105)  
a) Enhancement Map mit Subtraktionsartefakt (Pfeil)  
b) Nach Korrektur mit reduziertem Subtraktionsartefakt

Die Nutzung von Enhancement Maps würde die Zuverlässigkeit der Messungen erhöhen, da durch Visualisierung der Anreicherung die Bereiche besser erkannt werden könnten, die fälschlicherweise in die Dichtemessung mit einbezogen wurden, wie es bei der pleuraständigen Atelektase der Fall war.

Zum zweiten Problem der falsch hohen bzw. falsch niedrigen Nativmessung wurde bereits diskutiert, die Nativaufnahme nach dem ersten Scan ein weiteres Mal durchzuführen. So würden sich Messungenauigkeiten der Nativuntersuchung über eine Mittelwertbildung ausgleichen. Bei groben Dichtedifferenzen beider Aufnahmen hätte man einen Hinweis, dass man dem Ergebnis nicht vertrauen kann und weitere Untersuchungen zur Bestimmung der Dignität zu unternehmen sind.

Weitere Studien könnten zeigen, ob man durch Berechnung der Mittelwerte der zweimalig durchgeführten Nativmessung bessere Diagnoseresultate erzielen würde.



### 5.3. Vergleich mit bisherigen Publikationen

#### 5.3.1. Publikationen von Swensen et al.

Grundsätzlich kann man sagen, dass die Ergebnisse dieser Untersuchung mit der Grundaussage von Swensen übereinstimmen.

Nach Auswertung mit dem 2D-Verfahren, wie es Swensen vorgeschlagen hatte, reicherten alle als maligne diagnostizierten Herde mehr als 25 HU an. 11 dieser 12 Herde reicherten sogar über 30 HU an.

Betrachtet man das 3D-Verfahren, so konnte ähnliches beobachtet werden. Die malignen Herde zeigten bis auf eine Ausnahme ebenfalls eine Anreicherung über die 15 HU hinaus.

So konnte Swensens Theorie erneut untermauert werden. Auch wenn zum jetzigen Zeitpunkt die zur Verfügung stehenden Verfahren noch keine ausreichende Messgenauigkeit liefern, um die CT-gestützte Enhancement-Bestimmung vorbehaltlos einzusetzen, so kann man aber festhalten, dass die Grundidee von Swensen, dass fehlende Anreicherung einen malignen Prozess sehr unwahrscheinlich macht, richtig zu sein scheint.

Alternativen bei der Auswertung der Daten konnten anhand dieser Arbeit nicht belegt werden.

Swensen schlug vor, die maximale Anreicherung aus allen vier Messungen nach Kontrastmittelgabe für die Bestimmung der Dignität zu nutzen. Ein überraschendes Ergebnis zeigte sich in diesem Punkt nur bei den verschiedenen ROC-Analysen des 2D-Verfahrens. Mit Benutzung ausschließlich des 60-Sekunden-Wertes zeigte sich mit einer AUC von etwa 0,7 eine wesentlich günstigere ROC-Kurve als das bei Anwendung der Methode, die Swensen vorschlägt der Fall gewesen ist (AUC nur 0,52).

Dieses Ergebnis kann aber nicht ausreichen, um eine andere Vorgehensweise bei der Auswertung zu fordern. Als Begründung kann wieder die geringe Zahl

der Probanden mit bekannter Diagnose genannt werden, deren Daten für die ROC-Analysen genutzt werden konnten.

Wie oben bereits diskutiert kann aber dennoch wieder die Frage gestellt werden, ob hier tatsächlich vier Messungen nach Kontrastmittelgabe sinnvoll sind, um eine signifikante Anreicherung zu erkennen, wie Swensen in seiner Arbeit vorgeschlagen hat.

Man könnte argumentieren, dass mit häufigeren Messungen nur eine Verringerung der Spezifität bewirkt wird. Aufgrund des Messfehlers, der zwar beim 2D-Verfahren um einiges höher ist, jedoch auch beim 3D-Verfahren eine wichtige Rolle spielt, steigt mit jeder zusätzlichen Messung die Wahrscheinlichkeit, dass eine Dichtemessung auch bei nicht-anreichernden Herden über den Schwellenwert von 15 HU hinaus ausschlägt, und somit eine falsch positive Diagnose gestellt wird.

Es müsste also noch weiter untersucht werden, wann die Herde tatsächlich ihre maximale Anreicherung zeigen. Die Tatsache, dass ein Großteil des Enhancements auf die Perfusion und somit auf die ersten 60 Sekunden zurückfällt, lässt vermuten, dass in diesem Bereich auch die meisten der anreichernden Herde ihr Dichtemaximum erreichen. Ein Hinweis darauf könnte die ROC-Analyse des 2D-Verfahrens in dieser Arbeit sein, die für die 60 Sekunden Messung mit Abstand die besten AUC-Werte zeigte.

Ein Indiz dafür, dass die Messung nach 4 Minuten überflüssig sein könnte, ist in den Kurven der Anreicherungsmaxima zu finden. Nachdem alle Maximalwerte, die nur zufällig durch kleinere Schwankungen bei nicht anreichernden Herden entstanden sind, als auch die offensichtlichen Messfehler aus der Auswertung entfernt wurden, zeigte sich, dass nur etwa 9% der malignen Herde und 14 % der benignen Herde ihr Maximum bei der 240-Sekunden-Messung erreichten. Unter diesen Herden dürften immer noch Herde sein, deren Maximum durch einen Messfehler des 2D-Verfahrens verursacht wurde, so dass vermutlich der tatsächliche Anteil noch geringer ist. Daher ist die Hypothese erlaubt, dass zumindest die letzte Messung nach 240 Sekunden keine Relevanz für die Sensitivität besitzt, aber sich negativ auf die Spezifität auswirkt.

Überprüft man diese Hypothese an unseren Daten, so zeigen sich die in Tab. 7 zusammengestellten Ergebnisse:

| <b>Testparameter 2D</b> | <b>mit 240s-Wert</b> | <b>ohne 240s-Wert</b> |
|-------------------------|----------------------|-----------------------|
| Sensitivität:           | 100.0%               | 100.0%                |
| Spezifität:             | 20.0%                | 30.0%                 |
| Akkuranz:               | 51.5%                | 57.6%                 |
| PPV:                    | 44.8%                | 48.2%                 |
| NPV:                    | 100.0%               | 100.0%                |

Tab. 7: Testgütekriterien für das 2D-Verfahren jeweils ermittelt mit und ohne Berücksichtigung der letzten Messung nach 240 Sekunden.

Lässt man die Messung nach 240 Sekunden unberücksichtigt, so führt dies bei den 2D-Daten tatsächlich zu einer Verbesserung der Spezifität, der Akkuranz sowie des PPV, während Sensitivität und NPV unbeeinflusst bleiben. Ursache dafür ist, dass bei zwei benignen Herden in der letzten Messung noch eine Anreicherung von über 15 HU gemessen wurde, was zu falsch positiven Ergebnissen führte. Maligne Herde sind ohne den 240 Sekunden-Wert hingegen nicht unerkannt geblieben, so dass keine Einschränkung der Sensitivität festzustellen war.

Bei den 3D-Daten zeigte sich kein Unterschied bei den beiden Auswertungsvarianten.

Es würde sich lohnen, diesen Effekt, der sich bei unseren Daten bereits andeutete in größeren Studien weiter zu untersuchen.

Es wäre vorstellbar, einen Algorithmus zu entwickeln, in dem nur noch eine Messung nach Kontrastmittel erfolgt. Der Zeitpunkt dieser Messung sollte möglichst nahe an dem Punkt liegen, an dem bei den anreichernden Herden ihre maximale Dichteerhöhung festgestellt wird.

### 5.3.2. Publikation von Ohno und Schäfer et al.

Weitere Hinweise auf die Dignität könnten nicht nur die Anreicherungs-werte liefern, sondern zusätzlich noch die Dynamik der Anreicherung.

Interessant ist demnach die Frage, ob man neben der Tatsache, dass ein Herd signifikant anreichert oder nicht, auch noch die Information nutzen kann, wie er anreichert. Darunter kann man Aussagen fassen, wie steil der Anstieg der Dichte ist, bei welcher Messung der Herd sein Dichtemaximum erreicht und wie der Verlauf der Dichte sich darstellt.

Wie bereits in der Einführung erwähnt führte Ohno et al. (74) mit dem MRT neben der reinen Dichtemessung noch Untersuchungen über die Steilheit des Dichteanstieges, den so genannten Slope, durch. Man kann argumentieren, dass ein Herd, der viel mehr anreichert als ein anderer, zwangsläufig auch einen steileren Dichteanstieg haben muss. Dies gilt unter der Voraussetzung, dass die time to peak, also die Zeit bis zum Erreichen seines Dichtemaximums, identisch oder zumindest ähnlich ist. Daher müsse eine Korrelation zwischen Ausmaß des maximalen Enhancements und dem Slope bestehen.

Auf der anderen Seite scheint es aber dennoch Charakteristika bei den Verläufen der Dichte zu geben, insbesondere wenn man anreichernde gutartige Herde mit bösartigen Herden vergleicht. Schäfer et al. unterscheidet vier charakteristische Verläufe der Dichtekurve (79).

Es gibt Herde, die einen sehr steilen Dichteanstieg nach Kontrastmittelgabe zeigen, die dann allerdings wieder an Dichte verlieren. Dies kommt dadurch zustande, dass das Kontrastmittel wieder aus dem Tumor ausgeschwemmt wird (Wash-Out). Diese Tumoren sind laut Schäfer fast immer bösartig.

Andere zeigen einen nicht ganz so steilen Anstieg. Dann allerdings ist eine Plateauphase zu beobachten, das heißt, die Dichte stagniert auf einem bestimmten Niveau. Von diesen Herden sind noch etwa 67% maligne.

Eine dritte Gruppe von Herden weist einen langsamen, kontinuierlichen Dichteanstieg auf. In nur 7% der Fälle ist hier noch Malignität vorhanden. Unter

der vierten Gruppe werden Prozesse zusammengefasst, die kein Enhancement zeigen. Sie sind als gutartig zu betrachten.

Nach diesen Daten zeigen die bösartigen Tumoren also im Gegensatz zu gutartigen Herden einen Dichteverlauf, der steiler ist. Die Herde erreichen also früher ihr Dichtemaximum.

Als Kritikpunkt an dieser Arbeit kann die Auswahl der benignen Läsionen angeführt werden, die hier untersucht wurde. 17 der 24 benignen Herde waren entweder Hamartome oder Granulome, also Strukturen, die ohnehin kein Enhancement erwarten lassen. Dabei sind die gutartigen Herde für die Differentialdiagnose besonders interessant, die wie maligne Herde ebenfalls ein Enhancement zeigen. Dies wäre beispielsweise bei floriden Entzündungen der Fall, von denen Schäfer in seiner Arbeit 5 untersuchte. Dabei zeigten 2 von ihnen ein Enhancement vom Typ 2, was ja ein Charakteristikum von malignen Herden sein soll.

Die Ergebnisse von Schäfer sind nicht mit denen dieser Arbeit zu vereinbaren. Wenn man versucht, die einzelnen Herde einem bestimmten Anreicherungstyp zuzuordnen, so gab es auch eine Reihe von gutartigen Herden, die dem Anreicherungstyp 1 zuzuordnen waren. Sie zeigten einen steilen Anstieg der Dichte mit darauf folgendem Wash-Out, eine Dynamik, die Schäfer für maligne Herde charakterisiert hatte.

Die Ursachen, warum auch viele gutartige Herde einen steileren Dichteanstieg zeigen als bösartige Herde, dürfte mit in der Natur der Kontrastmittelanreicherung zu finden sein.

Maligne Herde nehmen vermehrt Kontrastmittel auf, weil ihre Perfusion aufgrund der durch den Tumor induzierten Neoangiogenese gesteigert ist. Gutartige Prozesse, die Kontrastmittel anreichern, sind in vielen Fällen floride Entzündungen. Hier hat keine Neoangiogenese stattgefunden, sondern eine durch Entzündungsmediatoren ausgelöste Vasodilatation sowie Permeabilitätserhöhung der Kapillaren im entzündeten Gewebe. Dadurch ist nicht nur die Perfusion des Gewebes erhöht (Vasodilatation), sondern ebenfalls

die Menge Diffusion des Kontrastmittels aus den Kapillaren in den extrakapillären Raum.

Hier kann man nun vermuten, dass die Summe beider Effekte eine schnellere Kontrastmittelanreicherung im entzündeten Gewebe erlaubt, als dies nur durch die vermehrte Gefäßversorgung bei einem malignen Tumor möglich ist. Bei malignen Tumoren ist zwar auch eine Schrankenstörung mit vermehrter Permeabilität zu beobachten, jedoch könnte diese nicht so stark ausgeprägt sein, wie bei einer akuten Entzündung. So wäre dann der steilere Anstieg der Dichtekurve zu erklären.

Warum sich CT- und MRT-Ergebnisse hier widersprechen, kann nicht erklärt werde. Sowohl beim MRT auftretende zusätzliche Effekte (wie der T2-Effekt), als auch die Pharmakokinetik von CT- und MRT-Kontrastmittel können nicht als Begründung für die differierenden Dichtekurven dienen.

Ob diese Unterschiede der Anreicherungszeit so charakteristisch ist, dass sie als Unterscheidungskriterium zwischen beiden Dignitäten dienen kann, kann hier somit nicht mit letzter Sicherheit beantwortet werden. Außerdem empfehlen sich für derartige Aussagen Untersuchungsdaten, die kernspintomographisch gewonnen wurden. Da die Magnetresonanztomographie in der Lage ist, wesentlich mehr Dichtemessungen pro Zeiteinheit durchzuführen, können so bessere Kurvenverläufe der Dichte dargestellt werde, so dass verlässlichere Aussagen gemacht werden können.

#### **5.4. Klinische Anwendungsmöglichkeiten der Verfahren**

Eine Einsatzmöglichkeit für eine derartige Methode wäre beispielsweise im Rahmen einer Screeningmaßnahme für Risikopopulationen gegeben.

Es sind schon eine Reihe von Studien (5, 16, 17, 21, 22, 25-27, 45, 65, 66, 90, 95, 96) durchgeführt worden, die eine Untersuchung darüber zum Ziel hatten, ob es sinnvoll wäre, bei bestimmten Personengruppen ein Screeningverfahren zur Entdeckung von Bronchialkarzinomen zu etablieren.

Als Zielgruppe würden sich hier langjährige Raucher anbieten, da sie bei weitem das größte Risiko tragen, an bösartigen Neoplasien der Atemwege, insbesondere am Bronchialkarzinom, zu erkranken.

Dabei wurden beispielsweise in einer Studie (16) 817 Raucher im Alter von 40 bis 78 Jahren untersucht, die im Schnitt einen Tabakkonsum von 45 pack-years aufwiesen und bei denen keine Lungenrundherde bzw. Symptome, die auf ein Bronchialkarzinom hindeuten könnten, vorher bekannt waren.

Die Untersuchung erfolgte mit einem Low Dose Spiral CT und erbrachte folgende Ergebnisse: In 43% der Fälle (350 der 817 Patienten) wurden nicht-kalzifizierte, teilweise multiple pulmonale Herde gefunden. Am Ende wurde bei 11 dieser Probanden ein bösartiger Tumor, teilweise noch in sehr frühem Stadium, diagnostiziert. Die Prävalenz von Bronchialkarzinomen in einer derartigen Raucherstichprobe betrug somit 1,3%.

Dieses Beispiel ist ein gewichtiges Argument, warum es in Zukunft sinnvoll ist, radiologische Verfahren, wie die in dieser Arbeit untersuchten, in die Tumordiagnostik der Lunge mit einzubeziehen.

Fast die Hälfte der Personen in der untersuchten Stichprobe hatten unklare pulmonale Rundherde. Wird so ein Rundherd klinisch entdeckt, muss dieser natürlich abgeklärt werden. Führt man nun bei allen diesen Patienten eine Biopsie durch, so verursacht das nicht nur einen enormen Kostenaufwand, auch das Risiko für den Patienten ist unverhältnismäßig hoch, gerade wenn man weiß, dass letztendlich nur bei einem Bruchteil der Fälle ein maligner Tumor entdeckt wird.

Gleichzeitig spricht eine Prävalenz von 1,3% für eine Screeningmaßnahme, insbesondere wenn man bedenkt, dass die Inzidenz des Bronchialkarzinoms bei Frauen in diesem Land ansteigend ist (59) und vor allem, dass ein auf diese Art im Frühstadium entdeckter Krebs im Vergleich zu den bereits symptomatisch gewordenen späten Karzinomen eine ausgezeichnete Prognose aufweist (23, 45, 80).

Aus diesem Grunde könnte es sich als überaus sinnvoll erweisen eine Screeningmethode, die zum Ziel hat, pulmonale Rundherde zu erkennen, mit

morphologischen (Verkalkungen) und funktionellen (Vaskularisierung) radiologischen Methoden zu kombinieren, die in der Lage sind, einen Großteil der Herde zuverlässig als gutartig zu erkennen.

Doch nicht nur bei Screeningverfahren, auch bei der zufälligen Entdeckung eines pulmonalen Rundherdes, beispielsweise bei Routine-Röntgenaufnahmen des Thorax beispielsweise vor einer Operation, könnten sich derart zwischengeschaltete radiologische Verfahren als lohnend erweisen und viele Patienten vor einer Biopsie bewahren.

### **5.5. Fazit**

Wenn man zusammenfassend das Ergebnis des Vergleiches der beiden relevanten Verfahren betrachtet, kann man folgendes Fazit ziehen: Insbesondere die Ergebnisse der Doppelmessungen zeigen, dass das 3D-Verfahren einen wesentlich geringeren Messfehler aufweist, so dass seine Ergebnisse reproduzierbarer sind, als die des 2D-Verfahrens.

Aufgrund der geringeren Schwankungsbreite der Messwerte erklärt sich auch die bessere Spezifität, da fälschlich gemessene Ausreißer über den jeweiligen Schwellenwert bei der 3D-Auswertung seltener vorkommen.

Das bei dem durchgeführten Verfahrensvergleich keine deutlicheren Vorteile gegenüber dem 2D-Verfahren in den Testgütekriterien und den ROC-Analysen zu verzeichnen waren, lag an den beschriebenen spezifischen Nachteilen des 3D-Verfahrens, die dazu führen können, dass das 3D-Verfahren anfälliger gegenüber bestimmten Fehlmessungen ist, als das 2D-Verfahren. Das erwähnte Beispiel der falsch negativen Diagnose zeigte die Schwächen dieser Methode. An dieser Stelle kann noch die Frage gestellt werden, ob man durch die Einbeziehung von den Enhancement-Maps, also den graphischen Darstellungen der Anreicherung mit der 3D-Methode, bessere Resultate erzielen kann.



Durch Verbesserung der für das 3D-Verfahren verwendeten Algorithmen kann noch eine Verbesserung der Sensitivität und des negativ prädiktiven Wertes erwartet werden. Wenn diese Parameter garantieren, dass möglichst alle malignen Prozesse zuverlässig erkannt werden können, kann im klinischen Einsatz dieses Verfahren genutzt werden, um bei negativer CT-Dynamik zunächst von einer Biopsie abzusehen und stattdessen in bestimmten zeitlichen Abständen Verlaufskontrollen durchzuführen.

Zur Zeit ist zu empfehlen, dass die 3D-Methode noch ergänzend zur 2D-Methode eingesetzt wird, die zwar eine erheblich höhere Schwankungsbreite der Messwerte aufweist, aber sich insgesamt als das robustere Verfahren erwiesen hat, sowie ggf. zusätzlich Enhancement-Maps erstellt werden.

Beide Methoden zusammen mit anderen radiologischen Verfahren zur Differentialdiagnostik (PET, MRT) sowie der anamnestischen Gesamtsituation (Alter des Patienten sowie seine Raucheranamnese) können in ihrer Gesamtheit eine Aussage ermöglichen, mit welcher Wahrscheinlichkeit sich hinter einem unklaren pulmonalen Rundherd ein maligner Tumor verbirgt.

Bei geringer a priori-Wahrscheinlichkeit eines malignen Tumors im Falle eines negativen Befundes der differentialdiagnostischen Verfahren würde sich eine Biopsie zugunsten von Verlaufskontrollen erübrigen.

## 6. Quellenverzeichnis

1. **American cancer society:** Cancer facts & figures 2001.  
*Atlanta: American cancer society (2001).*
2. **Armato III SG, Giger ML, Blackburn JT, Doi K, MacMaon H.**  
Three-dimensional approach to lung nodule detection in helical CT.  
*Proceedings of SPIE Medical Imaging 1999*; 3661: 553-559
3. **Baba Y, Yamashita Y, Onomichi M, et al.**  
Dynamic magnetic resonance imaging of head and neck lesions.  
*Top Magn Reson Imaging 1999*; 10: 125-129
4. **Bergquist TH, Bailey PB, Cortese DA, et al.**  
Transthoracic needle biopsy: accuracy and complications in relation to location and type of lesion.  
*Mayo Clin Proc 1980*; 55: 475-481
5. **Berlin NI, Buncher CR, Fontana RS, Frost JK, Melamed MR.**  
The National Cancer Institute cooperative early lung cancer detection program: results of the initial screen (prevalence): early lung cancer detection: introduction.  
*Am Rev Resp Dis 1984*; 130: 545-549
6. **Bernard A, the Thorax Group.**  
Resection of pulmonary nodules using video-assisted thoracic surgery.  
*Ann Thorac Surg 1996*; 61: 202-204
7. **Bland JM, Altman DG.**  
Statistical methods for assessing agreement between two methods of clinical measurement.  
*The Lancet i. 1986*: 307-310

8. **Calhoun P, Feldman PS, Armstrong P, et al.**  
The clinical outcome of needle aspiration of the lung when cancer is not diagnosed.  
*Ann Thorac Surg* **1986**; 41: 592-596
  
9. **Conces DJ Jr, Schwenk GR, Doering PR, et al.**  
Thoracic needle biopsy: Improved results utilizing a team approach.  
*Chest* **1987**; 91: 813-816
  
10. **Cotran RS, Kumar V, Collins T, eds.**  
Acute and chronic inflammation.  
*Robbins pathologic basis of disease. 6<sup>th</sup> ed. Philadelphia, Pa: Saunders; 1999*: 85-111
  
11. **Croisille P, Souto M, Cova M, Wood S, Afework Y, Kuhlmann JE, Zerhouni EA.**  
Pulmonary nodules: improved detection with vascular segmentation and extraction with spiral CT.  
*Radiology* **1995**; 197: 397-401
  
12. **Cudkowski L, Armstrong JB.**  
The blood supply of malignant pulmonary neoplasms.  
*Thorax* **1953**; 8: 152-156
  
13. **Dean PB, Niemi P, Kivisaari L, Korman M.**  
Comparative pharmacokinetics of gadolinium DTPA and gadolinium chloride.  
*Invest Radiol* **1988**; 23: S258-S260
  
14. **Dewan NA, Gupta NC, Redepenning LS, Phalen JJ, Frick MP**  
Diagnostic efficiency of PET-FDG imaging in solitary pulmonary nodules.  
*Chest* **1993**; 104: 997-1002

15. **Dewan NA, Reeb SD, Gupta NC, Gobar LS, Scott WJ.**  
 PET-FDG imaging and transthoracic needle lung aspiration biopsy in  
 evaluation of pulmonary lesions: a comparative risk benefit analysis.  
*Chest* **1995**; 108: 441-446
  
16. **Diederich S, Wormanns D, Semik M, Thomas M, Lenzen H, Roos N,  
 Heindel W.**  
 Screening with early lung cancer with low-dose spiral-CT: Prevalence in  
 817 asymptomatic smokers.  
*Radiology* **2002**; 222: 773-781
  
17. **Eddy DM.**  
 Screening for lung cancer.  
*Ann Int Med* **1989**; 111: 232-237
  
18. **Ercolani P, Valeri G, Amici F.**  
 Dynamic MRI of the breast  
*Eur J Radiol* **1998**; 27: 265-271
  
19. **Fan L, Qian J, Odry B, Hong S, Naidich D, Kohl G, Klotz E.**  
 Automatic segmentation of pulmonary nodules by using dynamic 3D  
 cross-correlation for interactive CAD systems.  
*Proceedings of SPIE Medical Imaging* **2002**; 4684: 1362-1369
  
20. **Fish GD, Stanley JH, Miller KS, et al.**  
 Postbiopsy pneumothorax: Estimating the risk by chest radiography and  
 pulmonary function tests.  
*AJR* **1988**; 150: 71-74
  
21. **Flehinger BJ, Kimmel M.**  
 The natural history of lung cancer in a periodically screened population.  
*Biometrics* **1987**; 43: 44-53

22. **Flehinger BJ, Kimmel M, Melamed MR.**  
Natural history of adenocarcinoma-large cell carcinoma of lung:  
conclusions from screening programs in New York and Baltimore.  
*J Natl Cancer Inst* **1988**; 80: 337-444
  
23. **Flehinger BJ, Kimmel M, Melamed MR.**  
The effect of surgical treatment on survival from early lung cancer.  
*Chest* **1992**; 101: 1013-1018
  
24. **Flier JS, Mueckler MM, Usher P, Lodish HF.**  
Elevated levels of glucose transport and transporter messenger RNA are  
induced by ras or src oncogenes.  
*Science* **1987**; 235: 1492-1495
  
25. **Fontana RS, Sanderson DR, Taylor WF, et al.**  
Early lung cancer detection: results of the initial (prevalence) radiologic  
and cytologic screening in the Mayo clinic study.  
*Am Rev Resp Dis* **1984**; 130: 561-565
  
26. **Fontana RS, Sanderson DR, Woolner LB, et al.**  
Screening for lung cancer: a critique of the Mayo Lung Project.  
*Cancer* **1991**; 67 (suppl): 1155-1164
  
27. **Frost JK, Ball WC, Levin M, et al.**  
Early lung cancer detection: results of the initial (prevalence) radiologic  
and cytologic screening in the Johns Hopkins study.  
*Am Rev Resp Dis* **1984**; 130: 549-554
  
28. **Frouge C, Guinebretiere JM, Contesso G, et al.**  
Correlation between contrast enhancement in dynamic magnetic  
resonance imaging of the breast and tumor angiogenesis  
*Invest Radiol* **1994**; 29: 1043-1049

29. **Gaera M, Barone M, Russi EG, et al.**  
Carcinomatous solitary pulmonary nodules: evaluation of the tumor-bronchi relationship with thin-section CT  
*Radiology* **1993**; 187: 535-539
  
30. **Gallagher BM, Fowler JS, Gutterson NI, et al.**  
Metabolic trapping as a principle of radiopharmaceutical design: some factors responsible for the biodistribution of [18F] 2-deoxy-2-fluoro-D-glucose.  
*J Nucl Med* **1978**; 19: 1154-1161
  
31. **Godwin DJ, Speckman JM, Fram EK, et al.**  
CT densitometry: Distinguishing benign from malignant pulmonary nodules.  
*Radiology* **1982**; 144: 349-351
  
32. **Goldsmith SJ , Kostakoglu L.**  
Role of nuclear medicine in the evaluation of the solitary pulmonary nodule.  
*Semin Ultrasound CT MR* **2000**; 21: 129-138
  
33. **Goo JM, Im JG, Do KH, et al.**  
Pulmonary tuberculoma evaluated by means of FDG-PET: findings in 10 cases  
*Radiology* **2000**; 216: 117-121
  
34. **Gupta NC, Aloof J, Gunnel E.**  
Probability of malignancy in solitary pulmonary nodules using fluorine-18-FDG and PET  
*J Nucl Med* **1996**; 37: 943-948

35. **Gupta NC, Frank AR, Dewan NA, et al.**  
Solitary pulmonary nodules: Detection of malignancy with PET with 2-[F-18]-fluoro-2-deoxy-D-glucose.  
*Radiology* **1992**; 184: 441-444
36. **Gückel C, Schnabel K, Deimling M, Steinbrich W.**  
Solitary pulmonary nodules: MR evaluations of enhancement patterns with contrast enhanced dynamic snapshot gradient-echo imaging.  
*Radiology* **1996**; 200: 681-686
37. **Häußinger KE, Kohlhäufel M.**  
Ätiologie und Epidemiologie des Bronchialkarzinoms.  
*MANUAL: Tumoren der Lunge und des Mediastinums, Tumorzentrum München* **2000**.
38. **Hanley JA, McNeil BJ.**  
The meaning and use of the area under a receiver operating characteristic (ROC) curve.  
*Radiology* **1982**; 143: 29-36
39. **Harter LP, Moss AA, Goldberg HI, et al.**  
CT guided fine needle aspirations for diagnosis of benign and malignant disease.  
*AJR* **1983**; 140: 363-367
40. **Hatabu H, Alsop DC, Listerud J, Bonnet M, Gefter WB.**  
T2 and proton density measurement of normal human lung parenchyma using submillisecond echo time gradient echo magnetic resonance imaging.  
*Eur J Radiol* **1999**; 29: 245-252

41. **Hatabu H, Chen Q, Stock KW, Gefer WB, Itoh H.**  
Fast magnetiv resonance imaging of the lung.  
*Eur J Radiol* **1999**; 29: 114-132
  
42. **Hatabu J, Gaa J, Kim D, et al.**  
Pulmonary perfusion and angiography: evaluation with breath-hold enhanced three-dimensional fast imaging steady-state MR imaging with short TR and TE.  
*AJR Am J Roentgenol* **1996**; 167: 653-655
  
43. **Hatabu J, Gaa J, Kim D, et al.**  
Pulmonary perfusion: quantitative assessment with dynamic contrast-enhanced MRI using ultra-short TE and inversion recovery turbo flash.  
*Magn Reson Med* **1996**; 36: 503-508
  
44. **Hatabu H, Tadamura E, Levin DL, et al.**  
Quantitative assessment of pulmonary perfusion with dynamic contrast-enhanced MRI  
*Magn Reson Med* **1999**; 42: 1033-1038
  
45. **Henschke CI, McCauley DI, Yankelevitz DF, Naidich DP, McGuinness G, Miettinen OS, Libby DM, Pasmantier MW, Koizumi J, Altorki NK, Smith JP.**  
Early lung cancer action project: overall design and findings from baseline screening.  
*The Lancet* **1999**; 354: 99-105
  
46. **Hill PC, Spagnolo SV, Hockstein MJ.**  
Pneumothorax with fine needle aspiration of thoracic lesions: Is spirometry a predictor?  
*Chest* **1993**; 104: 1017-1020



47. **Hsu WH, Ikezoe J, Chen CY, et al.**  
Color doppler ultrasound signals of thoracic lesions: correlation with resected histologic specimens.  
*Am J Respir Crit Care Med* **1996**; 153: 1938-1951
48. **Ichikawa T, Araki T.**  
Fast magnetic resonance imaging of liver.  
*Eur J Radiol* **1999**; 29: 186-210
49. **Keagy BA, Starek PJ, Murray GF, Battaglini JW, Lores ME, Wilcox BR.**  
Major pulmonary resection for suspected but unconfirmed malignancy.  
*Ann Thorac Surg* **1984**; 38: 314-316
50. **Knopp MV, Weiss E, Sinn HP, et al.**  
Pathophysiologic basis of contrast enhancement in breast tumors.  
*J Magn Reson Imaging* **1999**; 10: 260-266
51. **Kono M, Adachi S, Kusumoto M, Sakai E.**  
Clinical utility of Gd-DTPA-enhanced magnetic resonance imaging in lung cancer.  
*J Thorac Imaging* **1993**; 8: 18-26
52. **Kormano M, Dean PB.**  
Extravascular contrast material: the major component of contrast enhancement.  
*Radiology* **1976**; 121: 379-382
53. **Kostis W, Reeves A, Yankelevitz DF, Henschke CI.**  
Three-dimensional segmentation of solitary pulmonary nodules from helical CT scans.  
*Computer Assisted Radiology and Surgery* **1999** (CARS 1999): 203-207

54. **Kubota K, Matsuzawa T, Fujiwara T, et al.**  
 Differential diagnosis of solitary pulmonary nodules with positron emission tomography using [<sup>11</sup>C] L-methionine  
*J Comput Assist Tomogr* **1988**; 12: 794-796
55. **Kuhlman JE, Fishman EK, Kuhajda FP, et al.**  
 Solitary bronchoalveolar carcinoma: CT criteria.  
*Radiology* **1988**; 167: 379-382
56. **Kuriyama K, Tateishi R, Doi O, et al.**  
 CT-pathologic correlation in small peripheral lung cancers.  
*AJR* **1987**; 149: 1139-1143
57. **Kuriyama K, Tateishi R, Doi O, et al.**  
 Prevalence of air bronchograms in small peripheral carcinomas of the lung on thin section CT: Comparison with benign tumors.  
*AJR* **1991**; 156: 921-924
58. **Kusumoto M Kono M, Adachi S, et al.**  
 Gadopentetate dimeglumine-enhanced magnetic resonance imaging for lung nodules: differentiation of lung cancer and tuberculoma.  
*Invest radiol* **1994**; 29: 255-256
59. **Levi F, et al.**  
 Trends in mortality from lung cancer in the European Union 1955-1994.  
*Lancet* **1995**;354: 742-743.
60. **Li H, Boiselle PM, Shepard JO, Trotman-Dickenson B, McLoud TC.**  
 Diagnostic accuracy and safety of CT-guided percutaneous needle aspiration biopsy of the lung: comparison of small and large pulmonary nodules.  
*AJR Am J Roentgenol* **1996**; 167: 105-109

61. **Lillington GA.**  
Hazards of transthoracic needle biopsy of the lung.  
*Ann Thorac Surg* **1989**; 48: 163-164
62. **Littleton JT, Durizch ML, Moeller G, Herbert DE.**  
Pulmonary masses: contrast enhancement.  
*Radiology* **1990**; 177: 861-871
63. **Lowe VJ, Fletcher JW, Gobar L, et al.**  
Prospective investigation of positron emission tomography  
in lung nodules.  
*L Clin Oncol* **1998**; 16: 1075-1084
64. **Mack MJ, Hazelrigg SJ, Landreneau RJ, Acuff TE.**  
Thoracoscopy for the diagnosis of the indeterminate solitary pulmonary  
nodule.  
*Ann Thorac Surg* **1993**; 56: 825-830
65. **Melamed MR, Flehinger BJ, Zaman MB, Heelan RT,  
Perchick WA, Martini N.**  
Screening for early lung cancer: results of the Memorial Sloan-Kettering  
study in New York.  
*Chest* **1984**; 86: 44-53
66. **Melamed MR, Flehinger BJ, Zaman MB, Heelan RT, Perchick WA,  
Martini N.**  
Early lung cancer detection: results of the initial (prevalence) radiologic  
and cytologic screening in the Memorial Sloan-Kettering study.  
*Am Rev Resp Dis* **1984**; 130: 555-560

67. **Miller KS, Fish GB, Stanley JH, et al.**  
Prediction of pneumothorax rate in percutaneous needle aspiration of the lung.  
*Chest* **1988**; 93: 742-745
68. **Milne ENC.**  
Circulation of primary and metastatic pulmonary neoplasm: a post-mortem microarteriographic study.  
*Am J Roentgenol Radium Ther Nucl Med* **1967**; 100: 603-619
69. **Milne ENC.**  
Pulmonary metastases: vascular supply and diagnosis.  
*Int J Radiat Oncol Biol Phys* **1976**; 1: 739-742
70. **Milne ENC, Zerhouni EA.**  
Blood supply of pulmonary metastases  
*J Thoracic Imaging* **1987**; 2: 15-23
71. **Mori K, Saitou Y, Tominaga K, et al.**  
Small nodular lesions in the lung periphery : New approach to diagnosis with CT.  
*Radiology* **1990**; 177: 843-849
72. **Murray CJL, Lopez AD, et al.**  
The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020.  
*Cambridge, Mass: Harvard University Press, 1996.*
73. **Ney FG, Feist JH, Altemus LR, Ordinario VR.**  
The characteristic angiographic criteria of malignancy.  
*Radiology* **1972**; 104: 567-570

74. **Ohno Y, Hatabu H, Takenaka D, Adachi S, Kono M, Sugimura K.**  
Solitary pulmonary nodules: Potential role of dynamic MR imaging in management – initial experiment  
*Radiology* **2002**; 224: 503-511
75. **Patz EF Jr., Rossi S, Harpole DH, Herndon JE, Goodman PC.**  
Correlation of Tumor Size and Survival in Patients with Stage IA Non-small Cell Lung Cancer.  
*CHEST* 2000; 117: 1568-1571
76. **Prater HW, Weber WA, Romer W, Truman T, Ziegler SI, Schweigger M.**  
Controlled prospective study of positron emission tomography using the glucose analogue [18F]-fluorodeoxyglucose in the evaluation of pulmonary nodules.  
*Br J Surg* **1998**; 85: 1506-1511
77. **Proto AV, Thomas SR.**  
Pulmonary nodules studied by computed tomography.  
*Radiology* **1985**; 156: 149-153
78. **Reske SN, Kotzerke J.**  
FDG-PET for clinical use. Results of the 3rd German Interdisciplinary Consensus Conference, Onko-PET III, 21 July and 19 September 2000.  
*Eur J Nucl Med* **2001**; 28: 1707-1723
79. **Schäfer JF, Vollmar J, Schick P, Vontheim R, Seemann M, Aebert H, Dierkesmann R, Friedel G, Claussen CD.**  
Solitary pulmonary nodules: Dynamic Contrast-enhanced MR Imaging – Perfusion Differences in Malignant and Benign Lesions.  
*Radiology* **2004**; 232: 544-553

80. **Shah R, Sabanathan S, Richardson J, Mearns AJ, Goulden C.**  
Results of surgical treatment of stage I and II lung cancer  
*The Journal Of Cardiovascular Surgery* **1996**; 37: 169-172
81. **Shopland DR, et al.**  
Toward a tobacco-free society.  
*Semin Oncol* **1990**; 17: 402-412.
82. **Shopland DR, Eyre HJ, Pechacek TF.**  
Smoking attributable cancer mortality in 1991: is lung cancer now the  
leading cause of death among smokers in the United States?  
*J Natl Cancer Inst* **1991**; 83: 1142-1148.
83. **Siegelmann SS, Khouri NF, Scott WW jr., et al.**  
Pulmonary hamartoma: CT findings.  
*Radiology* **1986**; 160: 313-317
84. **Siegelmann SS, Zerhouni EA, Leo FP, Khouri NF, Stitik FP.**  
CT of the solitary pulmonary nodule.  
*AJR Am J Roentgenol* **1980**; 135: 1-13
85. **Sipkins DA, Cheresh DA, Kazemi MR, et al.**  
Detection of tumor angiogenesis in vivo by alpha Vbeta3-targeted  
magnetic resonance imaging.  
*Nat Med* **1998**; 4: 623-626
86. **St Anthony's DRG guidebook 1996.**  
*Reston, Va: St. Anthony's publishing* **1996**
87. **Statistisches Bundesamt Wiesbaden (2004).**  
Todesursachenstatistik 2002.  
Pressemitteilung vom 27. Januar 2004.

88. **Statistisches Bundesamt Wiesbaden (2004).**  
Pressestelle.  
Pressemitteilung vom 07. Juni 2004
89. **Statistisches Bundesamt Wiesbaden (2005).**  
Pressestelle.  
Pressemitteilung vom 17. Januar 2005
90. **Strauss GM, Gleason RE, Sugarbaker DJ.**  
Screening for lung cancer. Another look; a different view.  
*Chest* **1997**; 111: 754-768
91. **Swensen SJ, Brown LR, Colby TV, Weaver AL.**  
Pulmonary nodules: CT evaluation of enhancement with iodinated  
contrast material.  
*Radiology* **1995**; 194: 393-398
92. **Swensen SJ, Brown LR, Colby TV, Weaver AL, Midthun DE.**  
Lung nodule enhancement at CT: prospective findings  
*Radiology* **1996**; 201:447-455
93. **Swensen SJ, Brown LR, Colby TV, Weaver AL, Midthun DE.**  
Lung nodule enhancement at CT: prospective findings.  
*Radiology* **1996**; 201: 447-455
94. **Swensen SJ, Harms GF, Morin RL, Myers JL.**  
CT evaluation of the solitary pulmonary nodules: Value of 185-H  
reference phantom.  
*AJR* **1991**; 156: 925-929

95. **Swensen SJ, Jett JR, Hartman TE, Midthun DE, Sloan JA, Sykes AM, Aughenbaugh GL, Clemens MA.**  
Lung cancer screening with CT: Mayo clinic experience.  
*Radiology* **2003**; 226: 756-761
96. **Swensen SJ, Jett JR, Sloan JA, et al.**  
Screening for lung cancer with low-dose spiral computed tomography.  
*Am J Resp Crit Care Med* **2002**; 165: 508-513
97. **Swensen SJ, Morin RL, Schueler BA, et al.**  
Solitary pulmonary nodule: CT evaluation of enhancement with iodinated contrast media – a preliminary report  
*Radiology* **1992** 182: 343-347
98. **Swensen SJ, Viggiano RW, Midthun DE, et al.**  
Lung nodule enhancement at CT : Multicenter Study  
*Radiology* **2000**; 214: 73-80
99. **Tan BB, Flaherty KR, Kazerooni EA, et al.**  
The solitary pulmonary nodule.  
*Chest* **2003**; 123: p89S-p96S
100. **Viamonte M Jr.**  
Angiographic evaluation of lung neoplasms.  
*Radiol Clin North Am* **1965**; 3: 529-542
101. **Wiemker R, Zwartkuis A.**  
Optimal thresholding for 3D segmentation of pulmonary nodules in high resolution CT.  
*Computer Assisted Radiology and Surgery* **2001** (CARS 2001): 611-616



102. **Winning AJ, Mclvor J, Seed WA, et al.**  
Interpretation of negative results of fine needle aspiration of discrete pulmonary lesions.  
*Thorax* **1986**; 41: 875-879
103. **Wormanns D, Klotz E, Kohl G, Dregger U, Diederich S, Fischbach R, Heindel W.**  
Enhancement measurement of pulmonary nodules with multi-row detector CT – precision assessment of a 3D algorithm compared to the standard procedure.  
*Proc SPIE 5032 Medical Imaging* **2003**: Image processing: 795-801
104. **Wormanns D, Kohl G, Klotz E, Heindel W, Diederich S.**  
Clinical evaluation of the reproducibility of volume measurements of pulmonary nodules  
*Proc. SPIE* **2002**; Vol. 4684: 316 - 322
105. **Wormanns D, Diederich S.**  
Characterization of small pulmonary nodules by CT.  
*Eur Radiol* **2004**; 14: 1380-1391
106. **Yamashita K, Matsunobe S, Tsuda T, et al.**  
Intratumoral necrosis of lung carcinoma : a potential diagnostic pitfall in incremental dynamic computed tomography analysis of solitary pulmonary nodules ?  
*J Thorac Imaging* **1997**; 12: 181-187
107. **Yuan A, Chang DB, Yu CJ, Kuo SH, Luh KT, Yang PC.**  
Color doppler sonography of benign and malignant pulmonary masses.  
*AJR Am J Roentgenol* **1994**; 163: 545-549

108. **Zerhouni EA, Stitik FP, Siegelmann SS, et al.**  
CT of the pulmonary nodule: a cooperative study.  
*Radiology* **1986**; 160: 319-327
  
109. **Zhang M, Kono M.**  
Solitary pulmonary nodules: evaluation of blood flow patterns  
with dynamic CT  
*Radiology* **1997**; 205: 471-478
  
110. **Zhao B, Kostis W, Reeves A, Yankelevitz DF, Henschke CI.**  
Consistent segmentation of repeat CT scans for growth assessment in  
pulmonary nodules.  
*Proceedings of SPIE Medical Imaging* **1999**; 3661: 1012-1018
  
111. **Zhao B, Reeves AP, Yankelevitz DF, Henschke CI.**  
Three-dimensional multicriterion automatic segmentation of pulmonary  
nodules of helical computed tomography images.  
*Optical Engineering* **1999**; 38: 1340-1347

## **7. Danksagung**

An dieser Stelle sei einigen Personen, die in sehr unterschiedlicher Weise am Zustandekommen dieser Arbeit beteiligt waren, herzlich gedankt.

Zunächst geht dieser Dank an Dr. med. Dag Wormanns, nicht nur für die Überlassung des Themas, sondern auch dafür, dass er mir im Verlaufe dieser Arbeit jederzeit mit Rat und Tat zur Seite stand.

Großer Dank richtet sich auch an Dipl.-Phys. Ernst Klotz und seine Mitarbeiter der Firma Siemens Medical Solutions für die freundliche Zusammenarbeit beim Lösen der technischen Probleme.

Herrn Dipl.-Phys. G. Goder vom Institut für Medizinische Informatik und Biomathematik danke ich für die Unterstützung bei der statistischen Auswertung der Daten.

Nicht zuletzt gebührt meiner Freundin und Lebensgefährtin Kathrin Krause sowie meiner Familie besonderer Dank, die mich immer in jeglicher Form unterstützt haben. Meiner Schwester Christina Dregger danke ich für die Bereitstellung von Hard- und Software, mit der große Teile dieser Arbeit erstellt wurde.

## **8. Lebenslauf**

## **9. Anhang**

### **Abkürzungsverzeichnis**

|         |   |
|---------|---|
| Abb.    | Abbildung                                 |
| AUC     | Area Under The Curve                      |
| CT      | Computertomographie                       |
| ELCAP   | Early Lung Cancer Action Project          |
| Gd-DTPA | Gadolinium Diethylen-Triamin-Penta-Acetat |
| HU      | Hounsfield Unit                           |
| KM      | Kontrastmittel                            |
| MPR     | Multi Planar Reformatting                 |
| Mrd.    | Milliarden                                |
| MRT     | Magnetresonanztomographie                 |
| NPV     | Negative Predictive Value                 |
| NSCLC   | Non Small Cell Lung Carcinoma             |
| PET     | Positronenemmissionstomographie           |
| PPV     | Positive Predictive Value                 |
| ROC     | Receiver Operating Characteristic         |
| ROI     | Region Of Interest                        |
| SD      | Standard Deviation                        |
| SPN     | Solitary Pulmonary Nodule                 |
| TAF     | Tumor-Angiogenese-Faktor                  |
| Tab.    | Tabelle                                   |
| z.B.    | zum Beispiel                              |