

Simone Schultze
Dr. med.

Evaluation eines bildbasierten Scoringystems für die multiparametrische Magnetresonanztomographie der Prostata anhand der transperinealen Magnetresonanztomographie/Transrektal-Ultraschall-gesteuerten Biopsie

Fach/Einrichtung: DKFZ (Deutsches Krebsforschungszentrum)
Doktorvater: Priv.-Doz. Dr. med. Dipl.-Kfm. Matthias C. Röhke

Das Hauptziel der vorliegenden Arbeit war die Evaluation eines bildbasierten multiparametrischen MRT-Scoringschemas (PI-RADS) in Korrelation mit histopathologischen Ergebnissen aus transperinealen Magnetresonanztomographie/Transrektal-Ultraschall-gesteuerten Biopsien. Dieses Schema beurteilt anhand festgelegter Kriterien in vier verschiedenen multiparametrischen MRT-Sequenzen (T2-Gewichtung, diffusionsgewichtete Bildgebung, dynamische kontrastmittel-verstärkte MRT und Magnetresonanzspektroskopie) die Dignität einer Läsion innerhalb der Prostata und vergibt je nach Dignitätsgradeinschätzung einen Score auf einer Likert-Skala von 1-5. Neben der diagnostischen Güte der Scores der MRT-Einzelsequenzen wurde in dieser Arbeit die diagnostische Wertigkeit von verschiedenen Cut-Off-Scores im multiparametrischen Ansatz in den zweifach, dreifach und vierfach aggregierten Sequenzen untersucht. Die Arbeit unterschied dabei zwischen einem maximal vergebenen Score in der jeweiligen Sequenz-Aggregation als Cut-Off-Punkt und einem aus den Scores der einzelnen Sequenzen gebildeten Summenscore als Grenzwert und verglich beide Ansätze. Zudem wurde die diagnostische Güte des abschließend auf Basis des Gesamteindrucks subjektiv vergebenen Gesamtscores überprüft.

Dreiunddreißig (52%) der insgesamt 64 eingeschlossenen Patienten zeigten ein Prostatakarzinom in der Stanzbiopsie. Die DCE erwies sich für einen Cut-Off-Score von ≥ 3 als sensitivste und gleichzeitig als am wenigsten spezifische Einzelsequenz. Die MRSI war dagegen die spezifischste und gleichzeitig am wenigsten sensitive Einzelsequenz. Die Aggregation mehrerer Sequenzen zu einem multiparametrischen Ansatz führte bei Verwendung eines maximal vergebenen Scores von ≥ 3 als Grenzwert zwar zu einer Sensitivitätssteigerung im Vergleich zur alleinigen T2-gewichteten Sequenz bzw. zu den zweifach aggregierten Sequenzen, ging jedoch mit einem deutlichen Spezifitätsabfall einher. Die Einbeziehung der MRSI in die Aggregation führte sogar zu ausgeprägten Sensitivitätsverlusten. Eine Verwendung höherer oder niedrigerer Grenzwerte erzielte ebenfalls keine ausreichenden Sensitivitäten oder Spezifitäten. Die Aggregation aller vier Modalitäten erzielte für einen maximal vergebenen Score von ≥ 3 als Cut-Off-Wert keine weitere relevante Verbesserung der diagnostischen Güte im Vergleich zu den dreifach aggregierten Modalitäten.

Die Verwendung eines Summenscores als Grenzwert konnte im Vergleich zum maximal vergebenen Score von ≥ 3 als Cut-Off-Wert für die zweifach aggregierten Modalitäten zu einer leichten Steigerung der Spezifität führen, während die Sensitivität entweder unverändert blieb oder leicht abnahm. Für die dreifach aggregierten Sequenzen (T2w+DWI+DCE, T2w+DWI+MRSI, T2w+DCE+MRSI) fanden sich deutlich höhere Spezifitäten bei

Verwendung von Summenscores im Vergleich zum maximal vergebenen Score von ≥ 3 als Cut-Off-Wert. Die Sensitivität sank dabei nur leichtgradig ab, sodass die Verwendung von T2w+DWI+DCE für einen Summenscore von ≥ 9 als Cut-Off-Wert in der vorliegenden Arbeit das statistisch optimale Verhältnis aus Sensitivität (81,8%) und Spezifität (79%) erreichte. Die MRSI ist durch ihre geringe Sensitivität und ihren eingeschränkten diagnostischen Nutzen in Verbindung mit den anderen Sequenzen nicht a priori miteinzubeziehen. Zukünftig sind daher weitere Studien notwendig, um zu prüfen, ob sie unter gewissen technischen und methodischen Voraussetzungen für bestimmte Indikationen ergänzend eingesetzt werden kann.

Die Karzinom-Positivrate des Gesamtscores betrug auf alle evaluierten Areale bezogen für einen Score von 1-2, 3 und 4-5 jeweils 10,5%, 31% und 81,8%. Trotz der hohen negativen und positiven prädiktiven Werte der Gesamtscores von 1-2 und 4-5, zeigten die suboptimalen Resultate für den Gesamtscore von 3, dass die rein subjektive Vergabe dieses Scores nur limitiert nützlich erscheint. Daher sollte der Gesamtscore, der letztlich die Indikation zur weiteren diagnostischen Abklärung liefert, auf Basis eines standardisierten Schemas wie dem PI-RADS definiert und vergeben werden.

Die Analyse der Nebenziele ergab, dass 24 (62%) der 39 Patienten ohne Vorbiopsie Karzinom-positiv waren. Zwischen der Vorbiopsien-Anzahl und einem Karzinom-positiven Ergebnis bzw. dem Gleason Score bestand jedoch keine signifikante Korrelation. Weder der PSA-Wert bzw. der fPSA/PSA-Quotient noch deren Subgruppen korrelierten signifikant mit einem Karzinom-positiven Biopsie-Ergebnis bzw. der Gleason Score-Verteilung. Der PSA-Wert korrelierte nicht mit dem Gesamtscore. Eine suspekte DRU oder TRUS zeigte keine signifikante Korrelation zu einem Karzinom-positiven Biopsie-Ergebnis. Die vorliegende Arbeit zeigte somit, dass diese klinischen Parameter keine verlässlichen Prädiktoren für die Dignität einer Läsion in der untersuchten Kohorte darstellten.

Obwohl die TRUS-gesteuerte Sättigungsbiopsie derzeit die häufigste verwendete Methode in der Abklärung laborchemisch, klinisch oder sonographisch suspekter Befunde der Prostata darstellt, zeigen die vorliegenden Ergebnisse, dass die Anwendung des PI-RADS und die Gewebegewinnung anhand der transperinealen MRT/TRUS-gesteuerten Biopsie valide bildbasierte Verfahren von hoher diagnostischer Güte repräsentieren. Trotz des hohen negativen prädiktiven Wertes von 89,1%, sollten die Spezifität und der positive prädiktive Wert des PI-RADS aufgrund der relativ hohen Inzidenz des Prostatakarzinoms weiter verbessert werden. Ob dies durch Modifikationen der einzelnen Kriterien zur Score-Vergabe oder durch methodische bzw. technische Veränderungen erreicht werden kann, muss in größeren multizentrischen Studien untersucht werden.