Diagnostische Genauigkeit der multiparametrischen MRT bei Prostatakarzinom - eine Meta-Analyse

Dissertation
zur Erlangung des Akademischen Grades

doctor medicinae (Dr. med.)

vorgelegt dem Rat der Medizinischen Fakultät der Friedrich-Schiller-Universität Jena

von Christina Vargas Mora

geboren am 17.08.1987 in Plauen

Gutachter

1. **Prof. Dr. Peter Schlattmann, Jena**
2. **Prof. Dr. Ulf Teichgräber, Jena**
3. **Ass. Prof. Dr. Pascal Baltzer, Wien/ÖSTERREICH**

Tag der öffentlichen Verteidigung: 09.08.2021

Inhaltsverzeichnis:

1. Abkürzungsverzeichnis 1

2. Zusammenfassung 2

3. Einleitung 5

3.1 Prostatakarzinomdiagnostik 5

3.2 Die Meta-Analyse 7

3.3 Die multiparametrische MRT (mp-MRT) 8

4. Ziele der Arbeit 11

5. Methodik 12

5.1 Literaturrecherche 12

5.2 Studienauswahl 13

5.3 Datenaufarbeitung und Qualitätsbewertung 14

5.4 Datenanalyse 15

6. Ergebnisse 17

6.1 Studiencharakteristika 17

6.2 Publikationsbias 21

6.3 Diagnostische Genauigkeit 22

6.4 Meta-Regressionsanalyse 28

6.5 Subgruppenanalyse 31

6.6 Qualitätsbewertung 32

7. Diskussion 38

8. Schlussfolgerungen 45

9. Literatur- und Quellenverzeichnis 46

10. Anhang 54

10.1 Ehrenwörtliche Erklärung 54

10.2 Tabellen- und Bildverzeichnis 55

10.3 Protokoll 56

1. Abkürzungsverzeichnis

Abb.-	Abbildung
ADC-	Apparent-Diffusion-Coeffizient
AUC-	Area Under The Curve, Fläche unter der Kurve
KI-	Konfidenzintervall
DCE-	Dynamic Contrast-Enhanced Magnetic Resonance Imaging, dynamische Kontrastmittel-Perfusions-Magnetresonanztomographie
DWI-	Diffusion-Weighted Imaging, diffusionsgewichtete Sequenz der Magnetresonanztomographie
ESUR-	European Society of Urogenital Radiology, Europäische Gesellschaft für urogenitale Radiologie
FN-	falsch Negative
FP-	falsch Positive
mp-MRT-	multiparametrische Magnetresonanztomographie
MRSI-	Magnetic Resonance Spectroscopy Imaging, Magnetresonanzspektroskopie
NLR-	negatives Likelihood-Ratio
OR-	Odds-Ratio
PDRU-	positive digital-rektale Untersuchung
PI-RADS-	Prostate Imaging-Reporting and Data System, strukturiertes Befundungs-schema für die MRT der Prostata
PLR-	positive Likelihood-Ratio
PNB-	Prior Negative Biopsy, frühere Prostatabiopsie mit negativem Ergebnis
PSA-	prostataspezifisches Antigen
QUADAS-	Quality Assessment for Diagnostic Accuracy Studies, Bewertungsbogen zur Qualitätsanalyse von diagnostischen Genauigkeitsstudien
RN-	richtig Negative
RP-	richtig Positive
sROC-	summary Receiver Operating Characteristic Curve, Grenzwertoptimierungskurve
T2-	T2-gewichtete MRT
Tab.-	Tabelle
TRUS-	transrektaler Ultraschall

2. Zusammenfassung

Hintergrund:

Das Prostatakarzinom ist die häufigste Tumorerkrankung des Mannes. Die Früherkennung des Tumors ist wichtig, da die Überlebenschancen in frühen Stadien bei fast 100 % liegen. Hat der Tumor die Kapsel durchbrochen und distale Lymphknoten und benachbarte Organe infiltriert, liegt die Fünfjahresüberlebensrate nur noch bei 30 %. Als Screeningmethode haben sich der PSA- (prostataspezifisches Antigen) Test in Kombination mit der digital rektalen Untersuchung etabliert. Allerdings lassen sich damit nicht alle Tumore rechtzeitig erkennen. Der Goldstandard in der Prostatakarzinomdiagnostik ist zurzeit die ultraschall-gestützte Biopsie bei Patienten mit erhöhtem PSA oder auffälliger digital rektaler Untersuchung. Aber auch der transrektale Ultraschall kann falsch positive Resultate liefern, da sich tumoröse Prozesse, eine Prostatitis oder Infarkte ähnlich darstellen und somit nur schwer voneinander abgrenzbar sind. Auch die Magnetresonanztomographie wird zur gestützten Biopsie eingesetzt. Es gibt Studien, die darauf hindeuten, dass die multiparametrische MRT mit einer besseren diagnostischen Genauigkeit verbunden ist und so helfen kann, unnötige Biopsien zu vermeiden. Bei Patienten mit bereits vorangegangener negativer Biopsie, aber weiterhin bestehendem Tumorverdacht, wird die mp-MRT (multiparametrische Magnetresonanztomografie) aus den Sequenzen T2 (T2-gewichtete MRT-Sequenz), DWI (diffusionsgewichtete Sequenz), DCE (dynamische Kontrastmittel-Perfusions-Sequenz) in der Leitlinie der Europäischen Gesellschaft für Urologie, EAU von 2012 empfohlen. Seitdem sind weitere Studien zur diagnostischen Genauigkeit der MRT bei Prostatakarzinom mit zum Teil sehr variablem Ergebnis erschienen.

Fragestellung und Ziele:

Ziel dieser Arbeit war mit Hilfe einer Meta-Analyse die Primärstudien zur diagnostischen Genauigkeit der multiparametrischen MRT mit den Sequenzen T2, DWI und DCE bei Patienten unter Prostatakarzinomverdacht zusammenzufassen. So sollte die diagnostische Genauigkeit untersucht und eventuelle Einflussgrößen identifiziert werden. Dies kann Anregungen zur Verbesserung der Methode, Qualität und Aussagekraft der Primärstudien liefern. Weiterhin wurden die Ergebnisse mit der aktuellen Studienlage verglichen.

Methodik:

Für die Meta-Analyse wurden bis zum 28. Oktober 2014 systematisch die Online-datenbanken Pubmed, Embase, Isl Web of Knowledge und Scopus auf Primärstudien zu diesem Thema untersucht. Eingeschlossen wurden alle Studien zur diagnostischen Genauigkeit mit der Sequenzkombination T2, DWI und DCE zur Prostatakarzinom-diagnostik verglichen mit der Histopathologie aus systematischen Biopsieproben, publiziert in Englisch, Deutsch oder Spanisch. Um die methodische Qualität der Studien zu bewerten, wurde der Bewertungsbogen zur Qualitätsanalyse von diagnostischen Genauigkeitsstudien QUADAS-2 (Quality Assessment for Diagnostic Accuracy Studies) verwendet. Zwei unabhängige Reviewer beurteilten die Studien bezüglich der Ein- und Ausschlusskriterien, bewerteten die ausgewählten Studien mittels des QUADAS-2 Bogens auf ihre methodische Qualität und entnahmen oder berechneten aus den Primärstudien die zur Aufstellung einer Kontingenztafel notwendigen Daten. Alle Entscheidungen erfolgten im Konsensus. Aus den Kontingenztafeln wurden die studienspezifische Sensitivität und Spezifität, sowie die Gesamt-Sensitivität und Gesamt-Spezifität berechnet. Die Gesamt-Sensitivität und Gesamt-Spezifität wurden graphisch in einer sROC-Kurve (summary Receiver Operating Characteristic) dargestellt und so die Fläche unter der Kurve (AUC) ermittelt. Es wurden Tests auf einen Publikationsbias und auf Heterogenität zwischen den Studien durchgeführt sowie eine Regressions- und Subgruppenanalyse, um eventuelle Einflussgrößen zu ermitteln. Als mögliche Einflussgrößen wurden studienspezifische Charakteristika, wie Prospektivität der Studien, technische Aspekte der MRT, Patientencharakteristika oder Durchführungsunterschiede des Referenzstandards, untersucht.

Ergebnisse und Diskussion:

13 Studien mit insgesamt 1618 Patienten erfüllten die Einschlusskriterien. Die Gesamt-Sensitivität betrug 76 % (95 % KI (Konfidenzintervall), 0,67-0,83) und die Spezifität 81 % (95 % KI, 0,70-0,89). Weiterhin ergaben sich eine AUC von 85 und ein I^2 von 97 %. Die methodische Qualität der Primärstudien war insgesamt hoch. Die Ergebnisse zeigten eine gute Spezifität und eine akzeptable Sensitivität. Eine AUC von 85 ließ eine gute diagnostische Genauigkeit vermuten. Allerdings war die Heterogenität zwischen den Studien sehr hoch und die Streuung von Sensitivität und Spezifität in den einzelnen Studien ebenfalls. Die Studien zeigten Unterschiede bei der Verwendung von Endorektalspulen und der verwendeten Feldstärke.

Schlussfolgerung:

Die Standardisierung technischer Aspekte, eine Vereinheitlichung des PI-RADS-Summenscores sowie das Festlegen eines einheitlichen Grenzwertes, ab welchem PI-RADS-Summenscore die multiparametrische MRT als positiv zu werten ist, sollten erfolgen. Insgesamt scheint die Kombination aus T2, DWI und DCE vorteilhaft bei der gestützten Biopsie für Patienten mit bereits vorangegangener negativer Biopsie und weiterhin erhöhtem PSA-Wert zu sein.

3. Einleitung

3.1 Prostatakarzinomdiagnostik

2018 lag die geschätzte Inzidenz des Prostatakarzinoms in Europa bei 449 761 Männern. Das sind 21,8 % der Krebserkrankungen des Mannes, ausgenommen der nicht melanösen Hauttumore. Damit war das Prostatakarzinom auf Platz 1 der Tumorentitäten des Mannes, gefolgt von Lungenkarzinomen. Die geschätzte Mortalität lag 2018 bei 107 315 Fällen, das sind 10 %, und war damit vergleichsweise niedrig. Die Fünfjahresprävalenz dagegen war mit geschätzten 1 551 420 Fällen sehr hoch (International Agency for Research on Cancer 2018, Ferlay et al. 2018). Meist schreitet die Erkrankung nur langsam voran, was auch die im Vergleich zur Prävalenz niedrige Mortalität erklärt. Es gibt aber auch sehr aggressiv wachsende Tumore. Es ist wichtig, das Prostatakarzinom in einem frühen Stadium zu erkennen, da die Heilungschancen bei Tumoren, die die Kapsel noch nicht durchbrochen haben bei nahezu 100 % liegen. Hat der Tumor aber einmal in distale Lymphknoten und benachbarte Organe gestreut, liegt die Fünfjahresüberlebensrate nur noch bei ca. 49 % (Cancer Research UK 2018).

Mit der digital rektalen Untersuchung kann ein erfahrener Untersucher auf Grund der Oberflächenbeschaffenheit Hinweise auf tumoröse Prozesse bekommen. Allein mit dieser einfachen Untersuchungsmethode können bereits 18 % der Fälle diagnostiziert werden (Heidenreich et al. 2014). Allerdings werden mittels digital-rektaler Untersuchung vor allem ventral liegende Tumore übersehen und erst in einem fortgeschrittenen Stadium entdeckt.
Eine gängige Methode zum Prostatakarzinomscreening ist die Messung des PSA-Wertes im Blut. Dieser Marker hat aber nur eine geringe Spezifität, da er z. B. auch durch Entzündungen erhöht sein kann (Sciarra et al. 2011). Weiterhin schließt ein normaler PSA-Wert einen tumorösen Prozess nicht aus (Barentsz et al. 2012). Eine 2013 durchgeführte Meta-Analyse zeigte, dass der PSA-Test Prostatakarzinome weder in einem früheren Stadium erkennt, noch die Mortalität senkt (Lee et al. 2013). Ein weiteres diagnostisches Mittel ist der TRUS (transrektaler Ultraschall).

In dem TRUS ist das gesunde Drüsengewebe der peripheren Zone der Prostata, in der ca. 70 % der Tumore liegen, echoarm zu sehen. Tumore zeigen ebenfalls ein echoarmes

Signal. Das führt einerseits zu vermehrt falsch negativen Ergebnissen, da sich 30 – 50 % der Tumore isoechogen darstellen (Ghafoori et al. 2013), andererseits aber auch zu falsch positiven Resultaten, da sich eine Prostatitis oder Infarkte ebenfalls echoarm darstellen können (Taneja 2004). Zudem werden mit dem TRUS auch vermehrt Tumore im Zentrum der Drüse übersehen (Ghafoori et al. 2013).

Derzeit ist die beste Methode, um die Diagnose Prostatakarzinom zu sichern, die Biopsie. Allerdings sind auch hier die Ergebnisse häufig falsch negativ, weil die Stanzen aus nicht betroffenen Arealen oder aus oberflächlichen Gewebeschichten entnommen wurden. Weiterhin werden bei der Standardbiopsie mit 10 Stanzen Tumore übersehen, weil die Anzahl der Stanzen besonders für stark vergrößerte Drüsen nicht ausreicht (bis 30 % der Fälle) oder Tumore in der Übergangszone oder im anterioren fibromuskulären Stroma medial vor der Urethra übersehen werden (Haffner et al. 2011). Dies erfordert weitere Biopsien, was unangenehm für den Patienten ist. Weiterhin konnte gezeigt werden, dass die Wiederholungsbiopsien eine schlechtere Tumordetektionsrate aufweisen (Roehl et al. 2002).

Um die Resultate der Biopsie zu verbessern, werden MRT oder TRUS gestützte Biopsien durchgeführt. Die Leitlinie der Europäischen Gesellschaft für Urologie (EAU) empfiehlt die TRUS gestützte Biopsie als Goldstandard (Heidenreich et al. 2014). Allerdings ist auch hier der negative Vorhersagewert nicht hoch genug. Mit MRT und TRUS sollen suspekte Areale detektiert werden, um aus ihnen gezielt zusätzliche Stanzen zu entnehmen. Das ist eine Verbesserung, aber auch aufwendig und kostenintensiv und nach wie vor eine invasive Methode, die für den Patienten mit Schmerzen und einem Risiko für Komplikationen wie Blutungen und Infektionen verbunden ist (Nazir 2014). Eine 2014 veröffentliche Meta-Analyse untersuchte die MRT-gestützte Biopsie. Es konnte gezeigt werden, dass es eine vielversprechende Methode ist, aber dass für eine endgültige Beurteilung weitere Studien und vor allem ein Standard zur einheitlichen Durchführung der Methode nötig sind (Bjurlin et al. 2014).

Die MRT ist ein nicht invasives Schnittbildverfahren, welches sich besonders zur Darstellung von Weichteilgewebe eignet. Damit bietet es sich zur Prostatakarzinomdiagnostik an. Die konventionelle MRT (T1- und T2-gewichtete Sequenzen) zeigte nur eine vergleichbare Sensitivität und Spezifität wie der konventionelle

TRUS und konnte damit, auch auf Grund des höheren Zeit- und Kostenaufwandes und der schlechteren Verfügbarkeit der Methode, nicht überzeugen (Pummer et al. 2013). Aber es gibt einige Studien (Kitajima et al. 2010), die eine verbesserte diagnostische Genauigkeit der konventionellen MRT kombiniert mit mindestens einer zusätzlichen funktionellen Sequenz wie DWI, DCE oder MRSI vermuten lassen, weshalb die multiparametrische MRT (mp-MRI) auch von der ESUR (Europäischen Gesellschaft für urogenitale Radiologie) bei persistierendem Tumorverdacht nach einer oder mehr vorangegangenen negativen Biopsien empfohlen wird (Barentsz et al. 2012). Eine Meta-Analyse zur diagnostischen Genauigkeit von T2, DWI und DCE, die sieben Studien eingeschlossen hatte, zeigte eine hohe Spezifität (de Rooij et al. 2014). Allerdings zeichnet sich eine gute Methode auch durch eine hohe Sensitivität mit einer möglichst kleinen Rate an falsch negativen Ergebnissen aus. Die mp-MRT wurde in dieser Hinsicht bisher, auch auf Grund der geringen Studienanzahl, nur unzureichend untersucht. Da seither neue Studien zur mp-MRT veröffentlicht wurden, soll die diagnostische Genauigkeit mittels einer erneuten Meta-Analyse reevaluiert werden.

3.2 Die Meta-Analyse

Eine Meta-Analyse ist eine sekundäre Analyse. Sie gibt die Möglichkeit, die Analyseergebnisse primärer Studien zusammen zu fassen und zu reanalysieren, z. B. mit besseren statistischen Mitteln oder, um mit alten Daten neue Fragestellungen zu beantworten. Ursprünglich für Therapiestudien entwickelt (Van Houwelingen et al. 1993), wurden die Methoden der Meta-Analyse auch für diagnostische Genauigkeitsstudien angepasst (Vamvakas 2001). Sie wird genutzt, um aussagekräftige mittlere Schätzwerte zur diagnostischen Genauigkeit von Effektgrößen zu berechnen (Vamvakas 2001). Im Fall der diagnostischen Genauigkeitsstudien sind die Effektgrößen, die Sensitivität und Spezifität, mit der ein Test (in der vorliegenden Arbeit die multiparametrische MRT) eine Erkrankung (hier das Prostatakarzinom) diagnostiziert. Die Meta-Analyse bietet den Vorteil, dass man viele Einzelergebnisse zusammenfassen und so bessere Aussagen auch bei einer widersprüchlichen Datenlage treffen kann (Glass 1976). Dabei werden die Aussagen durch die aus den Einzelstudien zusammengefassten und damit größeren Patientenzahlen zuverlässiger. Zusätzlich können mittels einer Meta-Analyse Aussagen über Faktoren getroffen werden, die die Sensitivität und Spezifität beeinflussen und deren Variabilität zwischen den Studien erklären (Vamvakas 1998). So können mit ihrer Hilfe auch Schwachstellen in der Methodik und Qualität der Primärstudien aufgedeckt werden, was zu einer Verbesserung dieser in der Zukunft beiträgt (Vamvakas 1998).

Limitierend für eine Meta-Analyse sind eine mitunter schlechte methodische Qualität der Primärstudien und ein möglicher Publikationsbias (Vamvakas 1998). Aus diesem Grund wird die Qualität der Primärstudien gesondert untersucht, in dieser Analyse mit dem QUADAS-2 (Whiting et al. 2011) Bewertungsbogen und es wird auf einen Publikationsbias getestet.

In der hier vorliegenden Arbeit bietet die Meta-Analyse die Möglichkeit, die wenigen Einzelstudien mit zum Teil kleinen Patientenzahlen und sehr variablen Ergebnissen zur diagnostischen Genauigkeit der mp-MRT bei Prostatakarzinom zusammen zu fassen und so eine bessere Aussage über den Nutzen der Methode und eventuelle Einflussgrößen zu treffen.

3.3 Die multiparametrische-MRT (mp-MRT)

Die mp-MRT ist zur Zeit erst nach erfolgter negativer Biopsie, aber weiter bestehendem Tumorverdacht, zum Staging oder nach erfolgter Therapie mit erneutem PSA-Anstieg empfohlen (Barentsz et al. 2012). Die mp-MRT ist eine Kombination aus morphologischen und funktionellen Sequenzen, demzufolge aus einer hochauflösenden T2-gewichteten Sequenz und mindestens zwei funktionellen Sequenzen wie der DWI, DCE oder MRSI (Magnetresonanzspektroskopie) (Barentsz et al. 2012). Diese Kombination erlaubt eine bessere Beurteilung als die Kombination mit nur einer funktionellen Sequenz (Kitajima et al. 2010). Die T2-gewichtete Sequenz dient der anatomischen Darstellung der Prostata. Die DWI und MRSI erhöhen die Spezifität bei der Beurteilung der Läsionen (Villeirs et al. 2010) und die DCE erhöht die Sensitivität (Yoshizako et al. 2008). Die T2-gewichtete Sequenz wird in zwei bis drei Ebenen durchgeführt. Eine Endorektalspule ist unabhängig von der Feldstärke nicht zwingend erforderlich, jedoch eine Oberflächenspule. Tumore weisen in dieser Sequenz eine geringe Signalstärke auf. Die Methode ist sensitiv, aber nicht spezifisch, da sich z. B. eine Prostatitis oder eine Hämorrhagie genauso darstellen können. Um Fehlinterpretationen auf Grund von Hämorrhagien zu vermieden, sollen zwischen einer Biopsie und der MRT 6 Wochen vergehen oder zusätzlich ein T1-gewichtete Sequenz durchgeführt werden. Weiterhin ist die Diagnostik von Tumoren in der Übergangszone schwierig, da sie einer benignen Prostatahyperplasie ähneln können (Barentsz et al. 2012). Die DCE ist eine dynamische Kontrastmittel-Perfusions-MRT und untersucht die Vaskularisation der Prostata. Die Auswertung kann auf drei Weisen erfolgen: qualitativ,

semi-quantitativ oder quantitativ. Auch die normale Prostata ist gut vaskularisiert, deshalb müssen Aufnahmen während und nach der intravenösen Boluskontrastmittelgabe erfolgen.

Das genutzte Kontrastmittel ist auf Gadoliniumbasis. Die Methode ist hoch sensitiv, aber wenig spezifisch (Barentsz et al. 2012). Die DWI misst den Diffusionskoeffizient, ein Maß für die Stärke der Diffusion der Wasserstoffprotonen (Brown'sche Molekularbewegung) (Schwarzmüller-Erber und Silberstein 2010), und ermöglicht eine qualitative und quantitative Beurteilung des Tumors. Letztere erfolgt mit Hilfe des ADC (Apparent-Diffusion-Coeffizient), wobei sich Tumore durch niedrigere ADC-Werte als normales Prostatagewebe auszeichnen. Diese korrelieren auch gut mit dem Gleason-Score, sodass sich mit ihrer Hilfe Einschätzungen zur Tumoraggressivität treffen lassen können. Zur qualitativen Beurteilung wird der b-Wert herangezogen. Dieser gibt die Empfindlichkeit der Messung auf Diffusion an (Schwarzmüller-Erber und Silberstein 2010). Tumore fallen dabei mit einer hohen Signalstärke bei hohen b-Werten (800-1000) auf (Barentsz et al. 2012). Die MRSI wird eher zur Diagnostik bei Verdacht auf ein rekurrentes Karzinom oder zur Therapiekontrolle genutzt. Mittels MRSI können auch Aussagen zur Tumoraggressivität getroffen, aber nur wenig Informationen zur Morphologie bereitgestellt werden. Bei der MRSI erfolgt eine quantitative Betrachtung des Gewebes. Dazu sind drei Metabolite notwendig. Citrat für gesundes Gewebe, Kreatinin, was nur schwer von Cholin zu unterscheiden ist und Cholin als Marker für tumoröses Gewebe. Die Tumore fallen durch ein niedriges Citrat und ein hohes Cholin auf. Der Quotient wird wie folgt gebildet: $\frac{Citrat+Kreatinin}{Cholin}$. Dieser wird als positiv gewertet, wenn er in mindestens zwei Voxeln (Bildpunkten) zwei und drei Standardabweichungen über dem Mittelwert liegt. Nachteilig ist die Zeitintensivität der Methode. Außerdem ist bei 1,5 Tesla zwingend eine Endorektalspule notwendig. Das ist unangenehm für den Patienten und mit Mehrkosten verbunden (Barentsz et al. 2012). Ob es von Nutzen ist, die MRSI als 4. Sequenz zur mp-MRT hinzuzufügen oder ob es mit einem Mehrwert verbunden ist, die DWI oder DCE in der mp-MRT durch die MRSI zu ersetzen, ist noch unklar auf Grund der noch recht kleinen Studienlage.

Betrachtet man die technischen Anforderungen an die mp-MRT, zeigen 3 Tesla-Geräte bessere Auflösungen und zeichnen sich durch eine kürzere Aufnahmezeit aus, aber auch mehr Artefakte (Barentsz et al. 2012). Bei 1,5 Tesla hat sich die Kombination von Oberflächenspule mit einer Endorektalspule bewährt (Ghafoori et al. 2013). Da aber Endorektalspulen mit höheren Kosten und Unannehmlichkeiten für den Patienten

verbunden sind, werden sie nicht standardmäßig eingesetzt (Barentsz et al. 2012). Bei 3 Tesla-Geräten kann auf eine Endorektalspule verzichtet werden, bzw. sind die handelsüblichen Spulen bei dieser hohen Feldstärke noch nicht einsetzbar (Ghafoori et al. 2013).

Im Folgenden soll die Kombination aus T2+DWI+DCE untersucht werden. Außerdem wurde versucht, Rückschlüsse auf den Einfluss von z. B. der Feldstärke und der Benutzung einer Endorektalspule auf die diagnostische Genauigkeit zu ziehen.

4. Ziele der Arbeit

Diese Arbeit untersucht mittels einer Meta-Analyse die diagnostische Genauigkeit der multiparametrischen MRT, als nicht invasive Methode, in der Diagnostik des Prostatakarzinoms. Dabei soll die Testgenauigkeit der Sequenzkombination aus T2, DWI und DCE bei Patienten mit Verdacht auf ein Prostatakarzinom auf Grund von einem erhöhten PSA-Wert und/oder einer positiven digital rektalen Untersuchung mit dem Referenzstandard verglichen werden. Die Histopathologie, der in der systematischen ultraschallgestützten Biopsie gewonnen Stanzen, war der Referenzstandard. Es sollen die Heterogenität zwischen den Studien untersucht werden und mittels einer Meta-Regression mögliche Ursachen für die Heterogenität im Studiendesign, den Patientencharakteristika, den Charakteristika der mp-MRT oder des Referenzstandards identifiziert werden. Weiterhin sollen mit Hilfe einer Meta-Regression und einer Subgruppenanalyse Hinweise auf mögliche Einflussgrößen, wie z. B. der Feldstärke oder der Nutzung einer Endorektalspule und des PI-RADS-Scoring-Systems auf die Sensitivität und Spezifität der MRT gewonnen werden.

Die Ergebnisse sind relevant, da multiple Biopsien für den Patienten unangenehm und mit Komplikationen verbunden sein können. Aus diesem Grund könnten MRT-gestützte Biopsien in Zukunft vermehrt eingesetzt werden, um unnötige Biopsien zu vermeiden. Das hätte weiterhin positive ökonomische Auswirkungen.

5. Methodik

Die Durchführung der Meta-Analyse erfolgte gemäß den Anforderungen der PRISMA Checkliste (Moher et al. 2009). Es wurde ein Protokoll erstellt und eine systematische Literaturrecherche durchgeführt.

5.1 Literaturrecherche

Für die Recherche wurden bis zum 28. Oktober 2014 die Datenbanken „PubMed“, „Embase“, „Isl Web of Knowledge“ und „Scopus“ systematisch nach Artikeln, die sich mit der diagnostischen Genauigkeit der mp-MRT bei Prostatakarzinom beschäftigen, durchsucht. Eingeschlossen wurden Artikel, die in den Sprachen Deutsch, Englisch und Spanisch publiziert wurden. Seit 2014 verlagerte sich das wissenschaftliche Interesse eher auf den Nutzen der mp-MRT zur Identifizierung von klinisch signifikanten tumorösen Prozessen und der Einschätzung der Tumoraggressivität mittels mp-MRT, sodass zur hier untersuchten Fragestellung auch bei erneuter Datenbankrecherche keine neuen relevanten Ergebnisse zu erwarten sind.

Die Suchanfrage für „PubMed“ lautete wie folgt:

„(("prostate cancer"[All Fields] OR "prostatic disease"[All Fields] OR "prostatic neoplasm"[All Fields] OR "prostatic hyperplasia"[All Fields] OR "prostatic lesion"[All Fields] OR "prostatic carcinoma"[All Fields] OR "prostatic diseases"[MeSH Terms] OR "prostatic neoplasms"[MeSH Terms] OR "prostatic hyperplasia"[MeSH Terms]) AND ("magnetic resonance imaging"[All Fields] OR "magnetic resonance"[All Fields] OR "diffusion weighted"[All Fields] OR "dynamic contrast enhancement"[All Fields] OR "magnetic resonance spectroscopic imaging"[All Fields] OR MRSI[All Fields] OR DCE-MRI[All Fields] OR DW-MRI[All Fields] OR "T2-weighted"[All Fields] OR "multiparametric magnetic resonance imaging"[All Fields] OR "magnetic resonance imaging/methods"[MeSH Terms])) AND ("sensitivity and specificity"[MeSH Terms] OR "sensitivity"[All Fields] OR

"specificity"[All Fields] OR "diagnostic accuracy"[All Fields] OR "contingency table"[All Fields])”

Trotz intensiver Bemühungen gelang ein weiteres Filtern der Ergebnisse über die Suchanfrage nicht. Alle Ergebnisse aus den Datenbanken „Pubmed“, „Embase“ und „Isl Web of Science“ wurden in die bibliografische Datenbank „EndNote X7“ importiert. Die Ergebnisse aus Scopus wurden manuell erfasst und relevante Artikel ausgedruckt. Der Zugang zu Scopus wurde freundlicherweise vom Max-Planck-Institut für Menschheitsgeschichte Jena gewährt. Es wurden nicht nur Primärstudien gesichtet, sondern auch allgemeine und systematische Übersichtsarbeiten.

5.2 Studienauswahl

Gemäß den PRISMA Empfehlungen wurde die Suche von 2 unabhängigen Personen durchgeführt. Der Autor (C.B.) durchsuchte die Datenbankergebnisse auf Studien, die wahrscheinlich die Einschlusskriterien erfüllen. Eingeschlossen wurden nur Studien, die in Englisch, Deutsch oder Spanisch nach dem 31.12.1999 publiziert worden waren. Die Begrenzung auf Studien publiziert, in den zuvor genannten Sprachen, erfolgte auf Grund der Sprachkenntnisse der Autoren. Die Limitation durch das Publikationsdatum sollte eine bessere Vergleichbarkeit der Studien gewährleisten, da die technischen Anforderungen und Bedingungen zur Durchführung einer MRT sich zu diesem Zeitpunkt stark weiterentwickelten. Die Vorauswahl erfolgte an Hand der Zusammenfassung der Studie.

Danach wurde zu den vorausgewählten Arbeiten der Volltext von zwei Personen durchgearbeitet (Autor Christina Vargas Mora und zweiter Prüfer Dr. med. Sebastian Nagel) und nach Studien gesucht, in denen die diagnostische Genauigkeit der mp-MRT mit Kombination aus T2, DWI und DCE mit den histopathologischen Ergebnissen der Standardbiopsie als Referenzstandard auf Basis jedes einzelnen Patienten verglichen

wurde und nicht auf Grundlage einer pro-Prostatahälften- oder pro-Läsions-Analyse. Als Mindestfeldstärke der mp-MRT wurden 1,5 Tesla gefordert. Es wurden nur prospektive diagnostische Genauigkeitsstudien und keine randomisierten Studien eingeschlossen. Eine Berücksichtigung der Studien konnte auch erfolgen, wenn sie zwar retrospektiv waren, aber die mp-MRT noch vor der systematischen Biopsie als geforderter Referenzstandard durchgeführt wurde und bei der mp-MRT-Auswertung das histologische Ergebnis dem Radiologen nicht bekannt war. Als Standard-Biopsie wurde eine transperineale oder transrektale Biopsie mit mindestens acht Stanzen, meist Ultraschall gestützt, definiert.
Eingeschlossen wurden nur Studien mit Patienten mit hochgradigem Verdacht auf ein Prostata-Karzinom auf Grund von erhöhtem PSA-Wert und/oder positiver digital-rektaler Untersuchung ohne bereits positive Biopsie und wenn Absolutwerte für RP (Richtig Positive), FP (Falsch Positive), RN (Richtig Negative) und FN (Falsch Negative) oder genügend andere Daten genannt waren, die das Bilden eine 4-Felder-Tafel ermöglichten.

Ausgeschlossen wurden Tierstudien, Übersichtsarbeiten, Fallberichte, Expertenmeinungen, Briefe und Editorials. Studien in denen alle Patienten positiv auf ein Karzinom getestet wurden, Arbeiten, in denen ein Rezidiv erkannt werden sollte, Staging oder Follow-Up Studien oder mit Patienten unter aktiver Beobachtung, Studien mit einer Feldstärke kleiner 1,5 Tesla, mit per-Prostatahälften- oder per-Läsions-Analyse, Studien publiziert vor 2000, oder mit nur positiven MRT-Sequenzen oder in denen nicht alle Patienten die MRT in der geforderten Sequenzkombination erhielten oder die Sequenzen nicht in Kombination vom Radiologe ausgewertet wurden, sondern einzeln, Studien, in denen der Referenzstandard nicht eine systematische Biopsie war, sondern nur die gezielte Biopsie aus suspekten Arealen, Studien aus denen keine 4-Felder-Tafeln erhoben werden konnten.

5.3 Datenaufarbeitung und Qualitätsbewertung

Schließlich wurden 13 Studien eingeschlossen, aus denen die beiden Reviewer (Christina Vargas Mora, Dr. med. Sebastian Nagel) die Zahl der in der mp-MRT als RP, FP, FN und

RN diagnostizierten Personen berechneten oder die Daten zum Bilden einer 4-Felder-Tafel entnahmen. Es war nicht möglich, eine 3 x 2-Felder-Tafel (Schuetz et al. 2012) zu bilden, da nicht alle Studien offen legten, welche der als „nicht eindeutig“ eingestuften MRT's am Ende in der Referenzbiopsie einen Tumornachweis erbrachten oder nicht. Beide Reviewer bewerteten die eingeschlossenen Studien bezüglich ihrer Qualität mittels des Quadas-2 Bogens (Whiting et al. 2011). Die abschließende Bewertung erfolgte als Konsensentscheidung nach Diskussion aller sich eventuell ergebenden Diskrepanzen. Wenn es möglich war, wurde das PI-RADS (Prostate Imaging-Reporting and Data System) (ACR 2015) genutzt, um Cut-off-Level zu beschreiben. Als Cut-off-Level wurde in den Studien meist vier angegeben. Ließen Studien verschiedene mögliche Cut-off-Level zu, um eine 4-Felder-Tafel zu bilden, wurde mittels Youden-Index (Youden 1950), der Schwellenwert ermittelt, für den sich die beste diagnostische Genauigkeit der mp-MRT ergab. Das führte meist zu einem Cut-off von PI-RADS >= 4, außer in einer Studie. In dieser ergab sich für PI-RADS >= 3 die beste diagnostische Genauigkeit (Thompson et al. 2014). Trotzdem wurden von dieser Studie die Daten für PI-RADS >= 4 erhoben und genutzt, um eine bessere Vergleichbarkeit der Studien und der Ergebnisse zu gewährleisten. In einer Studie erhielten nicht alle Patienten alle mp-MRT Sequenzen, sodass die entsprechende 4-Felder-Tafel ohne diese Patienten gebildet wurde (Jambor et al. 2014). In einer weiteren Studie bioptierte man drei Patienten nicht systematisch, sondern nur gezielt an in der mp-MRT auffälligen Regionen. Diese drei Patienten wurden beim Erstellen der 4-Felder-Tafel nicht berücksichtigt (Sonn et al. 2014).

Weiterhin wurden Daten bezüglich der Bildinterpretation, des Referenzstandards, Charakteristika der Studienpopulation gesammelt. Wenn das Patientenalter, der PSA-Wert und der Gleason-Score als Medianwert gegeben waren, wurde der Mittelwert mit Hilfe der Wan-Methode (Wan et al. 2014) berechnet. Es wurde Formel 14 $\left(\overline{X} \approx \frac{q1+m+q3}{3}\right)$ verwendet, wenn 25. und 75. Perzentile gegeben waren und Formel 3 $\left(\overline{X} \approx \frac{a+2m+b}{4}\right)$, wenn Minimum und Maximum gegeben waren. Um fehlende Daten zu vervollständigen, wurden die entsprechenden Autoren angeschrieben.

5.4 Datenanalyse

Die Berechnungen wurden mit STATA 12.0 und dem Programm „MIDAS“ durchgeführt.

Um auf einen eventuellen Publikationsbias zu testen, wurde ein Funnel-Plot (Deeks et al. 2005) genutzt.

Um die individuelle und Gesamt-Sensitivität, –Spezifität und die 95 % Konfidenz-Intervalle zu berechnen, wurden die von uns gebildeten 4-Felder-Tafeln genutzt. Dann erfolgte die Berechnung mittels eines bivariaten logistischen Regressionsmodells mit zufälligen Effekten (bivariate random-effects model) (Chu et al. 2010). Dieses Model nimmt eine exakte Binomialverteilung für Sensitivität und Spezifität zwischen den Studien an (Riley et al. 2008). Auf Grundlage dieses Models, wurden mittels logistischer Regression die Durchschnitts-Sensitivität und -Spezifität mit jeweiliger Standard-Abweichung und 95 % Konfidenz-Intervall bestimmt. Es wurde außerdem die Variabilität der Sensitivität und Spezifität zwischen den Studien mittels der Kovarianzmatrix der zufälligen Effekte abgeschätzt. Mit Hilfe der via logistischer Regression berechneten Werte für Sensitivität und Spezifität und deren Varianz wurde eine Grenzwertoptimierungskurve (summary Receiver Operating Charachteristic Curve – sROC-Kurve) mit zusammenfassendem Test-Genauigkeitspunkt (summary Operating Point) für die Gesamt-Sensitivität und -Spezifität auf der Kurve, einem 95 %-Konfidenz-Ellipsoid und einem 95 %-Vorhersage-Ellipsoid generiert. Um die Heterogenität zu bewerten, wurde zusätzlich der Chi-Quadrat-Test verwendet und der I^2-Index berechnet (Higgins und Thompson 2002). Weiterhin wurde eine Meta-Regressionsanalyse durchgeführt, um zu untersuchen, wieviel der Heterogenität durch die Studienpopulationscharakteristika oder andere studienspezifische Einflussgrößen zu erklären ist. Kovariablen waren das Publikationsjahr, der mittlere PSA-Wert, der mittlere Gleason-Score, das mittlere Patientenalter, ob Patienten mit PDRU (positiv digital-rektaler Untersuchung) eingeschlossen wurden, ein prospektives Studiendesign, die mp-MRT Feldstärke, der Einsatz einer Endorektal- oder anderer Spulen, der Cut-off-Wert für eine auffällige mp-MRT, die Nutzung des PI-RADS-Scores zur mp-MRT Auswertung, die Kenntnis des auswertenden Radiologen über klinische Daten, die Anzahl der genommenen

Biopsieproben, ob zusätzliche Proben aus auffälligen Arealen entnommen wurden, ob Proben nur aus der peripheren oder auch der Transitionszone entnommen wurden. Abschließend wurde noch eine Subgruppenanalyse durchgeführt.

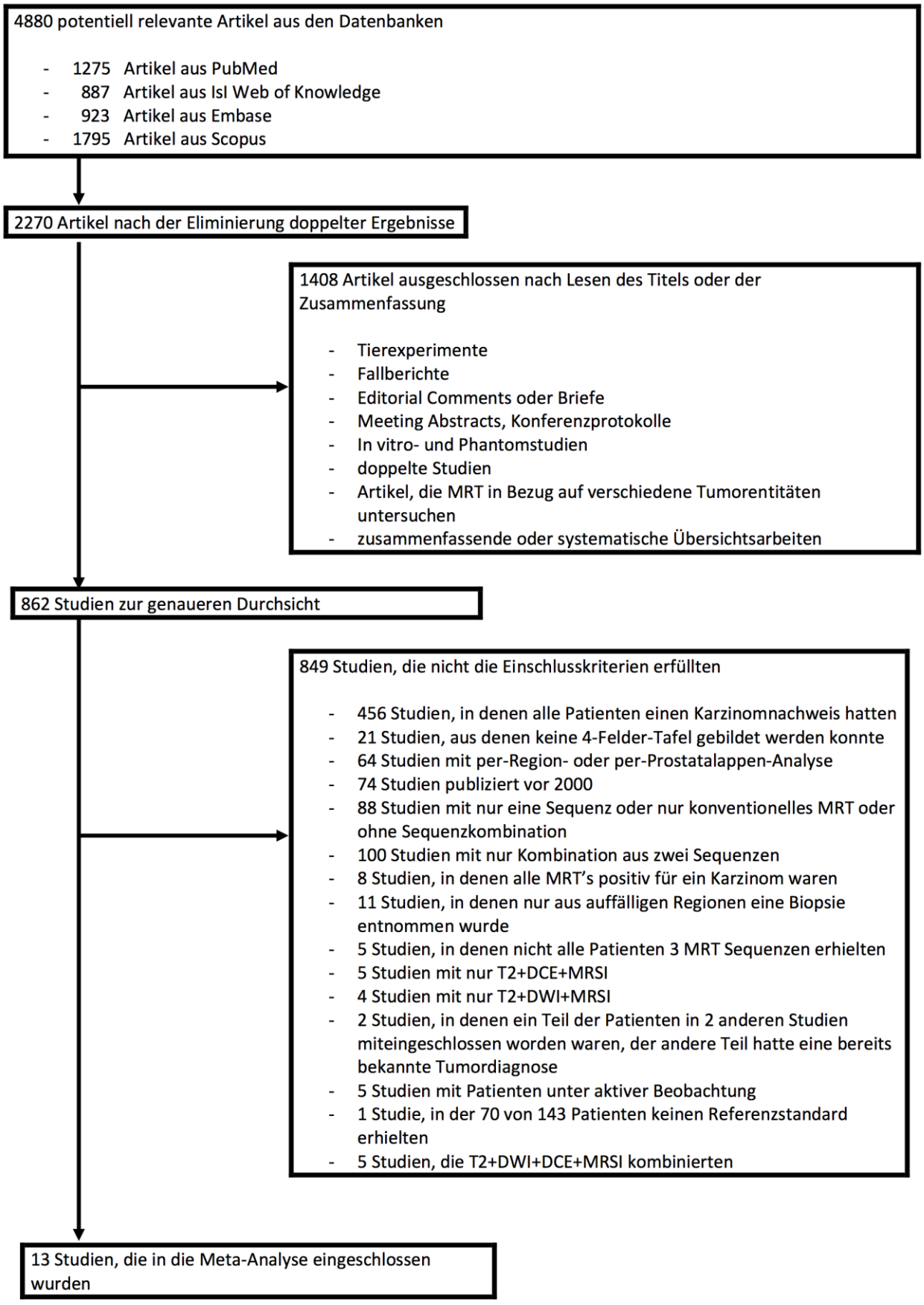
6. Ergebnisse

6.1 Studiencharakteristika

Beim Durchsuchen der elektronischen Datenbanken PubMed, Isl Web of Knowledge, Scopus und Embase, wurden 4880 potentiell relevante Artikel identifiziert. Nach dem Herausfiltern doppelter Studien konnten 2270 Studien in die Datenbank EndNote geladen werden. Anschließend wurden 1408 Studien nach dem Durchschauen der Titel und Zusammenfassungen aussortiert. Danach blieben 862 Studien übrig, die genauer analysiert wurden. Bei 518 Artikeln wurde der Volltext besorgt und durchgearbeitet. Davon erfüllten 849 Studien die Einschlusskriterien nicht. Die verbleibenden 13 Studien wurden in die Meta-Analyse eingeschlossen (Dikaios et al. 2014, Fiard et al. 2013, Iwamoto et al. 2014, Jambor et al. 2014, Kuru et al. 2013, Park et al. 2011, Pokorny et al. 2014, Porpiglia et al. 2014, Rosenkrantz et al. 2012, Sonn et al. 2014, Tamada et al. 2011, Tanimoto et al. 2007, Thompson et al. 2014). Das Durchsehen der Referenzen von Zusammenfassenden und

Systematischen Übersichtsarbeiten zum Thema erbrachte keine zusätzlichen Ergebnisse. Die im Vorfeld genannten Ein- und Ausschlusskriterien sowie das Flussdiagramm zur Studienauswahl sind in Abb. 1 dargestellt.

Abb. 1: Flussdiagramm zur Studienauswahl


4880 potentiell relevante Artikel aus den Datenbanken
- 1275 Artikel aus PubMed
- 887 Artikel aus Isl Web of Knowledge
- 923 Artikel aus Embase
- 1795 Artikel aus Scopus
2270 Artikel nach der Eliminierung doppelter Ergebnisse
1408 Artikel ausgeschlossen nach Lesen des Titels oder der Zusammenfassung
- Tierexperimente
- Fallberichte
- Editorial Comments oder Briefe
- Meeting Abstracts, Konferenzprotokolle
- In vitro- und Phantomstudien
- doppelte Studien
- Artikel, die MRT in Bezug auf verschiedene Tumorentitäten untersuchen
- zusammenfassende oder systematische Übersichtsarbeiten
862 Studien zur genaueren Durchsicht
849 Studien, die nicht die Einschlusskriterien erfüllten
- 456 Studien, in denen alle Patienten einen Karzinomnachweis hatten
- 21 Studien, aus denen keine 4-Felder-Tafel gebildet werden konnte
- 64 Studien mit per-Region- oder per-Prostatalappen-Analyse
- 74 Studien publiziert vor 2000
- 88 Studien mit nur eine Sequenz oder nur konventionelles MRT oder ohne Sequenzkombination
- 100 Studien mit nur Kombination aus zwei Sequenzen
- 8 Studien, in denen alle MRT's positiv für ein Karzinom waren
- 11 Studien, in denen nur aus auffälligen Regionen eine Biopsie entnommen wurde
- 5 Studien, in denen nicht alle Patienten 3 MRT Sequenzen erhielten
- 5 Studien mit nur T2+DCE+MRSI
- 4 Studien mit nur T2+DWI+MRSI
- 2 Studien, in denen ein Teil der Patienten in 2 anderen Studien miteingeschlossen worden waren, der andere Teil hatte eine bereits bekannte Tumordiagnose
- 5 Studien mit Patienten unter aktiver Beobachtung
- 1 Studie, in der 70 von 143 Patienten keinen Referenzstandard erhielten
- 5 Studien, die T2+DWI+DCE+MRSI kombinierten
13 Studien, die in die Meta-Analyse eingeschlossen wurden

Informationen zu Charakteristika der Studienpopulation sind in Tabelle 1 zu sehen.

Tab. 1: Charakteristika der Studienpopulation

Studie	Patientenzahl	Anzahl Ausgeschlossener	Studiendesign prospektiv	Mittlerer PSA	Mittlerer Gleason-Score	Mittleres Alter	Zahl d. Patienten mit PNB	PDRU
1 (Dikaios et al. 2014)	85	n. g.	nein	8,66	n. g.	63	n. g.	n. g.
2 (Fiard et al. 2013)	30	n. g.	ja	6,77	6,68	64	17	10
3 (Iwamoto et al. 2014)	238	95	nein	9,6	n. g.	69,2	0	n. g.
4 (Jambor et al. 2014)	54	1	ja	7,70	6,76	65,51	0	0
5 (Kuru et al. 2013)	347	n. g.	ja	9,85	n. g.	65,3	170	94
6 (Park et al. 2011)	44	10	ja	6,1	7,31	63	0	8
7 (Pokorny et al. 2014)	223	3	ja	11,6	n. g.	62,67	0	40
8 (Porpiglia et al. 2014)	170	17	ja	7,3	6,53	65	170	13
9 (Rosenkrantz et al. 2012)	42	n. g.	nein	8,1	n. g.	63	29	n. g.
10 (Sonn et al. 2014)	102	3	ja	7,9	n. g.	64,67	105	n. g.
11 (Tamada et al. 2011)	50	4	nein	6,84	7,5	70	n. g.	n. g.
12 (Tanimoto et al. 2007)	83	n. g.	ja	19,4	6,9	67,4	n. g.	n. g.
13 (Thompson et al. 2014)	150	15	ja	5,87	n. g.	61,27	12	44

PNB: vorherige negative Biopsie, PDRU: positive digital-rektale Untersuchung, n. g.: nicht genannt

Tab. 2: Charakteristika der multiparametrischen MRT

Studie	Cut-off	Feld-stärke in Tesla	Endo-rektal Spule	An-dere Spu-len	Radiologe kennt klinische Information	Anzahl auswerten-der Radiologen
1 (Dikaios et al. 2014)	PI-RADS >= 4	1,5	nein	Ja	n. g.	2
2 (Fiard et al. 2013)	PI-RADS-Summen-Score >= 5 = Tumorverdacht	3	nein	Ja	ja	1
3 (Iwamoto et al. 2014)	mind. in 2 Sequenzen Tumor-verdacht	1,5 oder 3	n. g.	n. g.	nein	n. g.
4 (Jambor et al. 2014)	PI-RADS >= 4	3	nein	Ja	nein	n. g.
5 (Kuru et al. 2013)	mind. in 2 Sequenzen Tumor-verdacht	3	nein	n. g.	ja	2
6 (Park et al. 2011)	mind. in 2 Sequenzen Tumor-verdacht	3	nein	Ja	ja	2
7 (Pokorny et al. 2014)	PI-RADS >= 4	3	nein	Ja	ja	3
8 (Porpiglia et al. 2014)	mind. in 2 Sequenzen Tumor-verdacht	1,5	ja	Ja	nein	1
9 (Rosenkrantz et al. 2012)	dichotomisiert	3	nein	ja	n. g.	2
10 (Sonn et al. 2014)	PI-RADS >= 4	3	nein	ja	nein	1
11 (Tamada et al. 2011)	in mind. 1 Sequenz Tumorverdacht	1,5	nein	ja	ja	2
12 (Tanimoto et al. 2007)	PI-RADS >= 3	1,5	nein	ja	n. g.	2
13 (Thompson et al. 2014)	PI-RADS >= 4	1,5 oder 3	nein	ja	ja	2

PI-RADS - Prostate Imaging-Reporting and Data System, n. g. - nicht genannt, mind. - mindestens

Tab. 3: Charakteristika des Referenzstandards (systematische Biopsie)

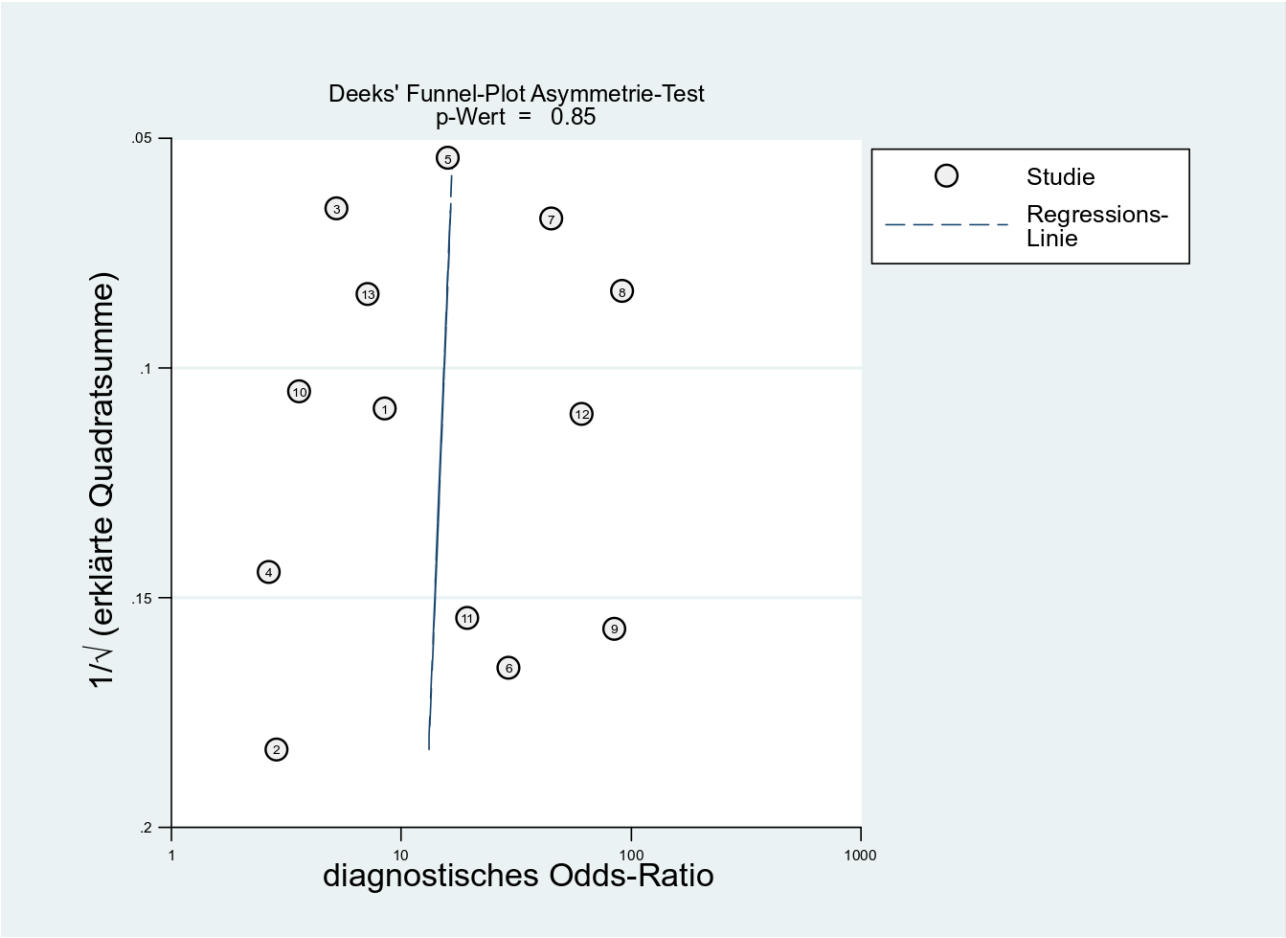
Studie	Anzahl der Proben	Ultraschall-gestützte Biopsie	Zusätzliche Proben	Biopsiezone
1 (Dikaios et al. 2014)	24	ja	nein	ganze Prostata
2 (Fiard et al. 2013)	12	ja	ja	ganze Prostata
3 (Iwamoto et al. 2014)	14	ja	ja	ÜZ und PZ
4 (Jambor et al. 2014)	12	ja	ja	ÜZ und PZ
5 (Kuru et al. 2013)	12	ja	ja	ganze Drüse
6 (Park et al. 2011)	10	ja	ja	PZ, zusätzliche Proben aus ÜZ, wenn MRI auffällig
7 (Pokorny et al. 2014)	12	ja	ja	ÜZ und PZ
8 (Porpiglia et al. 2014)	18	ja	nein	ÜZ und PZ
9 (Rosenkrantz et al. 2012)	16	n. g.	ja	n. g.
10 (Sonn et al. 2014)	12	ja	ja	ÜZ und PZ
11 (Tamada et al. 2011)	12	ja	nein	ÜZ und PZ
12 (Tanimoto et al. 2007)	n. g.	n. g.	nein	n. g.
13 (Thompson et al. 2014)	18	ja	ja	ÜZ und PZ

n. g. – nicht genannt, ÜZ – Übergangszone, PZ- periphere Zone

6.2 Publikationsbias

Der Funnel-Plot, in Abb. 2 zu sehen, ergab einen p-Wert von 0,85 und damit keine Hinweise auf einen Publikationsbias.

Abb. 2: Funnel-Plot zum Publikationsbias



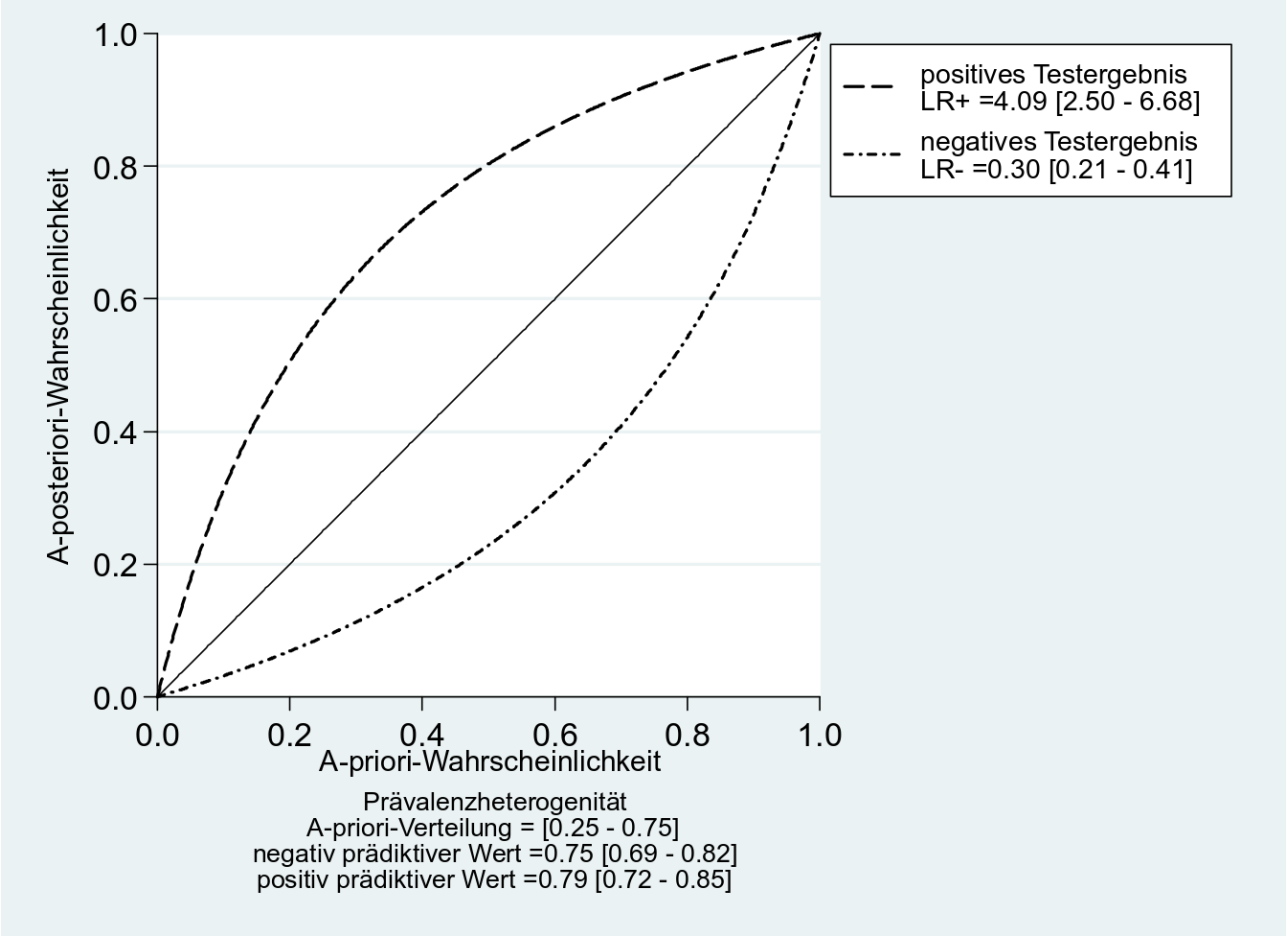
6.3 Diagnostische Genauigkeit

Die Analyse beinhaltete auf Grund der 13 eingeschlossenen Studien 1618 Patienten und davon 871 mit der Diagnose Prostata-Karzinom.

Die Gesamt-Sensitivität betrug 76 % (95 % Konfidenz-Intervall (KI), 0,67-0,83), die Gesamt-Spezifität 81 % (95 % KI, 0,70-0,89). Das PLR (positives Likelihood-Ratio) war 4,09 (95 % KI, 2,5-6,7), das NLR (negatives Likelihood-Ratio) 0,30 (95 % KI, 0,21-0,41), zu sehen in Abb. 3. Für das diagnostische OR (Odds-Ratio) ergab sich der Wert 14 (95 % KI, 7-27).

Der Test auf Heterogenität erbrachte die folgenden Ergebnisse: Q = 64.3 und Inkonsistenz I^2 von 97 % (95 % KI, 95-99). Für Q war das Ergebnis mit $p < 0,001$statistisch signifikant.

Abb. 3: Positives und negatives Likelihood-Ratio



Die gepoolten Ergebnisse sind in Tabelle 4 dargestellt. Tabelle 5 zeigt studienspezifische Schätzer.

Tab. 4: Gepoolte Schätzer

Methode	Sensitivität	95 % KI	Spezifität	95 % KI	PLR	95 % KI
T2+DWI+DCE	0,76	0,67-0,83	0,81	0,70-0,89	4,1	2,5-6,7
	NLR	95 % KI	diagnostisches OR	95 % KI	I^2	95 % KI
T2+DWI+DCE	0,30	0,21-0,41	14	7-27	97	95-99

KI: Konfidenzintervall, PLR: positive Likelihood-Ratio, NLR: negative Likelihood-Ratio, OR: Odds-Ratio, I^2: Inkonsistenz

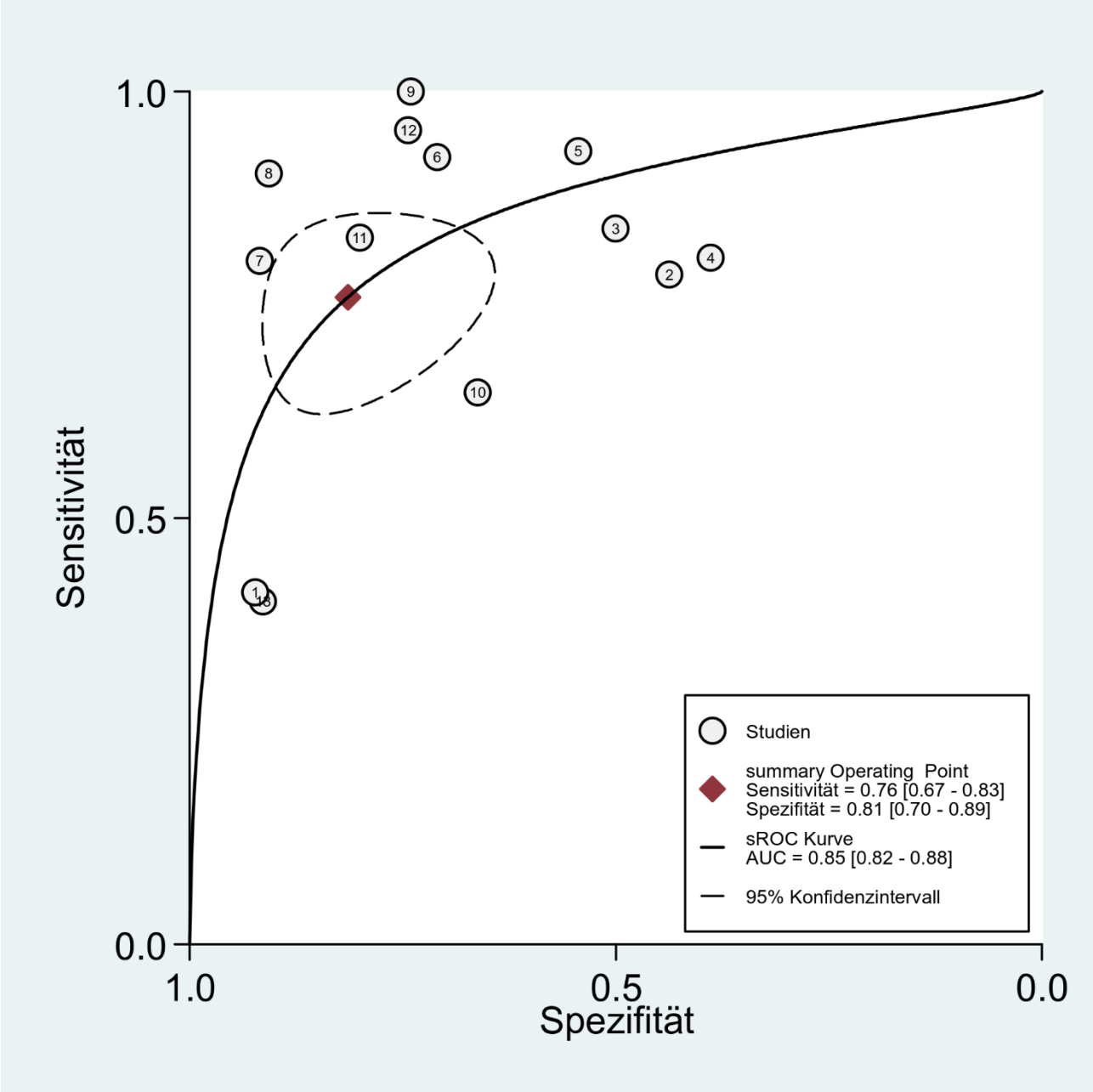
Tab. 5: Studienspezifische Schätzer

Studie	RP	FN	FP	RN	Sensitivität (%)	Spezifität (%)
1 (Dikaios et al. 2014)	19	27	3	36	41,30	92,31
2 (Fiard et al. 2013)	11	3	9	7	78,57	43,75
3 (Iwamoto et al. 2014)	89	17	66	66	83,96	50,00
4 (Jambor et al. 2014)	29	7	11	7	80,56	38,89
5 (Kuru et al. 2013)	186	14	67	80	93,00	54,42
6 (Park et al. 2011)	12	1	9	22	92,31	70,97
7 (Pokorny et al. 2014)	101	25	8	89	80,16	91,75
8 (Porpiglia et al. 2014)	47	5	11	107	90,38	90,68
9 (Rosenkrantz et al. 2012)	15	0	7	20	100,00	74,07
10 (Sonn et al. 2014)	22	12	23	45	64,71	66,18
11 (Tamada et al. 2011)	29	6	3	12	82,86	80,00
12 (Tanimoto et al. 2007)	42	2	10	29	95,45	74,36
13 (Thompson et al. 2014)	37	55	5	53	40,22	91,38

RP: Richtig Positive, FN: Falsch Negative, FP: Falsch Positive, RN: Richtig Negative

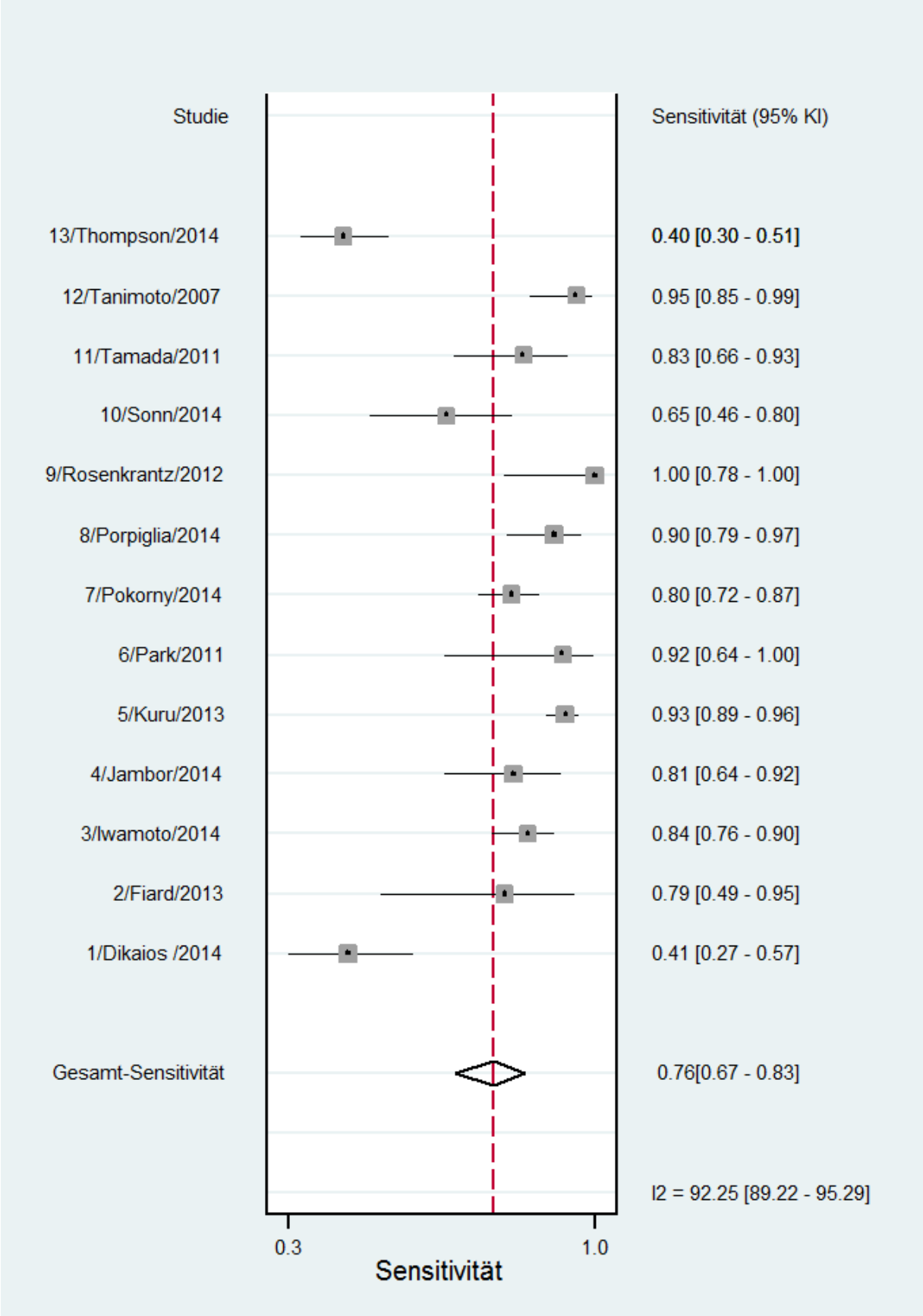
Bei Aufstellung der sROC-Kurve ergab sich eine Fläche unter der Kurve (Area Under the Curve (AUC)) von 0,85 (95 % KI, 0,82-0,88). Die sROC-Kurve ist in Abb. 2 dargestellt. Die studienspezifischen Sensitivitäten und Spezifitäten sind im Forest-Plot in Abb. 3 und 4 zu sehen.

Abb. 4: sROC-Kurve



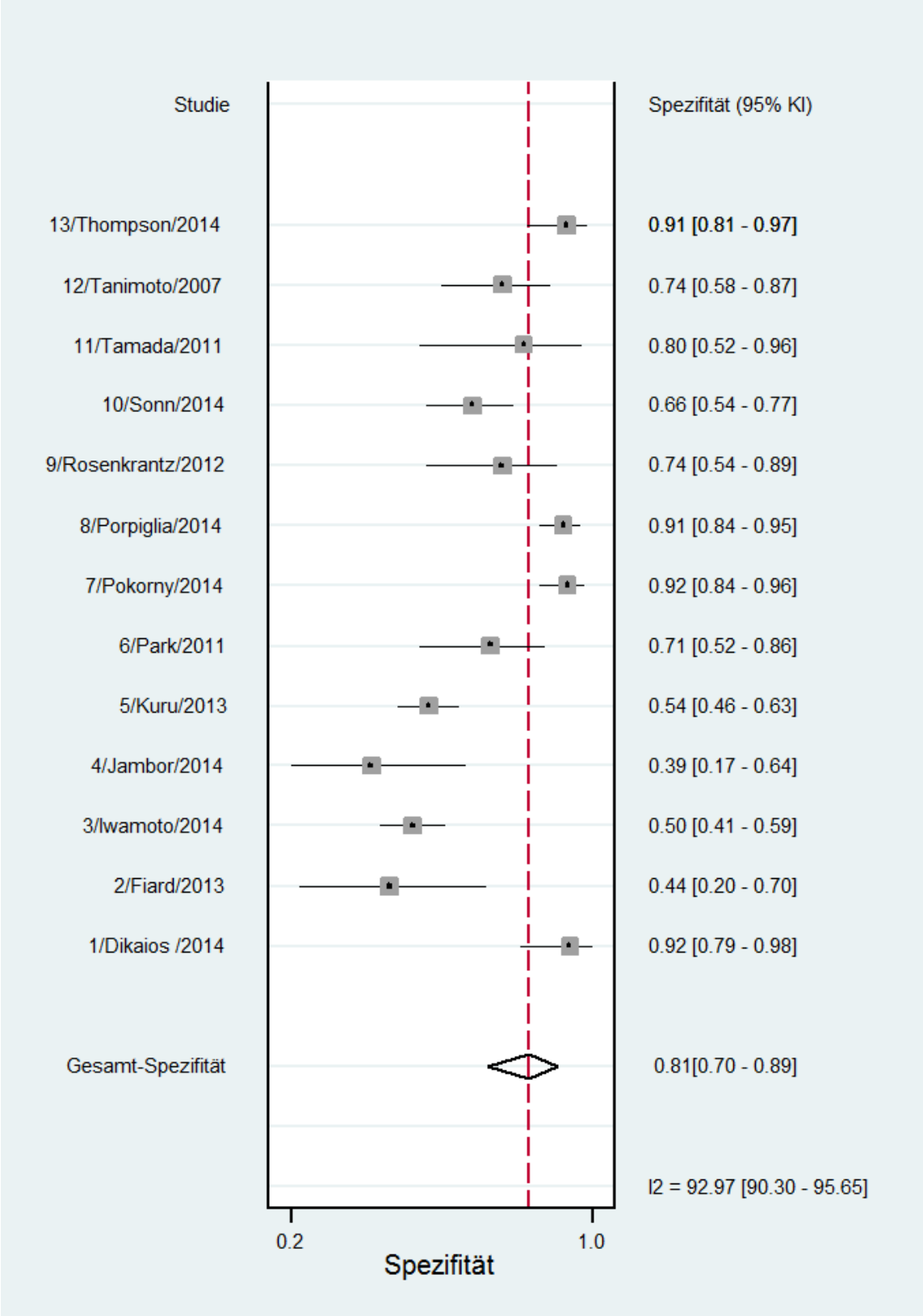
AUC – Fläche unter der Kurve, sROC- summary Receiver Operating Characteristic Curve

Abb. 5: Studienspezifische Sensitivität



KI - Konfidenzintervall, Q - Heterogenität, I^2 - Inkonsistenz

Abb. 6: Studienspezifische Spezifität



KI - Konfidenzintervall, Q - Heterogenität, I^2 - Inkonsistenz

6.4. Meta-Regressionsanalyse

Der Inkonsistenz-Index war mit I^2 = 97 % sehr hoch. Im Versuch einen Teil der hohen Heterogenität durch eventuelle Einflussgrößen zu erklären, wurde eine univariable Meta-Regression für mehrere mögliche Einflussfaktoren durchgeführt.

Als erstes wurde der zur Interpretation der MRT Sequenzen verwendete Cut-off-Wert untersucht. Für diesen waren die Ergebnisse signifikant ($p < 0,01$), erklärte aber nur 13 % der Heterogenität und lies I^2 = 84% unerklärt. Weiterhin erbrachten die möglichen Einflussfaktoren wie Anzahl der genommenen Biopsien, die Biopsiezone, die Tatsache, ob die auswertenden Radiologen Zugang zu klinischen Informationen hatten (wie z. B. PSA-Wert) und die Benutzung einer Endorektalspule signifikante Ergebnisse ($p < 0,01$). Eine ausreichende Erklärung für die hohe Heterogenität erbrachte keiner der Einflussfaktoren.

Andere mögliche Einflussgrößen wie das Publikationsjahr, der PI-RADS-Score, die Entnahme zusätzlicher Biopsieproben, eine Feldstärke von 3 Tesla und der PSA-Wert erklärten zwar viel Heterogenität, die Ergebnisse waren allerdings nicht signifikant. Die Ergebnisse der Meta-Regressionsanalyse sind in Tabelle 6 dargestellt.

Tab. 6: Meta-Regression - Faktoren, die möglicherweise Heterogenität erklären

Potentielle Einflussfaktoren	I^2 in %	p	Potentielle Einflussfaktoren	I^2 in %	p
Cut-off	84	<0,01	Publikationsjahr	43	0,17
Anzahl der Biopsieproben	85	<0,01	PI-RADS-Score	61	0,08
Biopsie-Zone	92	<0,01	zusätzliche Biopsieproben	54	0,11
Radiologe hat Kenntnis über klinische Daten	94	<0,01	3 Tesla Feldstärke	09	0,33

Endorektalspule	90	<0,01	PSA-Wert	12	0,32

PI-RADS - Prostate Imaging-Reporting and Data System, PSA - prostataspezifisches Antigen, I^2 - Inkonsistenz, p - Signifikanz

Schaut man sich in der Meta-Regressionsanalyse mögliche Einflussfaktoren auf Sensitivität und Spezifität an, so scheint eine Feldstärke von 3 Tesla die Sensitivität auf 70 % zu senken mit einem Signifikanzlevel von p = 0,01. Der gleiche Effekt, mit einer Signifikanz von p < 0,01, wurde für die Entnahme zusätzlicher Biopsieproben aus in der mp-MRT tumorverdächtig wirkenden Bereichen beobachtet. Die Benutzung des PI-RADS-Scores schien die Spezifität signifikant zu senken (p < 0,01). Ein ebensolcher Effekt, mit signifikanter Senkung der Spezifität (p = 0,01), konnte für das Publikationsjahr (Studien, die nach 2012 publiziert wurden, verglichen mit denen, die vor 2012 publiziert wurden) beobachtet werden. Im Gegensatz dazu, schien die Benutzung einer Endorektalspule die Spezifität zu erhöhen, auch wenn das Ergebnis nicht signifikant ausfiel (p = 0,06).

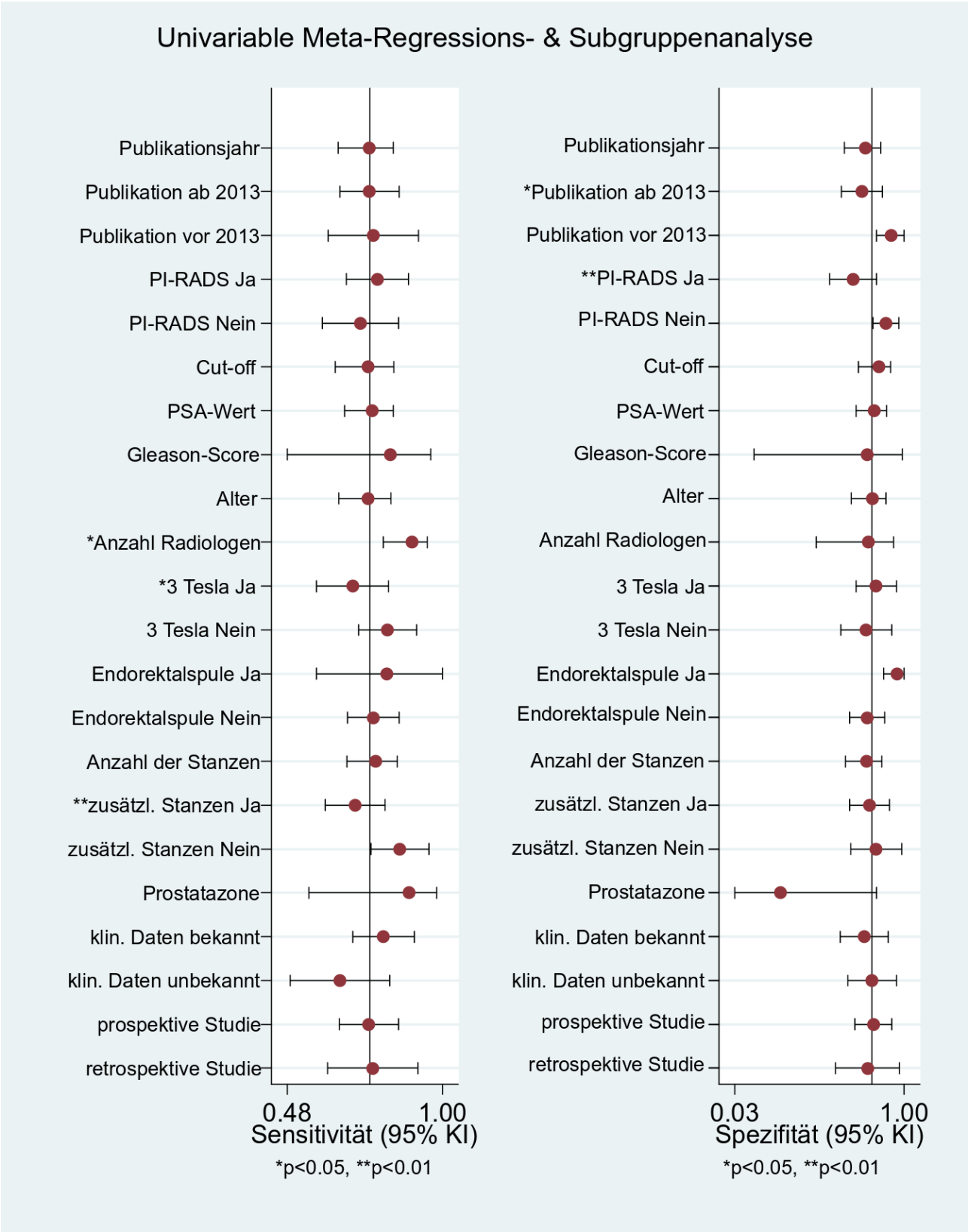
Die Ergebnisse sind in Tabelle 7 nummerisch und in Abb. 7 grafisch dargestellt.

Tab. 7: Mögliche Einflussfaktoren auf Sensitivität und Spezifität

Einflussfaktoren	Sensitivität + KI	p	Spezifität + KI	p
3 Tesla Feldstärke (7 Studien)	↓ 0,70 (0,58 - 0,82)	0,01	↑ 0,84 (0,72 - 0,95)	0,67
zusätzliche Biopsieproben (9 Studien)	↓ 0,71 (0,61 - 0,81)	<0,01	↓ 0,80 (0,69 - 0,92)	0,27
Anzahl von Radiologen	↑ 0,90 (0,80 - 0,95)	0,03	↓ 0,80 (0,50 - 0,94)	0,72
PI-RADS-Score (7 Studien)	↑ 0,78 (0,68 - 0,89)	0,37	↓ 0,71 (0,57 - 0,84)	<0,01
publiziert ab 2013 (9 Studien)	↔ 0,76 (0,66 -0,86)	0,22	↓ 0,76 (0,64 - 0,88)	0,01
Endorektalspule (1 Studie)	↑ 0,81 (0,58 – 1,00)	0,99	↑ 0,96 (0,88 - 1,00)	0,06

PI-RADS – Prostate Imaging-Reporting and Data System, KI – Konfidenzintervall

Abb. 7: Mögliche Einflussfaktoren auf Sensitivität und Spezifität



6.5 Subgruppenanalyse

Den in der Meta-Regression beobachteten Einflüssen auf Sensitivität und Spezifität sollte mittels Subgruppenanalyse weiter nachgegangen werden.
Die Benutzung des PI-RADS-Scores, ein PSA-Wert > = 9ng/ml und eine Feldstärke von 1,5 Tesla erhöhten die Sensitivität, wohingegen eine Feldstärke von 3 Tesla und die Kenntnis über klinische Angaben bei der Auswertung der MRT Sequenzen die Sensitivität senkte. Allerdings muss hier die kleine Anzahl an Studien in der Subgruppe und die Tatsache, dass alle 3 Tesla-MRT's ohne Endorektalspule durchgeführt worden, bedacht werden.
Die Benutzung des PI-RADS-Scores und eine Feldstärke von 1,5 Tesla senkten die Spezifität. Im Gegensatz dazu erhöhten ein PSA-Wert < 9 n g/ml, eine Feldstärke von 3 Tesla, Patienten mit wenigstens einer vorherigen negativen Biopsie und eine positive digital-rektale Untersuchung die Spezifität. Die Tabelle 8 zeigt die Ergebnisse.

Tab. 8: Subgruppenanalyse

	Anzahl der Studien	Sensitivität (%)	95 % KI	Spezifität (%)	95 % KI
PI-RADS	7	↑ 78	0,65-0,88	↓ 71	0,57-0,82
kein PI-RADS	6	↓ 72	0,61-0,81	↑ 90	0,78-0,96
PSA > = 9	4	↑ 78	0,62-0,89	↑ 82	0,77-0,86
PSA < 9	9	↓ 74	0,63-0,83	↓ 80	0,62-0,91
3 Tesla	7	↓ 70	0,70-0,81	↑ 83	0,71-0,91
1,5 Tesla*	6	↑ 82	0,70-0,89	↓ 78	0,59-0,90
keine Endorektalspule	11	↑ 77	0,67-0,85	↓ 79	0,66-0,88
PNB	7	↓ 74	0,62-0,83	↑ 82	0,64-0,92
PDRU	8	↑ 80	0,70-0,87	↑ 83	0,69-0,92
Kenntnis über klinische Daten	6	↑ 81	0,66-0,90	↓ 78	0,62-0,88

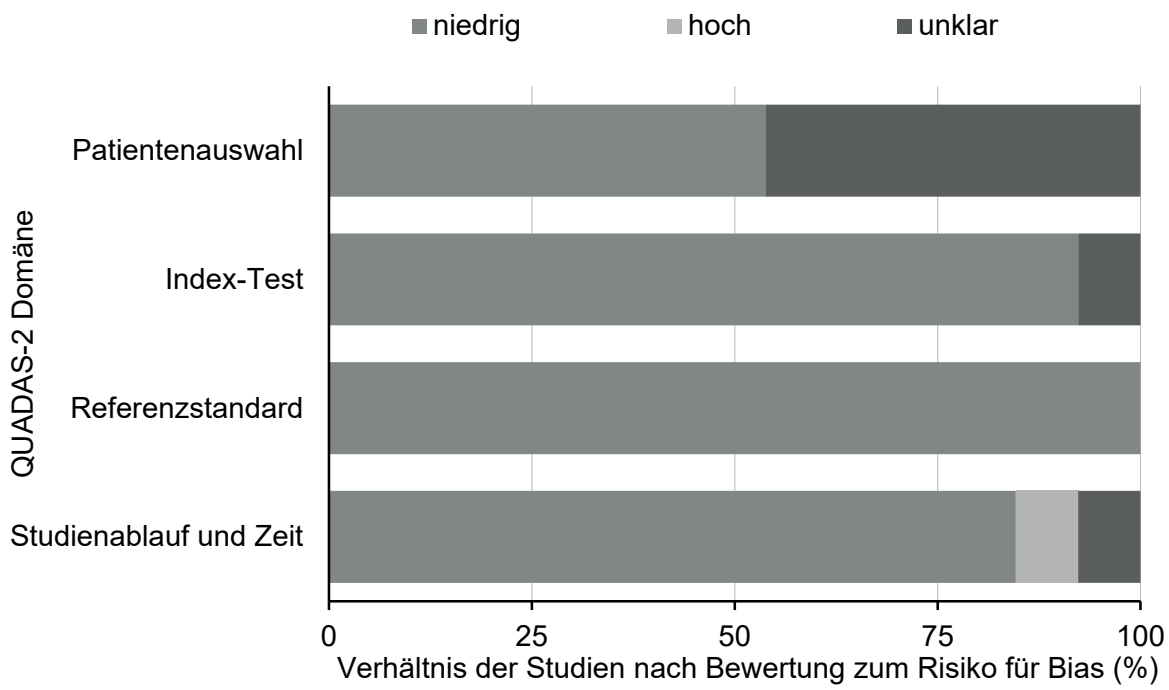
KI - Konfidenzintervall, PDRU - positive digital-rektale Untersuchung, PI-RADS - Prostate Imaging-Reporting and Data System, PNB - vorherige negative Biopsie, PSA - prostataspezifisches Antigen
*Einschließlich der Studien, die 1,5 und 3 Tesla-MRT benutzten

6.6. Qualitätsbewertung

Zwei unabhängige Reviewer bewerteten die individuelle Studienqualität mittels QUADAS-2 (Whiting et al. 2011). Im Falle von Uneinigkeit bezüglich der Bewertung wurden zwischen beiden Ihre zu Grunde liegenden Entscheidungskriterien diskutiert, bis ein Konsensus erreicht wurde.

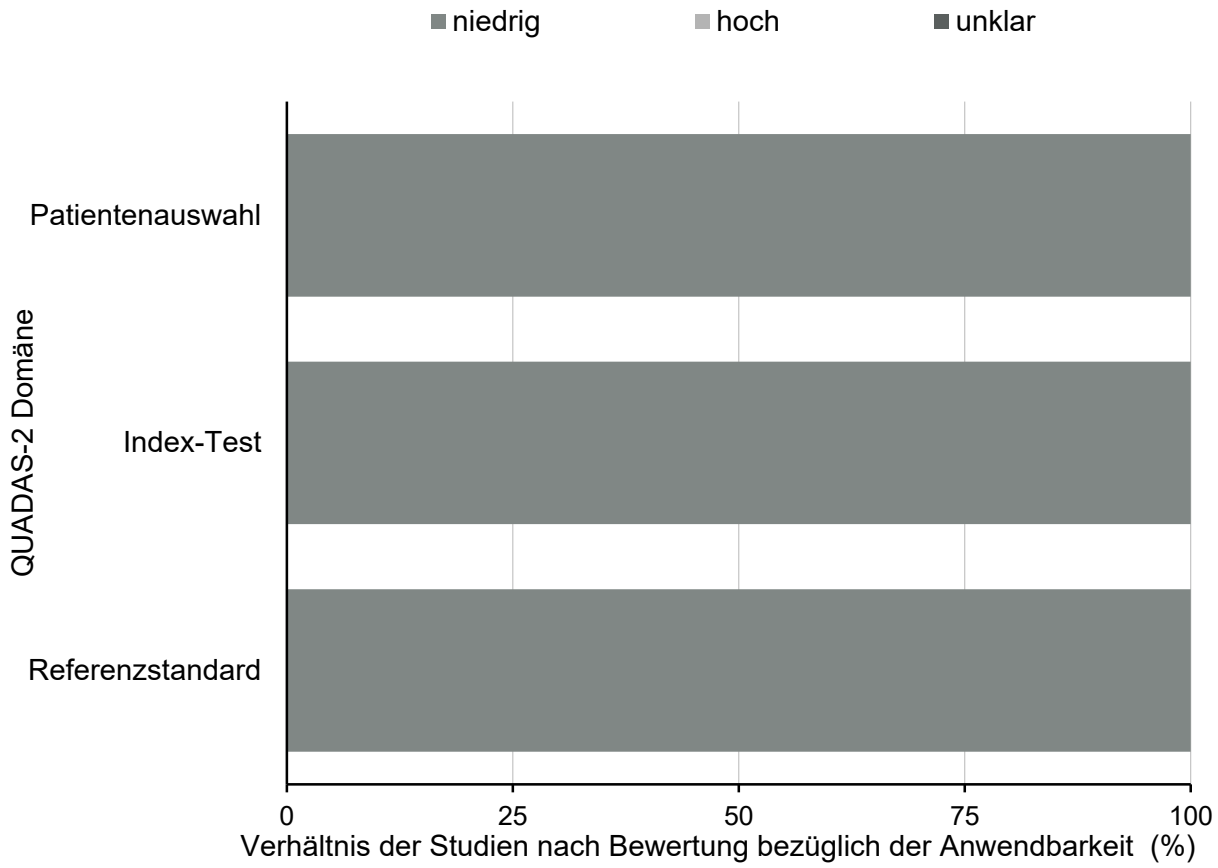
Die Auswertung mit Hilfe des QUADAS-2 Fragebogens zum Risiko für Bias sind in Abbildung 5 zu sehen. Insgesamt ist das Risiko für Bias als niedrig einzuschätzen. Zu erwähnen ist, dass im Bereich Patientenauswahl einige Studien nicht in die Tiefe gingen und hier manche Punkte unklar ließen (46 %). Im Bereich Studienablauf und Zeit wurden 7,7 % der Fragen mit einem hohen Risiko für Bias bewertet.

Abb. 8: Verhältnis der Studien nach Bewertung zum Risiko für Bias (%)



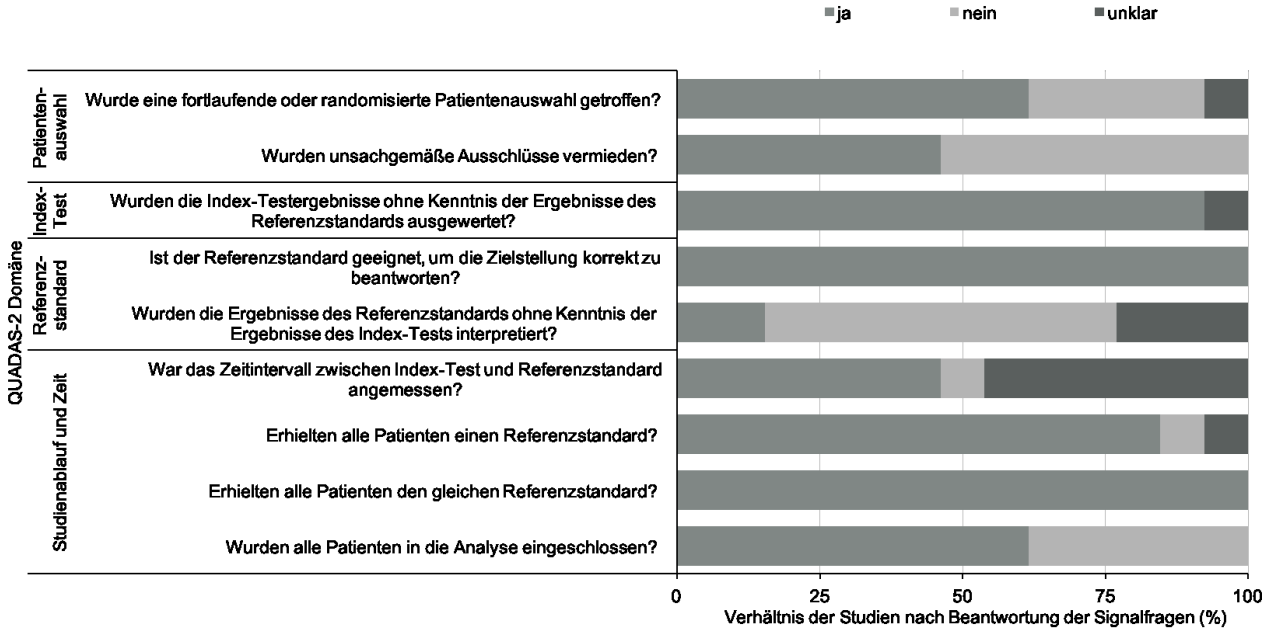
In Abbildung 6 wird klar, dass in Bezug auf die Anwendbarkeit der Studienergebnisse die Sicherheit besteht, dass die Fragestellung hinreichend gut geklärt wird und sich die Ergebnisse auf die Zielpopulation übertragen lassen.

Abb. 9: Verhältnis der Studien nach Bewertung bezüglich der Anwendbarkeit (%)



Das Säulendiagramm in Abbildung 7 stellt das Verhältnis der Studien nach Beantwortung der Signalfragen dar.

Abb. 10: Verhältnis der Studien nach Beantwortung der Signalfragen (%)



Die Kriterien für die Bewertung mittels QUADAS-2 können in den Tabellen 9-11 nachgelesen werden.

Tab. 9: Bewertungskriterien QUADAS-2 – Bereich 1

	Bereich 1: Patientenauswahl			
	A. Risiko für Bias			B. Bedenken bezüglich der Anwendbarkeit
	Wurde eine konsekutive oder randomisierte Studienpopulation eingeschlossen?	Vermied die Studie unangemessene Ausschlüsse?	Könnte die Patientenauswahl Bias verursacht haben?	Bestehen Bedenken, dass die eingeschlossenen Patienten nicht zur Review-Frage passen?
ja	prospektive Studie	andere	-	-
nein	retrospektive Studie, prospektive Studie mit randomisierten Patienten	Patienten mit vorangegangener negativer Biopsie, Patienten ohne Referenzstandard oder zurückgezogener Einwilligung, Patienten mit bereits diagnostiziertem Karzinom, Patienten mit unvollständigen Daten, Patientenausschluss ohne Nennen von Gründen	-	-
unklar	prospektiv, Patientenauswahl nicht erläutert, nur Anzahl eingeschlossener Patienten erwähnt	-	randomisiertes Kollektiv, retrospektive Studie, Ausschluss von Patienten mit vorangegangener negative Biopsie, Pateinten mit bestehendem Tumorverdacht	-
niedrig	-	-	andere	Patienten mit erhöhtem PSA-Wert
hoch	-	-	-	-

Tab. 10: Bewertungskriterien QUADAS-2 – Bereich 2-3

	Bereich 2: Index-Test			Bereich 3: Referenzstandard			
	A. Risiko für Bias		B. Bedenken bezüglich der Anwendbarkeit	A. Risiko für Bias			B. Bedenken bezüglich der Anwendbarkeit
	Wurden die Index-Testergebnisse ohne Kenntnis der Ergebnisse des Referenzstandards interpretiert?	Könnte die Durchführung oder Interpretation des Index-Tests zu Bias geführt haben?	Gibt es Bedenken, dass der Index-Test, dessen Durchführung, Interpretation von der Review-Frage abweichen?	Ist der Referenzstandard in der Lage die Fragestellung richtig zu beantworten?	Wurden die Ergebnisse des Referenzstandards ohne Kenntnis der Ergebnisse des Index-Tests interpretiert?	Könnte der Referenzstandard, seine Durchführung oder Interpretation zu Bias geführt haben?	Gibt es Bedenken, dass die Zielstellung, definiert durch den Referenzstandard, die Review-Frage nicht hinreichend beantwortet?
ja	prospektive Studie oder Radiologe blind für Ergebnisse der Referenz	-	-	systematische Biopsie	prospektive Studie ohne zusätzliche Stanzen, sodass alle Patienten die gleiche Anzahl an Stanzen erhielten	-	-
nein	-	-	-	-	wenn zusätzliche Stanzen aus suspekten Arealen genommen wurden	-	-
unklar	retrospektive Studie und es wurde nicht erwähnt, ob Radiologe blind für Referenz-Ergebnisse war	retrospektive Studie und es wurde nicht erwähnt, ob Radiologe blind für Referenz-Ergebnisse war	-	-	retrospektive Studie, es wurde nicht erwähnt, ob der Pathologe blind für Index-Ergebnisse war, schlecht beschriebener Referenzstandard, unklare Anzahl genommener Stanzen	retrospektive Studie, es wurde nicht erwähnt, ob der Pathologe blind für Index-Ergebnisse war, schlecht beschriebener Referenzstandard, unklare Anzahl genommener Stanzen	-
niedrig	-	prospektive Studie oder Radiologe blind für Referenz-Ergebnisse	mindestens ein Radiologe beurteilt mp-MRT Sequenzen auf Karzinomverdacht	-	-	andere	systematische Biopsie
hoch	-	-	-	-	-	-	-

Tab. 11: Bewertungskriterien QUADAS-2 – Bereich 4

	Bereich 4: Studienablauf und Zeit				
	A. Risiko für Bias				
	War das Intervall zwischen Index-Test und Referenzstandard angemessen?	Erhielten alle Patienten einen Referenzstandard?	Erhielten alle Pateinten den gleichen Referenzstandard?	Wurden alle Patienten in die Analyse eingeschlossen?	Könnte die Patientenauswahl Bias verursacht haben?
ja	bis zu 70 Tagen	alle erhielten eine Biopsie	auch wenn zusätzliche Stanzen aus auffälligen Arealen genommen wurden	auch wenn Patienten auf Grund der Ausschlusskriterien oder zurückgezogenem Einverständnis ausschieden	-
nein	Intervall > 4 Monate	beobachtendes Zuwarten, nicht alle Patienten erhielten eine Biopsie	beobachtendes Zuwarten	Patientenausschluss auf Grund unvollständiger Daten, fehlendem Referenzstandard oder ohne genannten Grund	-
unklar	3 - 4 Monate, Intervall nicht genannt	Patientenausschluss mit unvollständigen Daten; unklar, ob alle Patienten einen Referenzstandard oder den gleichen erhielten	-	-	wenn unklar blieb, ob Patienten auf Grund eines fehlenden Referenzstandards ausschieden
niedrig	-	-	-	-	andere
hoch	-	-	-	-	Ausschluss auf Grund von fehlendem Referenzstandard

7. Diskussion

Eine AUC von 85 lässt eine hohe diagnostische Genauigkeit erwarten. Die methodische Qualität der Ausgangsstudien war ebenfalls gut, aber es konnte auch eine Heterogenität unter den Studien von 97 % gezeigt werden. Auch die Sensitivität und die Spezifität variierten stark unter den Studien.

Die Ergebnisse einer Meta-Analyse aus dem Jahr 2012 (Wu et al. 2012), in der die diagnostische Genauigkeit von T2 kombiniert mit DWI an Hand von 10 Studien untersucht worden war, waren vergleichbar mit den Ergebnissen aus dieser Analyse. Wu und Kollegen ermittelten eine Sensitivität von 76 % (95 % KI, 0,65-0,84) und eine Spezifität von 82 % (95 % KI, 0,77-0,87). Das könnte darauf schließen lassen, das die DCE als zusätzliche dritte Sequenz keinen Mehrwert hat. Aber eine weitere Meta-Analyse mit sieben eingeschlossenen Studien (de Rooij et al. 2014), welche die Kombination aus den Sequenzen T2 + DWI + DCE untersuchte, kam zu besseren Ergebnissen mit einer Gesamt-Sensitivität von 0,81 (95 % KI, 0,75-0,85) und einer Gesamt-Spezifität von 0,91 (95 % KI, 0,67-0,98). Die meisten bei de Rooji und Kollegen eingeschlossenen Studien wurden in dieser Analyse auf Grund der verschiedenen Einschlusskriterien ausgeschlossen. In der hier vorliegenden Analyse wurden nur Patienten ohne eine gesicherte Prostatakarzinomdiagnose und Studien, die auf einem per-Patient- und nicht per-Region- oder per-Prostatahälften-Level ausgewertet wurden, eingeschlossen. Aus diesem Grund wurden nur zwei Studien in beide Analysen einbezogen (Tanimoto et al. 2007, Tamada et al. 2011). De Rooij und Kollegen führten auch eine Subgruppenanalyse durch, die wie in dieser Arbeit die Ergebnisse der mp-MRT auf einem per-Patient-Level betrachteten. Das macht sie mit den hier vorliegenden Ergebnissen besser vergleichbar. Dabei fanden sie eine Gesamt-Sensitivität von 0,78 (95 % KI, 0,65-0,87) und eine Gesamt-Spezifität von 0,88 (95 % KI, 0,80-0,94) und bemerkten ebenfalls eine hohe Heterogenität. Diese Ergebnisse unterscheiden sich kaum von den in dieser Analyse gewonnenen.

Eine weitere Meta-Analyse aus dem Jahr 2019 (Zhen et al. 2019) untersuchte ebenfalls die diagnostische Genauigkeit der mp-MRT. Allerdings lassen sich aus nachfolgend genannten Gründen die Ergebnisse kaum vergleichen. In der Meta-Analyse von Zhen und Kollegen

wurden 29 Studien untersucht. Neun der in der hier vorliegenden Arbeit eingeschlossenen Studien fanden bei Zhen und Kollegen keine Berücksichtigung, sodass die Literaturrecherche nicht als vollständig anzusehen ist. Nur vier (Kuru et al. 2013, Pokorny et al. 2014, Porpiglia et al. 2014, Thompson et al. 2014) der von uns eingeschlossen Studien wurden auch in der Meta-Analyse von 2019 berücksichtigt. Ähnlich wie in der Meta-Analyse von de Rooji und Kollegen, wurden auch die meisten bei Zhen und Kollegen untersuchten Studien auf Grund der anderen Ein- und Ausschlusskriterien in der hier vorliegenden Arbeit ausgeschlossen. So war z.B. das Bilden einer 4-Felder-Tafel in einigen Fällen nur auf einem per-Regions- oder per-Prostatahälften-Level möglich (Abd-Alazeez et al. 2014, Kitajima et al. 2010, Labanaris et al. 2010, Portalez et al. 2012, Wang et al. 2016). Teilweise wurden nur zwei statt drei MRT-Sequenzen kombiniert, oder nicht alle Patienten erhielten drei MRT-Sequenzen (Haffner et al. 2011, Ibrahiem et al. 2012, Javali et al. 2014, Kumar et al. 2007, Matsuoka et al. 2014, Numao et al. 2013, Rouse et al. 2011, Watanabe et al. 2012). Weiterhin wurden auch Studien ohne DCE oder DWI zu Gunsten von MRSI (Jagannathan und Indiran 2017, Petrillo et al. 2014, Sciarra et al. 2010), oder mit vier Sequenzen, also mit MRSI, untersucht (Busetto et al. 2013, Ferda et al. 2013, Hauth et al. 2015, Pepe et al. 2015a). In einer Studie war bei allen Patienten bereits ein Tumor bekannt (Popita et al. 2017) und es wurde die diagnostische Genauigkeit in Bezug auf klinisch signifikante Tumore untersucht.
So kommen Zhen und Kollegen zu einer hohen diagnostischen Genauigkeit mit einer AUC von 0,87 (95 % KI, 0,84-0,90), einer ebenfalls hohen Sensitivität von 0,87 (95 % KI, 0,81-0,91), aber deutlich niedrigeren Spezifität 0,68 (95 % KI, 0,56-0,79) und einer, in Anbetracht der oben genannten heterogenen Studiendesigns, der eingeschlossenen Arbeiten, entsprechenden jeweils hohen Heterogenität von $I^2 = 95{,}48$ % für die Sensitivität und $I^2 = 97{,}40$ % für die Spezifität, ohne diese in der Meta-Regressionsanalyse, oder Subgruppenanalyse hinreichend erklären zu können.

Mittels einer Meta-Regression sollten auch in der hier vorliegenden Arbeit Ursachen für die Heterogenität und Hinweise auf Einflussgrößen gefunden werden. Einen signifikanten Einfluss zeigten der Grenzwert, ab dem eine MRT als positiv betrachtet wird, die Nutzung einer Endorektalspule, die Anzahl der entnommenen Stanzen in der systematischen Biopsie, die Prostatazone, aus der die Stanzen entnommen wurden und ob den MRT

auswertenden Radiologen die klinischen Daten der Patienten bekannt waren. Einige der anderen Einflussfaktoren, wie das Publikationsjahr, der Gleason-Score, der mittlere PSA-Wert, die Feldstärke und ob zusätzliche Stanzen bei der Biopsie aus in der MRT suspekten Arealen entnommen wurden, waren nicht signifikant, ließen aber trotzdem einen Einfluss auf die Heterogenität zwischen den Studien vermuten.

In einigen der eingeschlossenen Studien wurde der PI-RADS-Score benutzt, um einen Grenzwert, ab dem eine MRT als positiv gewertet wird, zu finden. Dabei zeigte sich sowohl in der Meta-Regression als auch in der Subgruppenanalyse, dass die Benutzung des PI-RADS-Scores die Spezifität zu senken scheint. Das Ergebnis ist allerdings nicht signifikant. Es gibt mehrere Gründe, die dieses Ergebnis erklären könnten. Beim Durchsehen der eingeschlossenen Artikel fiel auf, dass es keinen Konsensus zu geben scheint, nach dem die MRT Sequenzen ausgewertet werden. Dabei zeigten sich folgende Probleme, die zukünftig vereinheitlicht werden sollten. Nicht alle Studien benutzen den von den ESUR Leitlinien empfohlenen PI-RADS-Score (Barentsz et al. 2012). Einige Studien dichotomisierten die Sequenzen in „auffällig“ und „unauffällig“, andere unterteilten in „unauffällig“, „zweifelhaft“ und „auffällig“. Hier sollte als Standard der PI-RADS-Score benutzt werden. Die Leitlinien geben eine klare Auskunft wie einzelne Sequenzen mit dem PI-RADS-Score zu bewerten sind. Wie die gemeinsame Auswertung bei multiparametrischer MRT erfolgen sollte, bleibt aber unklar. Es wird die Bildung eines Summenscores empfohlen, wobei nicht spezifiziert wird, wie dieser gebildet werden sollte und was ein möglicher Grenzwert sein könnte, ab dem die Sequenzen in ihrer Gesamtheit als positiv gewertet werden. In den Studien fiel diesbezüglich ein unterschiedliches Vorgehen auf. Einige werteten die mp-MRT als positiv, wenn mindestens eine, in anderen mindestens zwei, in wieder anderen alle drei Sequenzen positiv waren. Dabei ließen sie häufig unklar, wie mit den als „zweifelhaft“ eingestuften Sequenzen verfahren wurde. In den Studien die den PI-RADS-Score benutzen, herrschte ähnliche Uneinigkeit über die Bildung eines Summenscores. Wenn angegeben war, dass die MRT bei einem PI-RADS-Score > = 3 oder 4 als positiv eingestuft wurde, so blieb unerwähnt, ob dies in einer, zwei oder allen drei Sequenzen der Fall sein musste. Einige Studien lösten das Problem, indem sie einen Summenscore bildeten. Dabei wählte nur eine Studie (Fiard et al. 2013) die Methode nach Röthke (Rothke et al. 2013), um den Summenscore zu bilden, nutzte dann aber einen sehr niedrigen Cut-off Wert von > = 5 (entsprechend einem PI-RADS-Gesamtscore von 2).

Laut Röthke sollten die vergebenen Punkte 1 bis 5 aus den drei Sequenzen addiert werden. Dabei wird ein Wert von 7 bis 9 als „unklar“ eingestuft und mit einem PI-RADS-Gesamtscore von 3 bewertet. Ein PI-RADS-Summenscore von 10 bis 12 wird hingegen mit einem PI-RADS-Gesamtscore von 4 bewertet. Der PI-RADS-Gesamtscore 4 führte in der Studie von Schimmoller (Schimmoller et al. 2014) in 27 % der Fälle zu einem richtig positiven Ergebnis, eine Läsion mit einem PI-RADS-Gesamtscore 5 entsprach in 88 % der Fälle einem Prostatakarzinom. Das ist eine mögliche Methode, um die MRT Interpretation zu standardisieren. Nachdem eine Standardisierung bezüglich eines sinnvoll anwendbaren Scores erfolgt ist, sollte der bestmögliche Grenzwert untersucht und gefunden werden, bei dem die diagnostische Leistung der MRT optimal ist. Diese Inhomogenität in Bildinterpretation und Grenzwert, ist ein möglicher Grund für die hohe Heterogenität und die mäßige Sensitivität.

Wenn man die technischen Aspekte der MRT betrachtet, ist der aktuelle Stand der Wissenschaft, dass eine Kombination aus Endorektal- und Oberflächenspule mit einer Feldstärke von 1,5 Tesla die besten Ergebnisse liefert (Beyersdorff et al. 2003), mit einbeziehend das die handelsüblichen Endorektalspulen für 3 Tesla Scanner noch nicht einsetzbar sind. Eine Meta-Analyse zur diagnostischen Genauigkeit von DWI aus dem Jahr 2014 (Chen et al. 2014) beobachtete bessere Ergebnisse mit der Nutzung einer Endorektalspule für die DWI Sequenz. Auch wenn es nicht signifikant war, so konnte in der Meta-Regression doch eine Verbesserung der Spezifität mittels der Benutzung einer Endorektalspule gezeigt werden. Die Subgruppenanalyse verdeutlichte, dass ein Setting ohne Endorektalspule die Sensitivität verbessert und die Spezifität senkt. Die Meta-Analyse von Chen untersuchte weiterhin den Einfluss der Feldstärke und konnte eine hohe Sensitivität und Spezifität für 3 Tesla-MRT bei DWI Sequenzen zeigen. In der hier durchgeführten Meta-Regression zeigte sich ein signifikanter negativer Einfluss der Feldstärke 3 Tesla auf die Sensitivität. Betrachtet man die Ergebnisse der Einzelstudien, findet man einige Studien, die mit einer Feldstärke von 3 Tesla zu einer Sensitivität um die 90 % kommen (Park et al. 2011). Das beste Ergebnis erzielte die Studie von Porpiglia und Kollegen (Porpiglia et al. 2014). Diese kombinierte eine Feldstärke von 1,5 Tesla mit einer Endorektalspule. Die in dieser Arbeit durchgeführte Subgruppenanalyse kam zu ähnlichen Ergebnissen. Eine Feldstärke von 1,5 Tesla zeigte eine bessere Sensitivität und ein Studienaufbau ohne Endorektalspule senkte die Spezifität. Das lässt vermuten, dass die beste diagnostische Genauigkeit der mp-MRT mit einer Feldstärke von 1,5 Tesla und der

Kombination aus Endorektal- und Oberflächenspule oder einer Feldstärke von 3 Tesla und einer Oberflächenspule erreicht werden kann.

Eine weitere Limitation der Analyse ist der unzureichende Referenzstandard. Die TRUS gestützte Biopsie wird zwar von der Europäischen Gesellschaft für Urologie (EAU) (Heidenreich et al. 2014) als Goldstandard empfohlen, aber sie weist keinen ausreichend hohen negativ-prädiktiven Wert auf, weil einige Tumore bei der ersten Biopsie übersehen werden (Roehl et al. 2002). Weiterhin ist es ein bekanntes Problem, dass die Tumordetektionsrate in Wiederholungsbiopsien sogar noch niedriger ist (Maxeiner et al. 2014). Die hier eingeschlossenen Primärstudien untersuchten ebenfalls teilweise Patienten mit einer oder mehr vorangegangenen negativen Biopsien. Deshalb ist es möglich, dass einige der falsch positiven Ergebnisse in der MRT in Wahrheit richtig positive Ergebnisse waren und nur vom Referenzstandard übersehen wurden, was zu einer Unterschätzung der Sensitivität führen würde. Weiterhin könnte dies auch einen Teil der Heterogenität zwischen den Studien erklären. Die TRUS gestützte Biopsie scheint nicht der ideale Referenzstandard für diese Patientengruppe zu sein, wird es aber in Ermangelung von Alternativen, da eine radikale Prostataektomie nur für Patienten mit bereits positiver Biopsie in Frage kommt, bleiben. Dafür konnte in der Subgruppenanalyse eine verbesserte Spezifität für Patienten mit mindestens einer vorangegangenen negativen Biopsie gezeigt werden.

Chen und Kollegen (Chen et al. 2014) bemerkten eine höhere Sensitivität in Studien mit Patienten, die einen PSA-Wert > = 20 n g/ml aufwiesen. Wenn man bedenkt, dass keiner der Patienten aus den hier eingeschlossenen Studien so hohe PSA-Werte aufwies, könnte das auch die limitierte Sensitivität erklären. Außerdem erbrachte die dazu durchgeführte Subgruppenanalyse die gleiche Tendenz. Ein PSA-Wert > = 9ng/ml ging mit höherer Sensitivität und höherer Spezifität einher. Auch wenn es nicht auf Signifikanzniveau gezeigt werden konnte, so scheint der PSA-Wert auch zur Heterogenität zwischen den Studien beigetragen zu haben.

Insgesamt zeichnet sich hier die Tendenz ab, dass die mp-MRT hilfreich bei Patienten mit vorangegangener negativer Biopsie, aber weiterhin erhöhten PSA-Werten und nach wie vor bestehendem Tumorverdacht sein könnte, um die Biopsie zu stützen, wie von der ESUR-Leitlinie (Barentsz et al. 2012) empfohlen oder, um gezielt in der MRT auffällige Areale zu

bioptieren. Weiterhin kommen weiterführende aktuelle Studien zu dem Ergebnis, dass die mp-MRT von großem Nutzen zur Risikostratifizierung (Thompson et al. 2014) und zum Follow-Up bei Patienten mit einem niedrig malignen Karzinom unter aktiver Beobachtung zu sein scheint (Arumainayagam et al. 2013).

In dieser Arbeit wurde die Signifikanz der Tumore nicht mit betrachtet. Es ist bekannt, dass die mp-MRT in der Lage ist, sehr kleine und wenig aggressive Tumore in frühen Stadien zu entdecken, die nie eine klinische Relevanz erlangen oder den Patienten in seiner Lebensqualität und Lebensdauer beschränken werden. Das ist vor allem bei niedrigen Grenzwerten wie einem niedrigen PI-RADS-Score der Fall. Dafür zeichnet sich die mp-MRT bei höheren Grenzwerten, wie z. B. einem PI-RADS-Score von > = 4 mit einer geringeren Sensitivität für niedrig maligne Tumore, aber einer sehr hohen Detektionsrate von klinisch signifikanten und hoch malignen Tumoren, aus (Thompson et al. 2014). Für die sensitive und spezifische Diagnostik von signifikanten hoch malignen Prozessen an der Prostata als auch zur Risikostratifizierung scheint die mp-MRT das am besten geeignetste Verfahren zu sein, wie eine Studie aus dem Jahr 2011 bestätigt (Rouse et al. 2011). Eine weitere Studie aus dem Jahr 2015 fand eine diagnostische Genauigkeit von 100 % bei mp-MRT für klinisch signifikante Tumore (Pepe et al. 2015b).

Die klinisch insignifikanten Tumore werden nach erfolgter Diagnostik mittels aktiver Beobachtung (Active Surveillance) weiter kontrolliert, um erst bei deutlichem Progress eine Therapie einzuleiten. Die mp-MRT eignet sich zur Selektion der für „Active Surveillance“ geeigneten Patienten (Pelechano Gomez et al. 2014). Eine Studie aus dem Jahr 2012 untersuchte Niedrig-Risiko-Patienten und konnte zeigen, dass die mp-MRT in der Lage ist, versteckte signifikante Tumore in dieser Gruppe zu detektieren, sodass diese in einer gezielten Biopsie gesichert werden konnten (Margel et al. 2012). Für Auswahl, Bestätigung und Kontrollen der Patienten unter „Active Surveillance“ ist die mp-MRT die zuverlässigste nicht invasive Methode (Pelechano Gomez et al. 2014). Für die hier vorliegende Analyse wurden Studien, die Patienten unter „Active Surveillance“ enthielten, ausgeschlossen, da diese schon eine bioptisch gesicherte Karzinomdiagnose hatten.

Das in dieser Arbeit benutzte mathematische Modell für zufällige Effekte ist für diagnostische Genauigkeitsstudien adaptiert worden (Reitsma et al. 2005) und berücksichtigt die hohe Heterogenität zwischen den Studien. Es geht davon aus, dass die Studien normal verteilt

sind. Der Vorteil besteht darin, dass es berücksichtigt, dass sich Sensitivität und Spezifität oft indirekt proportional verhalten. Das konnte man an den hier eingeschlossenen Studien ebenfalls beobachten, wenn man den Grenzwert PI-RADS-Score > = 3 im Vergleich zum PI-RADS-Score > = 4 betrachtete, stieg die Spezifität und sank die Sensitivität (Thompson et al. 2014). Weiterhin kann das Modell mit einer breiten Streuung in Sensitivität und Spezifität umgehen, wie sie auch bei den in dieser Arbeit eingeschlossenen Studien vorlag, in dem es die dichotome Natur der Daten beibehält (Reitsma et al. 2005).

Trotzdem müssen die Ergebnisse dieser Arbeit auf Grund der relativ kleinen Studienzahl und der hohen Heterogenität zwischen den Studien zurückhaltend interpretiert werden. Dies spielt vor allem bei der Subgruppenanalyse eine Rolle. Weiterhin konnten auf Grund der teilweise zu geringen Anzahl nicht alle möglichen Einflussgrößen in der Meta-Regression oder der Subgruppenanalyse betrachtet werden. Aus diesem Grund konnten daraus lediglich Hinweise gewonnen werden. An Hand der Ergebnisse zeichnete sich eine Tendenz ab. Diese sollte aber in weiteren Studien unter standardisierten Studienbedingungen überprüft werden, um sicherere Aussagen treffen zu können.

8. Schlussfolgerung

In dieser Arbeit konnte eine gute Spezifität und eine akzeptable Sensitivität für die diagnostische Genauigkeit der multiparametrischen MRT, bestehend aus den Sequenzen T2, DWI und DCE, bei Prostatakrebs gezeigt werden. Die AUC von 85 deutete auf eine gute diagnostische Genauigkeit hin. Allerdings war die Heterogenität zwischen den Studien sehr hoch und die Streuung von Sensitivität und Spezifität in den einzelnen Studien ebenfalls, weshalb die Ergebnisse vorsichtig interpretiert werden sollten. Ebenso zeigten die Studien Unterschiede bei der Verwendung von Endorektalspulen und der verwendeten Feldstärke. Diesbezüglich sollte ein einheitlicher technischer Standard zur Durchführung der Magnetresonanztomographie gefunden werden. Am günstigsten schien die Kombination von 1,5 Tesla-Geräten mit einer Endorektal- und einer Oberflächenspule oder einer Feldstärke von 3 Tesla nur mit einer Oberflächenspule zu sein, was auch dem Tenor in der aktuellen Literatur entspricht. Außerdem sollte eine Vereinheitlichung des PI-RADS-Summenscores erfolgen und ein einheitlicher Grenzwert gefunden werden, ab welchem PI-RADS-Summenscore die multiparametrische MRT als positiv zu werten ist. Damit wären die Studien besser vergleichbar, die Ergebnisse zuverlässiger und es könnten bessere Einschätzungen erfolgen. Insgesamt zeichnete sich ein Vorteil der Kombination aus T2, DWI und DCE bei der gestützten Biopsie für Patienten mit bereits vorangegangener negativer Biopsie und weiterhin erhöhtem PSA-Wert ab.

9. Literatur- und Quellenverzeichnis

Abd-Alazeez M, Kirkham A, Ahmed HU, Arya M, Anastasiadis E, Charman SC, Freeman A, Emberton M. 2014. Performance of multiparametric MRI in men at risk of prostate cancer before the first biopsy: a paired validating cohort study using template prostate mapping biopsies as the reference standard. Prostate Cancer Prostatic Dis, 17 (1):40-46.

ACR, American College of Radiology 2015. PI---RADS v2, Prostate Imaging and Report and Data System: Version 2 [Webpage]. https://www.acr.org/-/media/ACR/Files/RADS/Pi-RADS/PIRADS-V2.pdf, retrieved on May 24th 2015.

Arumainayagam N, Ahmed HU, Moore CM, Freeman A, Allen C, Sohaib SA, Kirkham A, van der Meulen J, Emberton M. 2013. Multiparametric MR imaging for detection of clinically significant prostate cancer: a validation cohort study with transperineal template prostate mapping as the reference standard. Radiology, 268 (3):761-769.

Barentsz JO, Richenberg J, Clements R, Choyke P, Verma S, Villeirs G, Rouviere O, Logager V, Futterer JJ, European Society of Urogenital R. 2012. ESUR prostate MR guidelines 2012. Eur Radiol, 22 (4):746-757.

Beyersdorff D, Darsow U, Stephan C, Schnorr D, Loening S, Taupitz M. 2003. MRI of prostate cancer using three different coil systems: image quality, tumor detection, and staging. Rofo, 175 (6):799-805.

Bjurlin MA, Meng X, Le Nobin J, Wysock JS, Lepor H, Rosenkrantz AB, Taneja SS. 2014. Optimization of Prostate Biopsy: The Role of MRI Targeted Biopsy in Detection, Localization, and Risk Assessment. J Urol.

Busetto GM, De Berardinis E, Sciarra A, Panebianco V, Giovannone R, Rosato S, D'Errigo P, Di Silverio F, Gentile V, Salciccia S. 2013. Prostate cancer gene 3 and multiparametric magnetic resonance can reduce unnecessary biopsies: decision curve analysis to evaluate predictive models. Urology, 82 (6):1355-1360.

Cancer Research UK 2018. Prostate cancer survival statistics [Webpage]. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/prostate-cancer/survival, retrieved on Feb 04th 2020.

Chen J, Liu R, Tan P. 2014. The value of diffusion-weighted imaging in the detection of prostate cancer: a meta-analysis. European Radiology, 24 (8):1929-1941.

Chu HT, Guo HF, Zhou YJ. 2010. Bivariate Random Effects Meta-Analysis of Diagnostic Studies Using Generalized Linear Mixed Models. Medical Decision Making, 30 (4):499-508.

de Rooij M, Hamoen EHJ, Futterer JJ, Barentsz JO, Rovers MM. 2014. Accuracy of Multiparametric MRI for Prostate Cancer Detection: A Meta-Analysis. American Journal of Roentgenology, 202 (2):343-351.

Deeks JJ, Macaskill P, Irwig L. 2005. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J Clin Epidemiol, 58 (9):882-893.

Dikaios N, Alkalbani J, Sidhu HS, Fujiwara T, Abd-Alazeez M, Kirkham A, Allen C, Ahmed H, Emberton M, Freeman A, Halligan S, Taylor S, Atkinson D, Punwani S. 2014. Logistic regression model for diagnosis of transition zone prostate cancer on multi-parametric MRI. Eur Radiol.

Ferda J, Kastner J, Hora M, Hes O, Finek J, Topolcan O, Kreuzberg B. 2013. A role of multifactorial evaluation of prostatic 3T MRI in patients with elevated prostatic-specific antigen levels: prospective comparison with ultrasound-guided transrectal biopsy. Anticancer Res, 33 (6):2791-2795.

Ferlay J, Colombet M, Soerjomataram I, Dyba T, Randi G, Bettio M, Gavin A, Visser O, Bray F. 2018. Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. Eur J Cancer, 103:356-387.

Fiard G, Hohn N, Descotes JL, Rambeaud JJ, Troccaz J, Long JA. 2013. Targeted MRI-guided prostate biopsies for the detection of prostate cancer: initial clinical experience with real-time 3-dimensional transrectal ultrasound guidance and magnetic resonance/transrectal ultrasound image fusion. Urology, 81 (6):1372-1378.

Ghafoori M, Alavi M, Aliyari Ghasabeh M. 2013. MRI in prostate cancer. Iran Red Crescent Med J, 15 (12):e16620.

Glass GV. 1976. Primary, Secondary, and Meta-Analysis of Research. Educational Researcher, 5 (10):3-8.

Haffner J, Lemaitre L, Puech P, Haber GP, Leroy X, Jones JS, Villers A. 2011. Role of magnetic resonance imaging before initial biopsy: comparison of magnetic

resonance imaging-targeted and systematic biopsy for significant prostate cancer detection. BJU Int, 108 (8 Pt 2):E171-178.

Hauth E, Hohmuth H, Cozub-Poetica C, Bernand S, Beer M, Jaeger H. 2015. Multiparametric MRI of the prostate with three functional techniques in patients with PSA elevation before initial TRUS-guided biopsy. Br J Radiol, 88 (1054):20150422.

Heidenreich A, Bastian PJ, Bellmunt J, Bolla M, Joniau S, van der Kwast T, Mason M, Matveev V, Wiegel T, Zattoni F, Mottet N, Urology EAo. 2014. EAU guidelines on prostate cancer. part 1: screening, diagnosis, and local treatment with curative intent-update 2013. Eur Urol, 65 (1):124-137.

Higgins JP, Thompson SG. 2002. Quantifying heterogeneity in a meta-analysis. Stat Med, 21 (11):1539-1558.

Ibrahiem EI, Mohsen T, Nabeeh AM, Osman Y, Hekal IA, Abou El-Ghar M. 2012. DWI-MRI: single, informative, and noninvasive technique for prostate cancer diagnosis. ScientificWorldJournal, 2012:973450.

International Agency for Research on Cancer, World Health Organization, Globocan 2018. Prostate Cancer: Estimated Incidence, Mortality and Prevalnece Worldwide in 2018 [Webpage]. https://gco.iarc.fr/today/data/factsheets/populations/908-europe-fact-sheets.pdf, retrieved on Feb 04th 2020.

Iwamoto H, Yumioka T, Yamaguchi N, Inoue S, Masago T, Morizane S, Yao A, Honda M, Sejima T, Takenaka A. 2014. The efficacy of target biopsy of suspected cancer lesions detected by magnetic resonance imaging and/or transrectal ultrasonography during initial prostate biopsies: Comparison of outcomes between two physicians. Yonago Acta Medica, 57 (1):53-58.

Jagannathan D, Indiran V. 2017. Accuracy of Diffusion Weighted Images and MR Spectroscopy in Prostate Lesions - Our Experience with Endorectal Coil on 1.5 T MRI. J Clin Diagn Res, 11 (5):TC10-TC14.

Jambor I, Kahkonen E, Taimen P, Merisaari H, Saunavaara J, Alanen K, Obsitnik B, Minn H, Lehotska V, Aronen HJ. 2014. Prebiopsy multiparametric 3T prostate MRI in patients with elevated PSA, normal digital rectal examination, and no previous biopsy. J Magn Reson Imaging.

Javali TD, Dwivedi DK, Kumar R, Jagannathan NR, Thulkar S, Dinda AK. 2014. Magnetic resonance spectroscopy imaging-directed transrectal ultrasound biopsy increases prostate cancer detection in men with prostate-specific antigen between 4-10 ng/mL and normal digital rectal examination. Int J Urol, 21 (3):257-262.

Kitajima K, Kaji Y, Fukabori Y, Yoshida K, Suganuma N, Sugimura K. 2010. Prostate cancer detection with 3 T MRI: comparison of diffusion-weighted imaging and dynamic contrast-enhanced MRI in combination with T2-weighted imaging. J Magn Reson Imaging, 31 (3):625-631.

Kumar V, Jagannathan NR, Kumar R, Thulkar S, Gupta SD, Hemal AK, Gupta NP. 2007. Transrectal ultrasound-guided biopsy of prostate voxels identified as suspicious of malignancy on three-dimensional (1)H MR spectroscopic imaging in patients with abnormal digital rectal examination or raised prostate specific antigen level of 4-10 ng/ml. NMR Biomed, 20 (1):11-20.

Kuru TH, Roethke MC, Seidenader J, Simpfendorfer T, Boxler S, Alammar K, Rieker P, Popeneciu VI, Roth W, Pahernik S, Schlemmer HP, Hohenfellner M, Hadaschik BA. 2013. Critical evaluation of magnetic resonance imaging targeted, transrectal ultrasound guided transperineal fusion biopsy for detection of prostate cancer. J Urol, 190 (4):1380-1386.

Labanaris AP, Engelhard K, Zugor V, Nutzel R, Kuhn R. 2010. Prostate cancer detection using an extended prostate biopsy schema in combination with additional targeted cores from suspicious images in conventional and functional endorectal magnetic resonance imaging of the prostate. Prostate Cancer Prostatic Dis, 13 (1):65-70.

Margel D, Yap SA, Lawrentschuk N, Klotz L, Haider M, Hersey K, Finelli A, Zlotta A, Trachtenberg J, Fleshner N. 2012. Impact of multiparametric endorectal coil prostate magnetic resonance imaging on disease reclassification among active surveillance candidates: a prospective cohort study. J Urol, 187 (4):1247-1252.

Matsuoka Y, Numao N, Saito K, Tanaka H, Kumagai J, Yoshida S, Koga F, Masuda H, Kawakami S, Fujii Y, Kihara K. 2014. Combination of diffusion-weighted magnetic resonance imaging and extended prostate biopsy predicts lobes without significant cancer: application in patient selection for hemiablative focal therapy. Eur Urol, 65 (1):186-192.

Maxeiner A, Fischer T, Stephan C, Cash H, Slowinski T, Kilic E, Durmus T. 2014. Real-time MRI/US fusion-guided biopsy improves detection rates of prostate cancer in pre-biopsied patients. Aktuelle Urol, 45 (3):197-203.

Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. Open Med, 3 (3):e123-130.

Nazir B. 2014. Pain during transrectal ultrasound-guided prostate biopsy and the role of periprostatic nerve block: what radiologists should know. Korean J Radiol, 15 (5):543-553.

Numao N, Yoshida S, Komai Y, Ishii C, Kagawa M, Kijima T, Yokoyama M, Ishioka J, Matsuoka Y, Koga F, Saito K, Masuda H, Fujii Y, Kawakami S, Kihara K. 2013. Usefulness of pre-biopsy multiparametric magnetic resonance imaging and clinical variables to reduce initial prostate biopsy in men with suspected clinically localized prostate cancer. J Urol, 190 (2):502-508.

Park BK, Park JW, Park SY, Kim CK, Lee HM, Jeon SS, Seo SI, Jeong BC, Choi HY. 2011. Prospective evaluation of 3-T MRI performed before initial transrectal ultrasound-guided prostate biopsy in patients with high prostate-specific antigen and no previous biopsy. AJR Am J Roentgenol, 197 (5):W876-881.

Pelechano Gomez P, Casanova Ramon-Borja J, Collado Serra A. 2014. An update on imaging techniques to optimize active surveillance in prostate cancer. Arch Esp Urol, 67 (5):473-485.

Pepe P, Garufi A, Priolo G, Pennisi M. 2015a. Can 3-Tesla pelvic phased-array multiparametric MRI avoid unnecessary repeat prostate biopsy in patients with PSA < 10 ng/mL? Clin Genitourin Cancer, 13 (1):e27-30.

Pepe P, Dibenedetto G, Garufi A, Priolo G, Pennisi M. 2015b. Multiparametric pelvic MRI accuracy in diagnosing clinically significant prostate cancer in the reevaluation of biopsy microfocal tumor. Anticancer Res, 35 (1):395-399.

Petrillo A, Fusco R, Setola SV, Ronza FM, Granata V, Petrillo M, Carone G, Sansone M, Franco R, Fulciniti F, Perdona S. 2014. Multiparametric MRI for prostate cancer detection: performance in patients with prostate-specific antigen values between 2.5 and 10 ng/mL. J Magn Reson Imaging, 39 (5):1206-1212.

Pokorny MR, de Rooij M, Duncan E, Schroder FH, Parkinson R, Barentsz JO, Thompson LC. 2014. Prospective Study of Diagnostic Accuracy Comparing Prostate Cancer Detection by Transrectal Ultrasound-Guided Biopsy Versus Magnetic Resonance (MR) Imaging with Subsequent MR-guided Biopsy in Men Without Previous Prostate Biopsies. Eur Urol.

Popita C, Popita AR, Sitar-Taut A, Petrut B, Fetica B, Coman I. 2017. 1.5-Tesla Multiparametric-Magnetic Resonance Imaging for the detection of clinically significant prostate cancer. Clujul Med, 90 (1):40-48.

Porpiglia F, Russo F, Manfredi M, Mele F, Fiori C, Bollito E, Papotti M, Molineris I, Passera R, Regge D. 2014. The Roles of Multiparametric Magnetic Resonance Imaging, PCA3 and Prostate Health Index-Which is the Best Predictor of Prostate Cancer after a Negative Biopsy? Results of a Prospective Study. J Urol, 192 (1):60-66.

Portalez D, Mozer P, Cornud F, Renard-Penna R, Misrai V, Thoulouzan M, Malavaud B. 2012. Validation of the European Society of Urogenital Radiology Scoring System for Prostate Cancer Diagnosis on Multiparametric Magnetic Resonance Imaging in a Cohort of Repeat Biopsy Patients. Eur Urol.

Pummer K, Rieken M, Augustin H, Gutschi T, Shariat SF. 2013. Innovations in diagnostic imaging of localized prostate cancer. World J Urol.

Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. 2005. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol, 58 (10):982-990.

Riley RD, Thompson JR, Abrams KR. 2008. An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown. Biostatistics, 9 (1):172-186.

Roehl KA, Antenor JA, Catalona WJ. 2002. Serial biopsy results in prostate cancer screening study. J Urol, 167 (6):2435-2439.

Rosenkrantz AB, Mussi TC, Borofsky MS, Scionti SS, Grasso M, Taneja SS. 2012. 3.0 T multiparametric prostate MRI using pelvic phased-array coil: Utility for tumor detection prior to biopsy. Urol Oncol.

Rothke M, Blondin D, Schlemmer HP, Franiel T. 2013. [PI-RADS classification: structured reporting for MRI of the prostate]. Rofo, 185 (3):253-261.

Rouse P, Shaw G, Ahmed HU, Freeman A, Allen C, Emberton M. 2011. Multi-parametric magnetic resonance imaging to rule-in and rule-out clinically important prostate cancer in men at risk: a cohort study. Urol Int, 87 (1):49-53.

Schimmoller L, Quentin M, Arsov C, Hiester A, Kropil P, Rabenalt R, Albers P, Antoch G, Blondin D. 2014. Predictive power of the ESUR scoring system for prostate cancer diagnosis verified with targeted MR-guided in-bore biopsy. Eur J Radiol, 83 (12):2103-2108.

Schuetz GM, Schlattmann P, Dewey M. 2012. Use of 3x2 tables with an intention to diagnose approach to assess clinical performance of diagnostic tests: meta-analytical evaluation of coronary CT angiography studies. British Medical Journal, 345.

Schwarzmüller-Erber G, Silberstein E. 2010. Angewandte Magnetresonanztomographie: Grundlagen und Anwendungen. Wien: Facultas Verlags- und Buchhandels AG.

Sciarra A, Panebianco V, Salciccia S, Cattarino S, Lisi D, Gentilucci A, Alfarone A, Mariotti G, Passariello R, Gentile V. 2011. Modern role of magnetic resonance and spectroscopy in the imaging of prostate cancer. Urologic Oncology: Seminars and Original Investigations, 29 (1):12-20.

Sciarra A, Panebianco V, Ciccariello M, Salciccia S, Cattarino S, Lisi D, Gentilucci A, Alfarone A, Bernardo S, Passariello R, Gentile V. 2010. Value of magnetic resonance spectroscopy imaging and dynamic contrast-enhanced imaging for detecting prostate cancer foci in men with prior negative biopsy. Clin Cancer Res, 16 (6):1875-1883.

Sonn GA, Chang E, Natarajan S, Margolis DJ, Macairan M, Lieu P, Huang JT, Dorey FJ, Reiter RE, Marks LS. 2014. Value of Targeted Prostate Biopsy Using Magnetic Resonance-Ultrasound Fusion in Men with Prior Negative Biopsy and Elevated Prostate-specific Antigen. European Urology, 65 (4):809-815.

Tamada T, Sone T, Higashi H, Jo Y, Yamamoto A, Kanki A, Ito K. 2011. Prostate cancer detection in patients with total serum prostate-specific antigen levels of 4-10 ng/mL: diagnostic efficacy of diffusion-weighted imaging, dynamic contrast-enhanced MRI, and T2-weighted imaging. AJR Am J Roentgenol, 197 (3):664-670.

Taneja SS. 2004. Imaging in the diagnosis and management of prostate cancer. Rev Urol, 6 (3):101-113.

Tanimoto A, Nakashima J, Kohno H, Shinmoto H, Kuribayashi S. 2007. Prostate cancer screening: the clinical value of diffusion-weighted imaging and dynamic MR imaging in combination with T2-weighted imaging. J Magn Reson Imaging, 25 (1):146-152.

Thompson JE, Moses D, Shnier R, Brenner P, Delprado W, Ponsky L, Pulbrook M, Bohm M, Haynes AM, Hayen A, Stricker PD. 2014. Multiparametric Magnetic Resonance Imaging Guided Diagnostic Biopsy Detects Significant Prostate Cancer and Could Reduce Unnecessary Biopsies and Over Detection: A Prospective Study. J Urol.

Vamvakas EC. 1998. Meta-analyses of studies of the diagnostic accuracy of laboratory tests: a review of the concepts and methods. Arch Pathol Lab Med, 122 (8):675-686.

Vamvakas EC. 2001. Applications of meta-analysis in pathology practice. Am J Clin Pathol, 116 Suppl:S47-64.

Van Houwelingen HC, Zwinderman KH, Stijnen T. 1993. A bivariate approach to meta-analysis. Stat Med, 12 (24):2273-2284.

Villeirs GM, Oosterlinck W, Vanherreweghe E, De Meerleer GO. 2010. A qualitative approach to combined magnetic resonance imaging and spectroscopy in the diagnosis of prostate cancer. Eur J Radiol, 73 (2):352-356.

Wan X, Wang W, Liu J, Tong T. 2014. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Med Res Methodol, 14 (1):135.

Wang X, Wang JY, Li CM, Zhang YQ, Wang JL, Wan B, Zhang W, Chen M, Li SY, Wan G, Liu M. 2016. Evaluation of the Prostate Imaging Reporting and Data System for Magnetic Resonance Imaging Diagnosis of Prostate Cancer in Patients with Prostate-specific Antigen <20 ng/ml. Chin Med J (Engl), 129 (12):1432-1438.

Watanabe Y, Terai A, Araki T, Nagayama M, Okumura A, Amoh Y, Ishimori T, Ishibashi M, Nakashita S, Dodo Y. 2012. Detection and localization of prostate cancer with the targeted biopsy strategy based on ADC map: a prospective large-scale cohort study. J Magn Reson Imaging, 35 (6):1414-1421.

Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM, Group Q-. 2011. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med, 155 (8):529-536.

Wu L-M, Xu J-R, Ye Y-Q, Lu Q, Hu J-N. 2012. The Clinical Value of Diffusion-Weighted Imaging in Combination With T2-Weighted Imaging in Diagnosing Prostate Carcinoma: A Systematic Review and Meta-Analysis. American Journal of Roentgenology, 199 (1):103-110.

Yoshizako T, Wada A, Hayashi T, Uchida K, Sumura M, Uchida N, Kitagaki H, Igawa M. 2008. Usefulness of diffusion-weighted imaging and dynamic contrast-enhanced magnetic resonance imaging in the diagnosis of prostate transition-zone cancer. Acta Radiol, 49 (10):1207-1213.

Youden WJ. 1950. Index for rating diagnostic tests. Cancer, 3 (1):32-35.

Zhen L, Liu X, Yegang C, Yongjiao Y, Yawei X, Jiaqi K, Xianhao W, Yuxuan S, Rui H, Wei Z, Ningjing O. 2019. Accuracy of multiparametric magnetic resonance imaging for diagnosing prostate Cancer: a systematic review and meta-analysis. BMC Cancer, 19 (1):1244.

10. Anhang:

10.1 Ehrenwörtliche Erklärung

Hiermit erkläre ich, dass mir die Promotionsordnung der Medizinischen Fakultät der Friedrich-Schiller-Universität bekannt ist,

ich die Dissertation selbst angefertigt habe und alle von mir benutzten Hilfsmittel, persönlichen Mitteilungen und Quellen in meiner Arbeit angegeben sind,

mich folgende Personen bei der Auswahl und Auswertung des Materials sowie bei der Herstellung des Manuskripts unterstützt haben:
Prof. Univ.-Prof. Dr. med. habil., M.Sc. Applied Statistics Peter Schlattmann,
Prof. Dr. med. Marc Dewey,
Assoc. Prof. Priv.-Doz. Dr. Pascal Baltzer,
Dr. med. Sebastian Nagel,

die Hilfe eines Promotionsberaters nicht in Anspruch genommen wurde und dass Dritte weder unmittelbar noch mittelbar geldwerte Leistungen von mir für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen,

dass ich die Dissertation noch nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht habe und

dass ich die gleiche, eine in wesentlichen Teilen ähnliche oder eine andere Abhandlung nicht bei einer anderen Hochschule als Dissertation eingereicht habe.

Chemnitz, 02. Juli 2020 Christina Vargas Mora

10.2 Tabellen- und Bildverzeichnis

Abb. 1: Flussdiagramm zur Studienauswahl 18
Abb. 2: Funnel-Plot zum Publikationsbias 22
Abb. 3: Positives und negatives Likelihood-Ratio 23
Abb. 4: sROC-Kurve 25
Abb. 5: Studienspezifische Sensitivität 26
Abb. 6: Studienspezifische Spezifität 27
Abb. 7: Mögliche Einflussfaktoren auf Sensitivität und Spezifität 30
Abb. 8: Verhältnis der Studien nach Bewertung zum Risiko für Bias (%) 32
Abb. 9: Verhältnis der Studien nach Bewertung bezüglich der Anwendbarkeit (%) 33
Abb. 10: Verhältnis der Studien nach Beantwortung der Signalfragen (%) 34

Tab. 1: Charakteristika der Studienpopulation 19
Tab. 2: Charakteristika der multiparametrischen MRT 20
Tab. 3: Charakteristika des Referenzstandards (systematische Biopsie) 21
Tab. 4: Gepoolte Schätzer 23
Tab. 5: Studienspezifische Schätzer 24
Tab. 6: Meta-Regression - Faktoren, die möglicherweise Heterogenität erklären 28
Tab. 7: Mögliche Einflussfaktoren auf Sensitivität und Spezifität 29
Tab. 8: Subgruppenanalyse 31
Tab. 9: Bewertungskriterien QUADAS-2 – Bereich 1 35
Tab. 10: Bewertungskriterien QUADAS-2 – Bereich 2-3 36
Tab. 11: Bewertungskriterien QUADAS-2 – Bereich 4 37

10.3 Protokoll

Protokoll für ein Systematisches Review- eine Meta-Analyse

1. Studiensynopsis

1.1 Antragsteller:

Vargas Mora, Christina

1.2 Studientitel:

Diagnostische Genauigkeit der multiparametrischen MRT (mp-MRT) bei der Diagnostik des Prostatakarzinoms – eine Meta-Analyse

1.3 Voraussetzung:

Patienten mit suspekt erhöhten PSA-Werten (> 4ng/ml) und begründetem Verdacht auf ein Prostatakarzinom

1.4 Zielstellung:

Das systematische Review und Meta-Analyse soll die diagnostische Genauigkeit und den klinischen Wert der multiparametrischen MRT zur nicht-invasiven Diagnostik des Prostatakarzinoms untersuchen.

1.5 Tests:

Indextest: T2-weighted MRT (T2-MRT) und dynamic contrast enhanced MRT (DCE-MRT) und diffusion-weighted MRT (DWI-MRT) und/ oder magnetic resonance sprectroscopic imaging (MRSI) zur Diagnostik von bestehendem oder nicht bestehendem Prostatakarzinom

Referenz-Test: histopathologische Ergebnisse aus systematischer Biopsie oder nach Prostatektomie

1.6 Ein- und Ausschlusskriterien

Studienpopulation: Männer, mit begründetem Verdacht auf Prostatakarzinom

Index-Test: T2-MRT, DCE-MRT, DWI-MRT und/oder MRSI Sequenz
Referenzstandard/ Vergleichstest: Der Vergleich aus Kombination aus T2-MRT, DCE-MRT, DWI-MRT und/ oder MRSI mit den histopathologischen Ergebnissen aus Prostatektomie und/oder Biopsie (mindestens 10 Stanzen, und/ oder gestützt durch transrektalem Ultraschall (TRUS) und verglichen mit nur T2-MRT
Ergebnis: Diagnostische Genauigkeit der mp-MRT bei der Diagnostik des Prostatakarzinoms verglichen mit dem Referenzstandard und nur T2-MRI

1.7. Suchstrategie:

Potentielle Studien werden in Isl Web of Knowledge, Medline (via PubMed), Embase (via Ovid) und Scopus gesucht.

1.8 Qualitätsbeurteilung:

Zur Beurteilung der methodischen Studienqualität soll das QUADAS-2 Tool genutzt werden.

1.9 Datenerhebung:

Zwei unabhängige Sucher werden die Daten aus den Studien erheben.

1.10 Datenberechnung:

Es wird eine exakte Binomialverteilung (Chu et al. 2006) des bivariaten Mixed-Effects-Regression-Models (entwickelt von Van Houwelingen (Van Houwelingen et al. 1993, Van Houwelingen et al. 2002)) zu Grunde gelegt.

1.11 Stichprobengröße:

Es werden 5 bis15 Ausgangsstudien erwartet, basierend auf vorhergehenden Meta-Analysen und ersten Datenbankrecherchen.

1.12 Dauer:

Januar 2012 – Dezember 2014

1.13 Stichworte zur Datenbankrecherche:

Prostate cancer, hyperplasia, carcinoma, prostatic neoplasm, diagnostic accuracy, multiparametric MRI, T2-weighted MRI, diffusion weighted MRI, DWI, dynamic contrast enhanced MRI, DCE-MRI, magnetic resonance spectroscopic imaging, MRSI

2. Medizinische Fragestellung

2.1 Evidenz

Das Prostatakarzinom ist eine der häufigsten Tumorentitäten in Deutschland. Mittels digital-rektaler Untersuchung ist ein erfahrener Untersucher auf Grund der typischen Oberflächenbeschaffenheit in der Lage einen Verdacht zu äußern.

Vor allem in der vorderen Zone der Prostata werden Tumore übersehen, sodass häufig eine Diagnosestellung erst im fortgeschrittenen Stadium erfolgt. Auch das prostataspezifische Antigen (PSA) wird zu Diagnosestellung herangezogen.

Es kann aber auch bei Entzündungen der Prostata erhöht sein.

Zur Diagnosesicherung wird aktuell die Biopsie benutzt, aber auch hier kommt es zu falsch negative Ergebnissen, wenn Stanzen aus der falschen Stelle oder zu oberflächlich entnommen werden, was weitere schmerzhafte Biopsien nach sich zieht. Es werden auch ultraschall- und neuerdings MRT-gestützte Biopsien aus in der Bildgebung auffälligen Arealen durchgeführt, um bessere Ergebnisse zu erzielen. Es bleibt aber eine invasive Methode. Ein rein nichtinvasives Verfahren wäre also wünschenswert. Mit transrektalem Ultraschall können Tumore mit einem Durchmesser von 10 mm sicher diagnostiziert werden, kleinere nur in 20 Prozent der Fälle. Damit liefert er aber bessere Ergebnisse als PSA-Wert und digital-rektale Untersuchung. Die MRT liefert vergleichbare Ergebnisse.

Mit der Entwicklung neuerer Sequenzen wie DCE-MRT, DWI-MRT, MRSI ist es interessant zu untersuchen, ob mittels Kombination der Sequenzen eine bessere diagnostische Genauigkeit als nur mit T2-MRT oder mit Biopsie erreicht werden kann.

2.2 Notwendigkeit eines Systematischen Reviews – einer Meta-Analyse

Prostatakarzinom ist die in Deutschland am häufigsten diagnostizierte Tumorentität und die dritthäufigste Todesursache bei Männern. 22 % der jährlichen Krebsdiagnosen gehen auf Prostatakarzinome zurück. Es erkranken jährlich 44.000 Männer, davon sterben 11.000. Es ist wichtig das Prostatakarzinom in einem frühen Stadium zu diagnostizieren, da die Heilungschancen, wenn der Tumor die Kapsel noch nicht durchbrochen hat, bei 90 % liegen, sonst nur bei ca. 50 %.

Der aktuell beste Weg zur Diagnose ist die Biopsie. Nicht selten sind die Ergebnisse falsch negativ, da Stanzen zu oberflächlich oder an der falschen Stelle genommen werden. Dies macht zusätzliche schmerzhafte Biopsien nötig. Deshalb ist eine nichtinvasive alternative Methode mit hoher Sensitivität und Spezifität wünschenswert. Eine zuverlässige Methode zeichnet sich durch eine minimale Rate falsch negativer Ergebnisse und damit hoher Sensitivität aus. Die mp-MRT ist bezüglich ihrer diagnostischen Genauigkeit bei Prostatakarzinom noch nicht hinreichend untersucht worden. Aus diesem Grund soll hier eine Meta-Analyse erstellt werden, um die per-Patient Genauigkeit und das klinische Potential in der Diagnosestellung des Prostatakarzinoms bei Männern mit Karzinomverdacht bei hohen PSA-Werten und negativem transrektalen Ultraschall oder negativer digital-rektaler Untersuchung verglichen mit der Histopathologie als Referenzstandard. Die Ergebnisse haben Relevanz, da es ökonomische Konsequenzen haben könnte und auch schmerzhafte Wiederholungsbiopsien vermeiden helfen könnte. Weiterhin kann ein Vergleich der verschiedenen mp-MRT Kombinationen helfen, die ökonomischste und sensitivste Sequenzkombination zu identifizieren. Da klinisch mit höchster Relevanz einzustufen, wurde die per-Patient-Analyse gewählt und nicht die per-Region-Analyse.

2.3 Strategien zur Verbreitung der Ergebnisse

Eine Publikation in einem medizinischen Fachjournal soll angestrebt werden. Die Meta-Analyse könnte das Potential haben, klinische Standards in der Diagnostik des Prostatakarzinoms zu beeinflussen. Leitlinien (wie z. B. „Interdisziplinäre Leitlinie der Qualität S3 zur Früherkennung, Diagnose und Therapie der verschiedenen Stadien des Prostatakarzinoms", der „Deutsche Gesellschaft für Urologie e. V." (Deutsche Gesellschaft für Urologie e. V. 2009) könnten eventuell durch die Ergebnisse dieser Meta-Analyse beeinflusst werden.

3. Rechtfertigung der Designaspekte

3.1 Auswahlkriterien für Primärstudien

Als prospektive diagnostische Genauigkeitsstudie sollen prospektive Studien mit Männern, mit V. a. Prostatakarzinom, die sich einer mp-MRT im Vergleich zum Referenzstandard unterzogen haben, zusammengefasst werden, um zusammenfassende Schätzgrößen zur diagnostischen Performance der mp-MRT zu ermitteln. Aus diesem Grund werden nur diagnostische Genauigkeitsstudien und keine Studien mit randomisiertem Patientenkollektiv

eingeschlossen. Der Referenzstandard sollte eine Histopathologie aus Prostatektomie oder systematischer Biopsie sein, publiziert in Englisch, Deutsch oder Spanisch.
Weiterhin müssen absolute Zahlen für Richtig Positive (RP) und Richtig Negative (RN), Falsch Positive (FP) und Falsch Negative (FN) auf einer per-Patient-Basis genannt sein oder hinreichende Daten, um absolute Zahlen berechnen und eine 4-Felder-Tafel bilden zu können.
Ausschlusskriterien sind: retrospektives Studiendesign, Tierstudien, Übersichtsarbeiten, Briefwechsel, Fallberichte, Expertenmeinungen und Editorials, Überlappungen mit anderen größeren Studien, MRT-basierte Biopsie, Studien, die potentielle Rezidive untersuchen, Follow-up- oder Staginguntersuchungen.

3.2 Population

Zur Vermeidung von Bias müssen die Studien ein geeignetes Patientenkollektiv einschließen (Männer, mit hochgradigem Verdacht auf Prostatakrebs). Weiterhin sind ein zuverlässiger Referenzstandard wie die histopathologischen Ergebnisse aus systematischer Biopsie oder Prostatektomie notwendig, um Sensitivität und Spezifität der Methode ermitteln zu können.

3.3 Intervention

Es soll untersucht werden, ob eine Kombination der MRT Sequenzen die diagnostische Performance verbessert und wie viele bzw. welche Sequenzen kombiniert werden sollten, wie z. B. T2-MRT und DCE-MRT und DWI-MRT und/ oder MRSI.

3.4 Vergleichsgrößen

Es erfolgt ein Vergleich verschiedener Kombinationen von MRT Sequenzen, wie T2-MRT und DCE-MRT, DWI-MRT und/ oder MRSI mit den histopathologischen Ergebnissen aus radikaler Prostatektomie oder systematischer Biopsie (mit mindestens 10 Stanzen, oder ultraschallgestützter Biopsie, um eine ausreichende Sicherheit und Sensitivität des Referenzstandards zu gewährleisten). Weiterhin soll geschaut werden, ob die diagnostische Genauigkeit der mp-MRT, der der T2-MRT überlegen ist. Die histopathologischen Ergebnisse werden der Referenzstandard sein, um die diagnostische Genauigkeit der mp-MRT beurteilen zu können.

3.5 Ergebnisse

Untersucht wird die Diagnostische Genauigkeit der mp-MRT (T2-MRT, DCE-MRT, DWI-MRT und/ oder MRSI). Um aus den Ergebnissen Empfehlungen für die klinische Entscheidungsfindung ableiten zu können, wird die „Post-Test Probability“ für Populationen mit unterschiedlichen Prävalenzen („Pre-Test Probabilities“) mittels Bayes Theorem berechnet.

3.6 Design der Primärstudien

Es wird eine Feldstärke von wenigstens 1,5 Tesla gefordert und eine Sequenzkombination aus T2-MRT, DCE-MRT, DWI-MRT und/ oder MRSI. Die statistische Analyse sollte auf einem per-Patient-Level und nicht auf einem per-Läsion- oder per-Region-Level erfolgen. Das Patientenkollektiv sollte Männer mit hochgradigem Verdacht auf ein Prostatakarzinom umfassen. Der Abstand zwischen vorangegangener Prostatabiopsie und mp-MRT sollte mindestens 6 Wochen betragen.

Die Anforderung an die mp-MRT Sequenzen sollten folgende sein:

- T2-MRT: 1,5 Tesla mit Benutzung einer Endorektal- oder einer externen Oberflächenspule und coronaren, axialen und sagittalen Schnitten, 3 mm Schichtdicke, oder 3 Tesla mit oder ohne Endorektal- oder Oberflächenspule
- DCE-MRT: semiquantitative Parameter wie integrierte Fläche unter der Gadolinium-Konzentrations-Zeit Kurve, wash-in Gradient, maximale Signalintensität, time-to-peak enhancement, washout und Enhancementbeginn, 1,5 Tesla mit Endorektal- oder Oberflächenspule oder 3 Tesla mit oder ohne Endorektal- oder Oberflächenspule
- DWI-MRT: 1,5 Tesla mit Endorektal- oder Oberflächenspule, b > = 850 sec/mm², ADC (Apparent Diffusion Coefficient), oder 3 Tesla mit oder ohne Endorektal- oder Oberflächenspule
- MRSI: dreidimensionale MR spectroscopic imaging Sequenzen mit einer Auflösung von 0,4 cm^3, mit hinreichender signal-to-noise-Ratio bei 1,5 Tesla mit oder Endorektal- oder Oberflächenspule oder 3 Tesla mit oder ohne Endorektal- oder Oberflächenspule, 3D-multivoxel-hydrogen1-MRSI (Voxel 0,6 - 0,7mm)

3.7 Suchstrategien

Potentielle Studien werden in Isl Web of Knowledge, Medline (via PubMed), Embase (via Ovid) und Scopus gesucht. Zusätzliche methodische Suchfilter für diagnostische Testgenauigkeit werden nicht benutzt, da sie auf Grund der hohen Versäumnisraten nicht

empfohlen sind (Doust et al. 2005, Leeflang et al. 2006, Leeflang et al. 2008). Weiterhin werden die Literaturverzeichnisse der ausgewählten Studien durchsucht. Die Suche und Auswahl der Studien wird von zwei unabhängigen Suchern durchgeführt. Doppelte Ergebnisse werden eliminiert. Zuerst werden Titel und Zusammenfassungen durchgeschaut und basierend darauf, mit Blick auf Ein- und Ausschlusskriterien, Studien aussortiert oder zur weiteren Durchsicht gespeichert. Das Maß für die Interrater-Reliabilität soll Cohens Kappa sein. In der zweiten Phase werden von den verbleibenden Artikeln die Volltexte durchgesehen. Der „Quality of Reporting of Meta-analyses (QUOROM) Guideline“ (Moher et al. 1999) folgend, werden die Schritte der Studienauswahl in einem Flussdiagramm darstellen werden: a) Anzahl der Suchergebnisse je Datenbank, b) Löschen Doppelter Treffer, c) Anzahl ausgeschlossener Studien an Hand des Titels oder der Zusammenfassung, d) Anzahl ausgeschlossener Studien auf Grund nicht erfüllter Einschlusskriterien mit vollständiger Auflistung der Ausschlussgründe jeder Studie in einem separaten Dokument, e) eingeschlossene Studien

3.8 Datenerhebung

Zwei unabhängige Sucher werden die Daten aus den Studien zum Bilden der 4-Felder-Tafel erheben und eine Qualitätsbewertung auf Basis der Volltexte durchführen. Bei Diskrepanzen muss ein Konsensus, gegebenenfalls mit Hilfe eines Dritten, erreicht werden. Sollten Daten fehlen oder es potentielle Überschneidungen geben, werden die zuständigen Autoren per E-Mail kontaktiert werden.

Bei fehlender Antwort, trotz zweitem Anschreiben, wird die Studie ausgeschlossen werden. Die erhobenen Daten werden Folgendes umfassen: generelle und detaillierte methodische Studiencharakteristiken, Charakteristika der Studienpopulation, Details zum Index-Test und dessen technologischer Standards und Details zum Referenzstandard.

3.9 Qualitätsbeurteilung:

Die methodische Studienqualität wird mittels der „Quality Assessment of Diagnostic Accuracy Studies (QUADAS) Checkliste“ (Whiting et al. 2003) überprüft werden. Diese 15 qualitätsbezogenen Punkte sollen helfen, auf mögliche Ursachen für Varianz und Bias zu untersuchen. Es erfolgt eine individuelle Anpassung der Guidelines zu jedem Scoring-Punkt auf unser Review, wie von Whiting empfohlen (Whiting et al. 2006).

Da eine Gesamtbewertung der Studienqualität schwierig ist und so laut Autoren des „QUADAS 2 Tools“ nicht in einer Meta-Analyse erfolgen sollte (Whiting et al. 2005), werden

deren Empfehlungen folgend individuelle qualitätsbezogenen Punkte als Kovariate im Random-Effects-Model untersucht werden, um sie als mögliche Ursache für Heterogenität und Bias zu identifizieren. Dazu soll auch der Einfluss der „QUADAS“ Punkte wie adäquater Referenzstandard und der Test-Review-Bias auf Sensitivität und Spezifität untersucht werden. Weiterhin soll das „Standards for Reporting of Diagnostic Accuracy“ (STARD) Protokoll (Bossuyt et al. 2003) in der Qualitätsanalyse aller eingeschlossenen Studien angewandt werden.

4. Statistische Analyse

Da eine Meta-Analyse eine Beobachtungsstudie ist, wird die Analyse Bias Wahrscheinlichkeiten und Störfaktoren umfassen.

4.1. Ergebnispräsentation und graphische Datenausgabe

Die Analyse wird in Tabellen richtig und falsch positive sowie richtig und falsch negative Ergebnisse für jede Studie auflisten und basierend darauf Sensitivität und Spezifität und das dazugehörige Konfidenzintervall berechnen. Die Ergebnisse werden graphisch in einem Forest-Plot dargestellt. Andere beschreibende Variablen können z. B. Publikationsjahr und Patientencharakteristika sein.

4.2 Publikationsbias

Als erstes werden die Daten in einem Funnel-Plot visualisiert, einem Streudiagramm aus Studiengröße und diagnostischen Odds-Ratios. Eine weitere Analyse des Publikationsbias wird der Egger’s Test sein oder eine andere passende Methode.

4.3 Statistische Methode zur Berechnung des Gesamteffekts

Die bivariate „summary ROC”-Kurven-Analyse wird mit „STATA“ (Version IC 12, StataCorp) und dem „MIDAS“ Module durchgeführt werden. Die Erstellung von Diagrammen und Grafiken wird ebenfalls mittels dieses Programmes erfolgen. Zum Publikationsbias wird ein Funnel-Plot erstellt und ein Regressionstest auf Asymmetrie durchgeführt werden (Deeks et al. 2005).

5. Datenberechnung

Es wird eine exakte Binomialverteilung (Chu et al. 2006) des bivariaten Mixed-Effects Regression Models (entwickelt von Van Houwelingen (Van Houwelingen et al. 1993, Van

Houwelingen et al. 2002)) zu Grunde gelegt. Dieses Model wurde für Therapiestudien-Meta-Analysen entwickelt und für diagnostische Genauigkeits-Tests adaptiert (Reitsma et al 2005, Riley et al. 2007). Dieses bivariate statistische Model wandelt Sensitivität und Spezifität einer Einzelstudie nicht in einen Indikator für die diagnostische Testgenauigkeit, sondern erhält die zweidimensionale Datenstruktur, sodass auch die Korrelation zwischen beiden Berücksichtigung findet.

Basierend auf diesem Model werden mean-logit-Sensitivität und mean-logit-Spezifität mit Standardabweichung und 95 % Konfidenzintervall erhoben, die Interstudienvariabilität von logit-Sensitivität und logit-Spezifität und deren Kovarianz. Diese Ergebnisse werden rücktransformiert in eine „Receiver Operating Curve“(ROC) Skala mit Gesamt-Sensitivität, Gesamt-Spezifität und dem diagnostischen Odds-Ratio. Aus den so berechneten Schätzern wird dann eine „summary ROC” Kurve erstellt für die mp-MRT mit einem „summary Operating Point“ für Sensitivität und Spezifität auf der Kurve und einem 95 % Konfidenzintervallellipsoid.

5.1 Meta-Regression

Wird Heterogenität zwischen den Studien beobachtet, soll die mögliche Ursache dieser mittels einer Meta-Regressionsanalyse aufgedeckt werden.

Dabei werden studienspezifische Kovariablen wie Feldstärke, PSA-Wert oder, ob der Radiologe blind für die Ergebnisse des Referenzstandards war, in der Meta-Regression untersucht.

6. Expertise

	Name	Institut	Rolle
	Marc Dewey	Klinik für Radiologie, Charité-Universitätsklinikum Berlin Charitéplatz 1, 10117 Berlin	Klinische Expertise
	Peter Schlattmann	Institut für Medizinische Statistik, Informatik,	Methodische Expertise

		und Datenwissenschaften FSU Jena Universitätsklinikum Jena Bachstraße 18, 07743 Jena,	

7. Referenzen

Bossuyt PM, Reitsma JB, Bruns DE. 2003. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. Ann Intern Med, 138 (1):40-4.

Chu H, Cole SR. 2006. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. J Clin Epidemiol, 59 (12):1 331-2; author reply 1332-3.

Deeks JJ, Macaskill P, Irwig L. 2005. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J Clin Epidemiol, 58 (9):882-93.

Deutsche Gesellschaft für Urologie e. V. 2009. Interdisziplinäre Leitlinie der Qualität S3 zur Früherkennung, Diagnose und Therapie der verschiedenen Stadien des Prostatakarzinoms. Ärztliches Zentrum für Qualität in der Medizin ÄZQ Gemeinsame Einrichtung von Bundesärztekammer und Kassenärztlicher Bundesvereinigung.

Doust JA, Pietrzak E, Sanders S, Glasziou PP. 2005. Identifying studies for systematic reviews of diagnostic tests was difficult due to the poor sensitivity and precision of methodologic filters and the lack of information in the abstract. J Clin Epidemiol, 58 (5):444-9.

Leeflang MM, Scholten RJ, Rutjes AW, Reitsma JB, Bossuyt PM. 2006. Use of methodological search filters to identify diagnostic accuracy studies can lead to the omission of relevant studies. J Clin Epidemiol, 59 (3):234-40.

Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM. 2008. Systematic reviews of diagnostic test accuracy. Ann Intern Med, 149 (12):889-97.

Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. 1999. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. Lancet, 354 (9193):1896-900.

Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. 2005. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol, 58 (10):982-90.

Riley RD, Abrams KR, Sutton AJ, Lambert PC, Thompson JR. 2007. Bivariate random-effects meta-analysis and the estimation of between-study correlation. BMC Med Res Methodol, 7:3.

Van Houwelingen HC, Zwinderman KH, Stijnen T. 1993. A bivariate approach to meta-analysis. Stat Med, 12 (24):2273-84.

Van Houwelingen HC, Arends LR, Stijnen T. 2002. Advanced methods in meta-analysis: multivariate approach and meta-regression. Stat Med, 21 (4):589-624.

Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. 2003. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol, 3:25.

Whiting PF, Weswood ME, Rutjes AW, Reitsma JB, Bossuyt PN, Kleijnen J. 2006. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. BMC Med Res Methodol, 6:9.

Whiting P, Harbord R, Kleijnen J. 2005. No role for quality scores in systematic reviews of diagnostic accuracy studies. BMC Med Res Methodol, 5:19.

10.4 Danksagung

Besonderen Dank möchte ich meinem Doktorvater Prof. Dr. med. Peter Schlattmann aussprechen für die Überlassung des Themas, die Betreuung der Arbeit und stets freundliche Hilfe sowie für die Einführung in die Welt der statistischen Berechnungen. Weiterhin Danke ich Priv.-Doz. Dr. med. Pascal Baltzer für die persönlichen und radiologischen Ratschläge und die kritische Durchsicht der Arbeit, Prof. Dr. med. Marc Dewey für seine Ideen und Anmerkungen aus radiologischer Sicht und besonders Dr. med. Sebastian Nagel für die Geduld, Zeit, Gründlichkeit und Motivation als zweiter unabhängiger Reviewer. Die Mithilfe und Zeit der eingangs genannten Personen und deren bereitwillige Teilung Ihres Erfahrungsschatzes haben die Anfertigung dieser Promotionsschrift erst möglich gemacht.

Weiterhin möchte ich meiner Familie danken. An erster Stelle meinen Eltern, die mir Ausbildung und Studium ermöglichten und mich stets mit moralischem Beistand und Halt ermuntert und unterstützt haben, meinen Großeltern und meinem Mann. Sie gaben mir die Motivation zur Anfertigung und Vollendung meiner Dissertation.

Ein besonderer Dank gilt auch meiner Freundin Franziska Steinmetz für ihre geduldige und verständnisvolle Unterstützung und die vielen Stunden mit offenem Ohr.