

Inter-Reader Variability Using PI-RADS v2 Versus PI-RADS v2.1: Most New Disagreement Stems from Scores 1 and 2

Vergleich der Interrater-Reliabilität von PI-RADS v2 und PI-RADS v2.1: Neue Differenzen entspringen meistens den Bewertungen 1 und 2

Authors

Nick Lasse Beetz¹, Matthias Haas¹, Alexander Baur¹, Frank Konietzschke², Akash Roy³, Charlie Alexander Hamm¹, Madhuri Monique Rudolph¹, Seyd Shnayien¹, Bernd Hamm¹, Hannes Cash⁴, Patrick Asbach¹, Tobias Penzkofer¹

Affiliations

- 1 Department of Radiology, Charite University Hospital Berlin, Germany
- 2 Department of Biometry and Clinical Epidemiology, Charite University Hospital Berlin, Germany
- 3 Biostatistics and Bioinformatics, Duke University School of Medicine, DURHAM, United States
- 4 Department of Urology, Charite University Hospital Berlin, Germany

Key words

PI-RADS v2.1, mpMRI, prostate cancer, inter-reader agreement

received 24.11.2021

accepted 29.12.2021

published online 11.05.2022

Bibliography

Fortschr Röntgenstr 2022; 194: 852–861

DOI 10.1055/a-1752-1038

ISSN 1438-9029

© 2022, Thieme. All rights reserved.

Georg Thieme Verlag KG, Rüdigerstraße 14, 70469 Stuttgart, Germany

Correspondence

Dr. Nick Lasse Beetz

Radiology, Charite University Hospital Berlin, Augustenburger Platz 1, 13353 Berlin, Germany

Tel.: +49/30/4 50 62 78 39

nick-lasse.beetz@charite.de

Supplementary material is available under <https://doi.org/10.1055/a-1752-1038>

ABSTRACT

Purpose To analyze possible differences in the inter-reader variability between PI-RADS version 2 (v2) and version 2.1 (v2.1) for the classification of prostate lesions using multiparametric MRI (mpMRI) of the prostate.

Methods In this retrospective and randomized study, 239 annotated and histopathologically correlated prostate lesions

(104 positive and 135 negative for prostate cancer) were rated twice by three experienced urologists using PI-RADS v2 and v2.1 with an interval of at least two months between readings. Results were tabulated across readers and reading timepoints and inter-reader variability was determined using Fleiss' kappa (κ). Thereafter, an additional analysis of the data was performed in which PI-RADS scores 1 and 2 were combined, as they have the same clinical consequences. **Results** PI-RADS v2.1 showed better inter-reader agreement in the peripheral zone (PZ), but poorer inter-reader agreement in the transition zone (TZ) (PZ: $\kappa = 0.63$ vs. $\kappa = 0.58$; TZ: $\kappa = 0.47$ vs. $\kappa = 0.57$). When PI-RADS scores 1 and 2 were combined, the use of PI-RADS v2.1 resulted in almost perfect inter-reader agreement in the PZ and substantial agreement in the TZ (PZ: $\kappa = 0.81$; TZ: $\kappa = 0.80$).

Conclusion PI-RADS v2.1 improves inter-reader agreement in the PZ. New differences in inter-reader agreement were mainly the result of the assignment of PI-RADS v2.1 scores 1 and 2 to lesions in the TZ. Combining scores 1 and 2 improved inter-reader agreement both in the TZ and in the PZ, indicating that refined definitions may be warranted for these PI-RADS scores.

Key Points:

- PI-RADSv2.1 improves inter-reader agreement in the PZ but not in the TZ.
- New differences derived from PI-RADSv2.1 scores 1 and 2 in the TZ.
- Combined PI-RADSv2.1 scores of 1 and 2 yielded better inter-reader agreement.
- PI-RADSv2.1 appears to provide more precise description of lesions in the PZ.
- Improved inter-reader agreement in the PZ stresses the importance of appropriate lexicon description.

Citation Format

- Beetz N, Haas M, Baur A et al. Inter-Reader Variability Using PI-RADS v2 Versus PI-RADS v2.1: Most New Disagreement Stems from Scores 1 and 2. Fortschr Röntgenstr 2022; 194: 852–861

ZUSAMMENFASSUNG

Ziel Analyse der möglichen Differenzen in der Interrater-Reliabilität zwischen PI-RADS Version 2 (v2) und Version 2.1

(v2.1) für die Klassifizierung von Prostata-Läsionen in der multiparametrischen MRT (mpMRT) der Prostata.

Methoden In dieser retrospektiven und randomisierten Studie wurden 239 bereits annotierte und histopathologisch korrelierte Läsionen der Prostata (104 positiv und 135 negativ für Prostatakrebs) zweimal von radiologisch-fachärztlichen Experten für urologische Bildgebung bewertet: zuerst nach PI-RADS v2 und nach einem zeitlichen Intervall von mindestens 2 Monaten anhand der aktualisierten Version PI-RADS v2.1. Die Ergebnisse der Bewerter und Bewertungszeitpunkte wurden kreuztabelliert und die Interrater-Reliabilität wurde mittels Fleiss' kappa (κ) berechnet. In einer weiteren Analyse wurden die beiden Wertungen mit derselben klinischen Konsequenz PI-RADS-Score 1 und 2 kombiniert und deren Interrater-Reliabilität erneut getestet.

Ergebnisse PI-RADS v2.1 zeigte eine bessere Interrater-Reliabilität in der peripheren Zone (PZ), aber eine schlechtere Interrater-Reliabilität in der Transitionszone (TZ) (PZ: $\kappa = 0,63$ vs. $\kappa = 0,58$; TZ: $\kappa = 0,47$ vs. $\kappa = 0,57$). Die Kombination der PI-RADS-Bewertungen 1 und 2 bei PI-RADS v2.1 resultierte in einer nahezu perfekten Interrater-Reliabilität in der PZ und in einer substantiellen Interrater-Reliabilität in der TZ (PZ: $\kappa = 0,81$; TZ: $\kappa = 0,80$).

Schlussfolgerung PI-RADS v2.1 verbessert die Interrater-Reliabilität in der PZ. Neue Differenzen in der Interrater-Reliabilität bei PI-RADS v2.1 entsprangen hauptsächlich den Läsionen in der TZ, die mit einem Score von 1 oder 2 bewertet wurden. Durch die Kombination der klinisch gleichwertigen PI-RADS-Bewertungen 1 und 2 wird die Interrater-Reliabilität sowohl in der TZ als auch PZ verbessert, sodass eine verbesserte Definition für diese Bewertungen angebracht sein könnte.

Kernaussagen:

- PI-RADS v2.1 verbessert Interrater-Reliabilität in der PZ, aber nicht in der TZ.
- Neue Unterschiede entspringen den PI-RADS v2.1-Bewertungen 1 und 2 in der TZ.
- Die Kombination der PI-RADS v2.1-Bewertungen 1 und 2 verbessert die Interrater-Reliabilität.
- PI-RADS v2.1 ermöglicht offenbar eine präzisere Deskription von Läsionen in der PZ.
- Verbesserung der Interrater-Reliabilität in der PZ betont die Bedeutung von geeigneten Deskriptoren.

Introduction

Multiparametric magnetic resonance imaging (mpMRI) of the prostate for suspected prostate cancer aims at detecting clinically significant prostate cancer (csPCa) and has gained support from international guidelines in order to reduce overdiagnosis and overtreatment [1–3]. The PROMIS and PRECISION trials have shown that mpMRI not only reduces unnecessary biopsies and overdiagnosis of clinically non-significant prostate cancer (cnsPCa) but also improves the detection of csPCa [4, 5]. Additionally, mpMRI is superior in detecting anteriorly located prostate cancers, which are frequently missed by systematic biopsies [6, 7]. Consequently, MRI/ultrasound-fusion biopsies have shown better prediction of the final Gleason score, which is crucial for initial treatment recommendation [8, 9].

In order to standardize prostate MRI examination and reporting, the European Society of Urogenital Radiology (ESUR) introduced Prostate Imaging Reporting and Data System (PI-RADS) version 1 (v1), the first structured reporting scheme for evaluating prostate MRI in pre-therapy patients, in 2012 [10]. Limitations of PI-RADS v1 included poor inter-reader agreement, as well as an unsharp definition of positive and negative examinations regarding suspected prostate cancer [11, 12]. In collaboration with the AdMeTech Foundation and American College of Radiology (ACR), the ESUR updated the reporting scheme to PI-RADS version 2 (v2) in 2015, successfully improving diagnostic accuracy and acceptance among urologists and radiologists [13–15].

An update, PI-RADS version 2.1 (v2.1), was published in 2019, aiming to further improve inter-reader and intra-reader agreement [16]. It introduces three major changes to the scoring sys-

tem: first, typical benign prostate hyperplasia (BPH) nodules (completely encapsulated and round) in the transition zone (TZ) are now classified as PI-RADS score 1. Second, atypical BPH nodules in the TZ are assigned PI-RADS score 2 with possible upgrading to PI-RADS score 3 based on the results of diffusion-weighted imaging (DWI) and corresponding apparent diffusion coefficient (ADC) maps. Only atypical BPH nodules (without or with incomplete encapsulation) or a homogeneous mildly hypointense area between nodules are assigned a score of 2. Third, in the peripheral zone (PZ), a score of 2 is now defined as a wedge-shaped or linear DWI-hyperintense or ADC-hypointense lesion, whereas a score of 3 is defined as a focal DWI hyperintense and/or focal ADC-hypointense lesion.

Although the PI-RADS score does not lead to definite clinical recommendations and other factors including prostate-specific antigen (PSA) level and clinical history have to be considered, PI-RADS scores 4 and 5 indicate a high likelihood of csPCa and favor prostate biopsy. In contrast, PI-RADS scores 1 and 2 suggest a low likelihood of csPCa and prostate biopsy is usually not performed [17, 18].

Given the significance of the PI-RADS score for clinical decision making, further improvement of inter-reader agreement of the updated PI-RADS v2.1 is desired. Therefore, this randomized and controlled study with a retrospective dataset aims to compare inter-reader agreement between PI-RADS v2 and PI-RADS v2.1. In addition, we tested if combining PI-RADS score 1 and score 2 may have an impact on inter-reader agreement.

Materials and Methods

Study design

In this single-center study, three highly experienced urologists interpreted annotated prostate lesions on mpMRI from a retrospective dataset in a randomized fashion, first according to PI-RADS v2 and second according to PI-RADS v2.1. The institutional review board approved this study.

Reference standard

All patients underwent 10-core systematic and 2–8 targeted MRI/TRUS fusion-guided biopsies with histopathological analysis, which served as a reference standard. Prostate cancer was graded using the Gleason classification system [19]. Gleason scores of $\geq 3 + 4$ were considered to indicate csPCa [20, 21]. Reported benign findings included chronic and acute prostatitis, BPH, and normal prostate tissue. The biopsies were obtained using one of two biopsy devices (Hitachi Medical Systems, HI VISION Preirus, or Toshiba, Aplio 500) and 18-gauge needles. All biopsies were executed by a team of experienced urologists and urologists at our tertiary university center.

Lesion annotation and reviewing software

For each prostate mpMRI dataset, 2–4 lesions representative for the relevant histopathological findings were defined and annotated in consensus by two radiologists (N.L.B. and T.P.) based on a review of histopathological and imaging data. These annotations included normal appearing prostate tissue, benign changes, and cancerous lesions. Unclear or ambiguous areas (e. g., if the targeted biopsy in a specific area was non-cancerous but an adjacent systematic biopsy yielded prostate cancer) were not used as possible review targets.

A dedicated proprietary reviewing software was used to ensure a randomized and controlled rating process as well as blinding of readers to clinical data and patient information. Annotated lesions were automatically presented in random order. The readers had access to the complete examination and were allowed to scroll through the dataset without limitation.

PI-RADS v2 and v2.1 reading

In the first round the three readers were instructed to independently rate the annotated lesion according to PI-RADS version 2. After an interval of at least 8 weeks, the readers were instructed to independently re-evaluate the same lesions according to PI-RADS version 2.1. During the review process, the annotated lesions were automatically displayed in all MRI sequences. The reading results were recorded using a digital questionnaire within the reviewing software.

Each mpMRI dataset was read by three senior board-certified urologists from a high-volume university center. All three readers (A.J.D.B., M.H., and P.A.) are attending-level urologists with several years (>7 years) of experience in prostate MRI and hold the highest available certificate for prostate mpMRI from the German Society of Radiology (DRG Q2 certificate). Each

of them has read and supervised at least 1000 prostate mpMRI datasets within the last 5 years.

In the first round, the three readers were instructed to independently rate the annotated lesions according to PI-RADS v2. After an interval of at least two months, the readers were instructed to re-evaluate the same annotated lesions according to PI-RADS v2.1. In both rounds, the annotated lesions were presented randomly to control for potential bias. The readers were handed reporting charts describing the rating system according to PI-RADS v2 and PI-RADS v2.1 for reference (► Fig. 1 and Fig. S1).

Inter-reader agreement was assessed for the primary scoring results, followed by a secondary analysis to assess how inter-reader agreement is affected when PI-RADS score 1 and score 2 are combined.

Patient cohort, patient characteristics, and biopsy results

The study included patients with suspected prostate cancer due to an abnormal digital rectal examination or an elevated PSA level. The availability of a prostate mpMRI and a subsequent in-house targeted MRI/TRUS fusion biopsy combined with a 10-core systematic biopsy were further inclusion criteria.

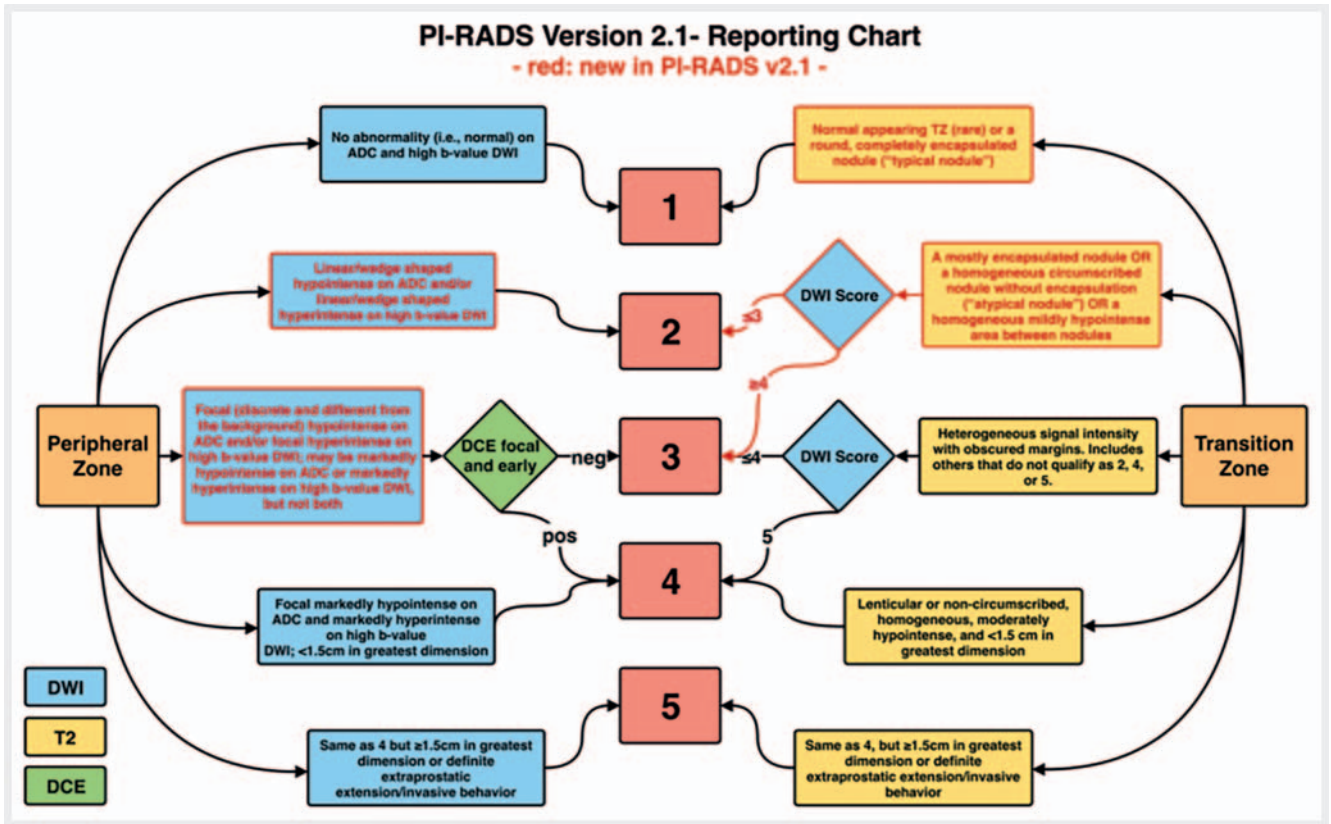
Exclusion criteria were prior prostate-related therapies such as brachytherapy, radiation, or hormonal treatment. Additionally, six patients were excluded because of non-diagnostic image quality due to motion artifacts and hip replacement. A total of 102 patients with a mean age of 68.7 ± 8.1 years and a mean PSA level of 10.1 ± 9.3 ng/dl were included. Overall, 239 previously annotated lesions were rated in the study.

The cohort is a subset of patients from a consecutive MRI/TRUS biopsy cohort recruited at a tertiary university hospital who met the inclusion criteria. The subset included 107 patients, of whom 8 were on active surveillance and 33 had undergone biopsy before.

In the TZ, 68 (65%) of the annotated lesions were normal prostate tissue, 2 (2%) inflammation, 12 (12%) corresponded to a Gleason score 3 + 3, 7 (7%) to a Gleason score 3 + 4, 4 (4%) to a Gleason score 4 + 3, 8 (8%) to a Gleason score 4 + 4 and 1 each (1%) to a Gleason score 5 + 3, 5 + 4 and 5 + 5. In the PZ, 59 (44%) of the annotated lesions were normal prostate tissue, 6 (4%) inflammation, 25 (19%) corresponded to a Gleason score 3 + 3, 12 (9%) to a Gleason score 3 + 4, 2 (1%) to a Gleason score 3 + 5, 7 (5%) to a Gleason score 4 + 3, 14 (10%) to a Gleason score 4 + 4, 6 (4%) to a Gleason score 4 + 5, and 4 (3%) to a Gleason score 5 + 4. The histopathological biopsy results of all 239 lesions are compiled in ► Table 1. The classifications according to the International Society of Urological Pathology (ISUP) are compiled in ► Table 2.

Image acquisition

A standardized mpMRI protocol in accordance with the ESUR guidelines was used for prostate imaging in all patients [22]. The protocol included axial T1-weighted imaging ($3.0 \times 0.6 \times 0.6$ mm, 32 cm FoV), axial and coronal T2-weighted imaging ($3.0 \times 0.47 \times 0.47$ mm, 18 cm FoV), DWI with generation of ADC maps ($3.0 \times 1.4 \times 1.4$ mm, 17 cm FoV, with b-values of 0, 50, 500, 1000, and calculated $b = 1400$ s/mm²), and dynamic contrast-en-



► **Fig. 1** PI-RADS reporting chart: Readers were handed reporting charts describing the rating system according to PI-RADS v2 and v2.1. Shown here is the reporting chart according to PI-RADS v2.1 for the second reading round with major changes being highlighted in red. The reporting chart describing the rating system according to PI-RADS v2 for the first reading round can be found in the supplementary material (online).

► **Table 1** Histopathological results of 239 lesions used to assess PI-RADS in terms of inter-reader agreement in a) the transition zone and b) the peripheral zone. GS = Gleason score.

a) Transition zone			b) Peripheral zone		
Biopsy result	Number	Percentage (%)	Biopsy result	Number	Percentage (%)
Total	104	100	Total	135	100
Normal tissue/BPH	68	65	Normal tissue/BPH	59	44
Inflammation	2	2	Inflammation	6	4
GS 3 + 3	12	12	GS 3 + 3	25	19
GS 3 + 4	7	7	GS 3 + 4	12	9
GS 4 + 3	4	4	GS 3 + 5	2	1
GS 4 + 4	8	8	GS 4 + 3	7	5
GS 5 + 3	1	1	GS 4 + 4	14	10
GS 5 + 4	1	1	GS 4 + 5	6	4
GS 5 + 5	1	1	GS 5 + 4	4	3

hanced (DCE) imaging (3.0 × 1.4 × 1.4 mm, 18.6 cm FoV, 5 s temporal resolution) after intravenous injection of Gd-DO3A-butrol (Gadovist, Bayer Healthcare) at a rate of 3 ml/s. All mpMRI examinations were performed on one of two identical 3-Tesla MRI scanners (Skyra, Siemens Healthineers).

Statistical analysis and performance correlation

Agreement analysis: Fleiss' kappa (κ) and Krippendorff's alpha (α) were used to assess inter-reader agreement. Both measures were analyzed together to determine inter-reader agreement between

► **Table 2** Classification of the biopsies listing PCa lesions by ISUP prostate cancer grade group and location. ISUP = International Society of Urological Pathology. PCa = prostate cancer.

a) Transition zone			b) Peripheral zone		
Grade group	Number	Percentage (%)	Grade group	Number	Percentage (%)
Total	34	100	Total	70	100
ISUP 1	12	35	ISUP 1	25	36
ISUP 2	7	21	ISUP 2	12	17
ISUP 3	4	12	ISUP3	7	10
ISUP 4	9	26	ISUP 4	16	23
ISUP 5	2	6	ISUP 5	10	14

► **Table 3** Overview of PI-RADS scores assigned by all three readers using version 2 and 2.1; TZ = transition zone, PZ = peripheral zone.

Score	PI-RADS v2			PI-RADS v2.1		
	TZ	PZ	Overall	TZ	PZ	Overall
1	8	65	73	107	104	211
2	203	100	303	94	69	163
3	22	42	64	33	33	66
4	14	118	132	10	118	128
5	65	80	145	68	81	149

all three readers. Additionally, inter-reader agreement for combined PI-RADS scores 1 and 2 was calculated. Fleiss' kappa results were interpreted as follows: 'κ' values of 0.41–0.60 indicate moderate agreement, 'κ' values of 0.61–0.80 indicate substantial agreement and 'κ' values >0.81 indicate almost perfect agreement. Krippendorff's alpha (α) was designed to compute inter-reader agreement between multiple ratings (readers 1, 2, 3). Values range between $\alpha = 1$ for perfect reliability and $\alpha = 0$ for absence of reliability.

Accuracy analysis: The study has been modeled as a factorial diagnostic trial involving the factors Reader (levels: readers 1, 2, 3) and Modality (PI-RADS versions 2.1 and 2.0). Diagnostic accuracies of the methods were assessed with the areas under the receiver operating characteristic curves (AUC) for each reader and modality combination. It should be noted that we assumed all lesions to be independent. A clustered-data AUC analysis revealed the same conclusions [23].

R statistical software was used for all data analysis (version 3.6.1; www.r-project.org).

Results

PI-RADS classification

In both reading rounds, a majority of 91 % of annotated lesions were assigned PI-RADS scores of 1 or 2 for low likelihood of csPCa and PI-RADS scores 4 or 5 for high likelihood of csPCa based on mpMRI.

Only 9 % of annotated lesions were PI-RADS score 3 for indeterminate likelihood of csPCa based on mpMRI (64 lesions using PI-RADS v2 vs. 66 lesions using PI-RADS v2.1). An overview of PI-RADS scores assigned by all three readers is compiled in ► **Table 3**.

Distribution of PI-RADS v2 and v2.1 classification for different histopathological findings

Histopathological normal or benign findings were rated 8x score 1, 180x score 2, 9x score 3, and 7x score 5 according to PI-RADS v2, whereas when using PI-RADS v2.1 the same lesions were rated 101x score 1, 79x score 2, 11x score 3, and 5x score 4, and 8x score 5. Inflammation was rated 6x score 2 using PI-RADS v2 and 4x score 1 and 2x score 2 using PI-RADS v2.1. cnsPCa was rated 10x score 2, 9x score 3, 4x score 4, and 13x score 5 using PI-RADS v2, whereas with PI-RADS v2.1 the lesions were rated 1x score 1, 12x score 2, 7x score 3, 3x score 4, and 13x score 5. csPCa was rated 7x score 2, 4x score 3, 10x score 4, and 44x score 5 using PI-RADS v2, whereas with PI-RADS v2.1 the lesions were rated 1x score 1, 1x score 2, 15x score 3, 2x score 4, and 47x score 5. The results are shown in ► **Table 4**.

Inter-reader agreement

Inter-reader agreement for lesions in the PZ was substantial using PI-RADS v2.1 and moderate using PI-RADS v2 ($\alpha = 0.63$ and $\kappa = 0.63$ vs. $\alpha = 0.58$ and $\kappa = 0.58$). In the TZ the inter-reader agree-

► **Table 4** Distribution of PI-RADS v2 and v2.1 classification for different histopathological findings for a) the TZ and b) the PZ. One histological finding showed atypical small acinar proliferation in the PZ which is not shown in this table. TZ = transition zone, PZ = peripheral zone, csPCa = clinically significant prostate cancer, cnsPCa = clinically non-significant prostate cancer.

PI-RADS	Benign/normal		Inflammation		cnsPCa		csPCa	
	v2	v2.1	v2	v2.1	v2	v2.1	v2	v2.1
a) TZ								
Score 1	8	101	0	4	0	1	0	1
Score 2	180	79	6	2	10	12	7	1
Score 3	9	11	0	0	9	7	4	15
Score 4	0	5	0	0	4	3	10	2
Score 5	7	8	0	0	13	13	44	47
b) PZ								
Score 1	59	93	3	7	1	4	2	0
Score 2	71	41	10	6	13	16	5	3
Score 3	22	17	1	0	9	6	8	10
Score 4	25	23	1	2	43	41	49	52
Score 5	0	3	0	0	9	8	71	70

ment was moderate for both PI-RADS v2.1 and PI-RADS v2 ($\alpha = 0.47$ and $\kappa = 0.47$ vs. $\alpha = 0.57$ and $\kappa = 0.57$).

Alluvial plots for the visualization of differences in scores assigned by the three readers using PI-RADS v2 versus PI-RADS v2.1 show that, for the reading using PI-RADS v2.1, most inter-reader differences were attributable to lesions assigned PI-RADS scores 1 and 2 (► Fig. 2). We therefore performed an additional analysis combining these two scores, as both scores have identical clinical consequences. Examples of lesions showing improved inter-reader agreement in the PZ and decreased inter-reader agreement in TZ using PI-RADS v2.1 are shown in ► Fig. 3A-D.

Using PI-RADS v2.1, agreement differences of ≥ 2 score points appeared in 17 out of 135 cases (12%) in the PZ and in 14 out of 104 in the TZ (13%). In the TZ there were 5 cases (5%) in which csPCa was rated with a difference of ≥ 2 score points: 4 lesions were rated PI-RADS v2.1 score 5 by two raters and score 3 by one rater, and another lesion was scored PI-RADS v2.1 score 5 by two raters and score 2 by one rater. In contrast, in the PZ no disagreement with a difference of ≥ 2 score points was observed for csPCa. All differences of ≥ 2 score points separated for benign/normal prostate tissue, inflammation, cnsPCa, and csPCa are shown in ► Table 5.

Inter-reader agreement for combined PI-RADS scores 1 and 2

Analysis of inter-reader agreement with combined PI-RADS scores of 1 and 2 for PZ lesions yielded almost perfect inter-reader agreement using PI-RADS v2.1 and PI-RADS v2 ($\alpha = 0.81$ and $\kappa = 0.81$ vs. $\alpha = 0.81$ and $\kappa = 0.81$). For TZ lesions, inter-reader agreement was nearly perfect using PI-RADS v2.1 and substantial agreement was achieved using PI-RADS v2 ($\alpha = 0.80$ and $\kappa = 0.80$ vs. $\alpha = 0.73$ and $\kappa = 0.73$).

Comparison of diagnostic performance of PI-RADS v2 and v2.1

A Gleason score $\geq 3 + 4$ was considered as csPCa. To measure the diagnostic performance of the readers for the detection of csPCa, the AUC were calculated. In the TZ there was a significant improvement in the diagnostic performance of PI-RADS v2.1 vs. PI-RADS v2.0 (AUC: 0.915 [0.869–0.962] vs. AUC: 0.889 [0.839–0.939], difference AUC: 0.026 [0.0001–0.052]). However, for the PZ there was no difference in diagnostic performance (AUC: 0.900 [0.857–0.947] vs. AUC: 0.905 [0.839–0.939]).

Discussion

In this study we compared inter-reader agreement between PI-RADS v2 and v2.1 in two reading rounds with randomized presentation of a set of predefined lesions in mpMRI datasets. When PI-RADS v2.1 was used and all scores were analyzed separately, inter-reader agreement improved from moderate to substantial in the PZ but worsened in the TZ compared with PI-RADS v2. Visualization of the results revealed that most differences in inter-reader agreement occurred in the assignment of the two lower PI-RADS scores of 1 and 2, which indicate low csPCa likelihood and usually have the same clinical consequences. In a further analysis combining the PI-RADS scores 1 and 2, PI-RADS v2.1 showed improved (almost perfect) inter-reader agreement in the TZ and comparable inter-reader agreement in the PZ.

Major changes introduced with the PI-RADS v2.1 update relate to the TZ, and it is therefore plausible to assume that the readers were mostly divided over rating a lesion as “typical” or “atypical” BPH nodules, even though it might be expected that further experience with PI-RADS v2.1 could cause inter-reader agreement to level off in the future. While the modifications regarding the

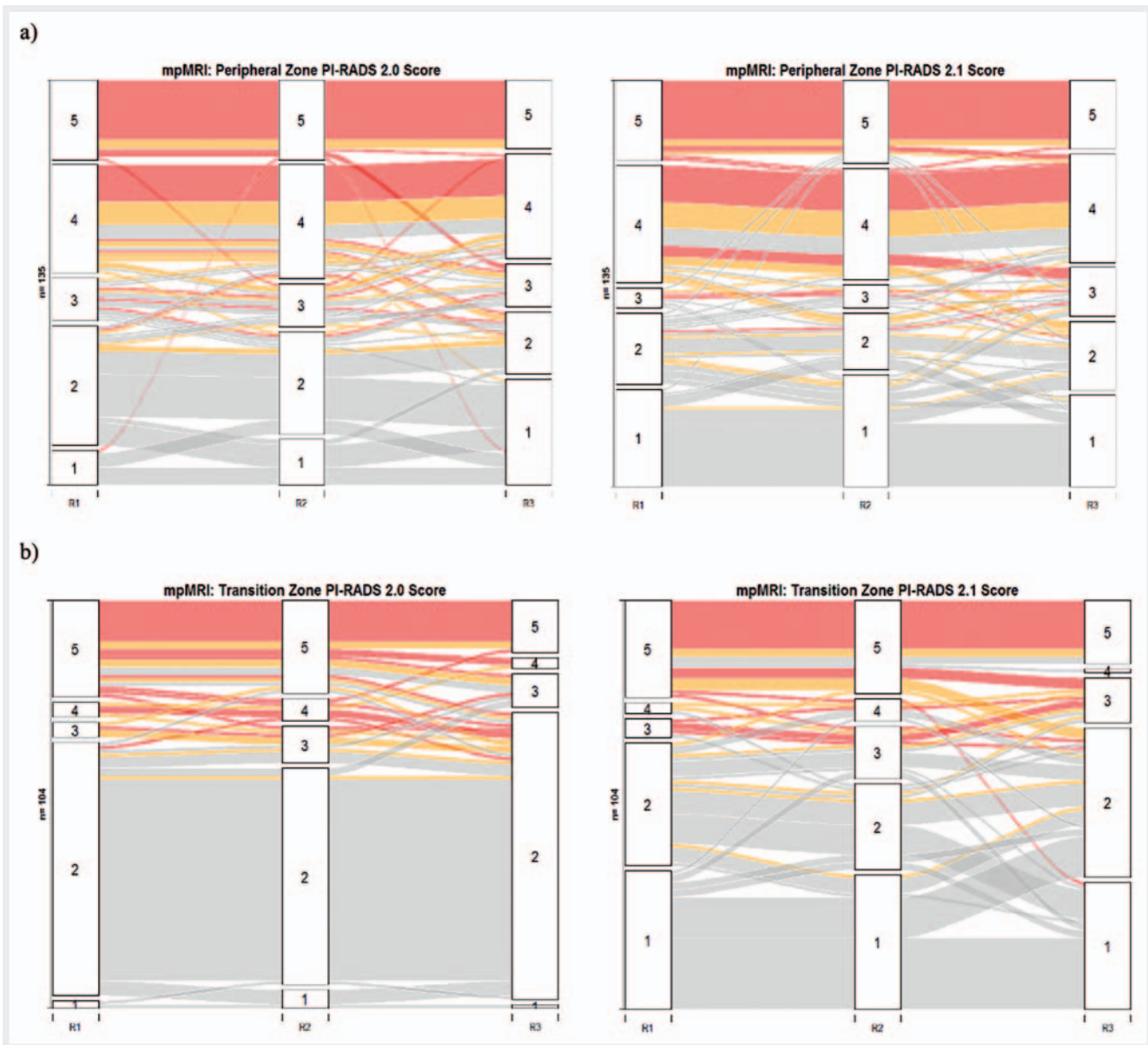


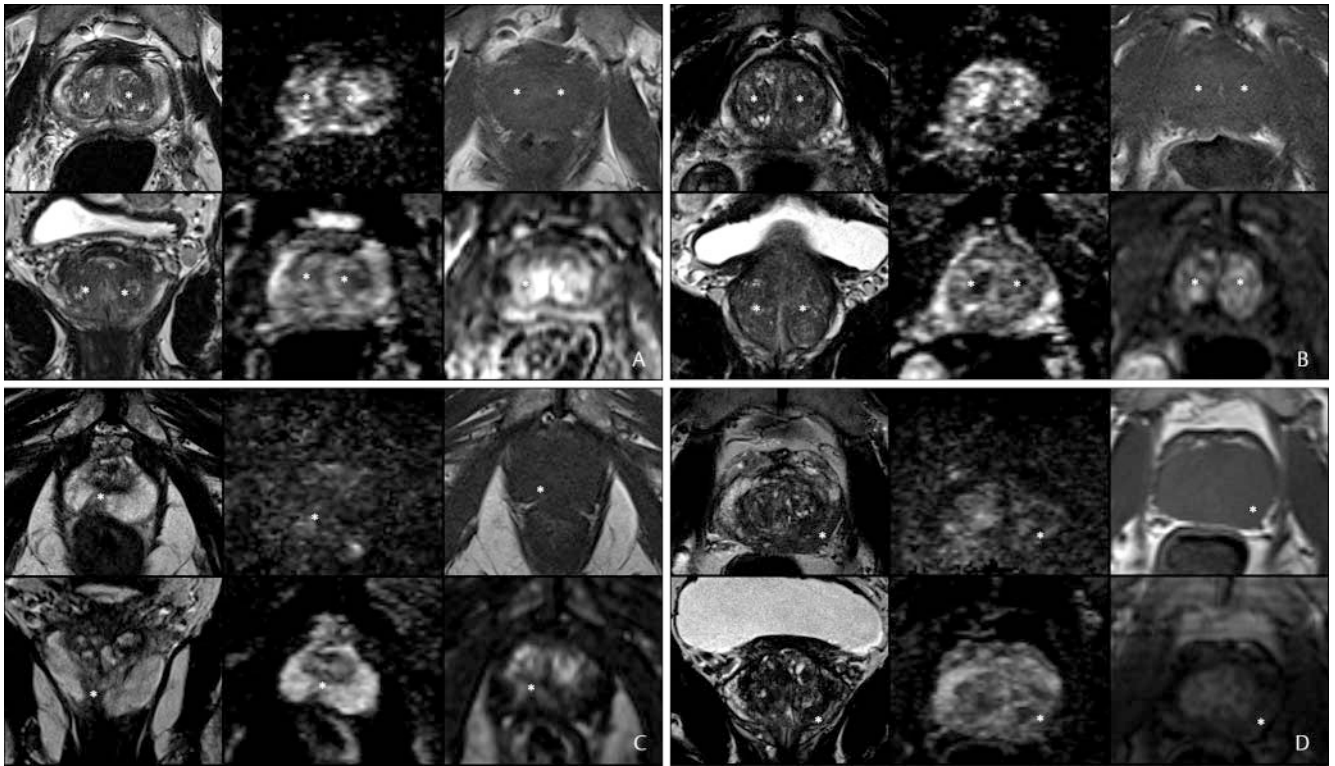
Fig. 2 Alluvial plots visualizing different scores assigned by the three readers for PZ lesions using PI-RADS v2 (left) and v2.1 (right), note the busy changes between scores 1 and 2 in the PI-RADS 2.1 reading. Biopsy-based lesion status is encoded as follows: red = csPca (Gleason $\geq 3 + 4$), orange = non-significant PCa (Gleason 3 + 3) and gray = no PCa. The y-axis represents the PI-RADS score (R1–3 = reader number 1 to 3).

PZ are relatively minor, our results indicate that PI-RADS v2.1 appears to provide a more precise description of lesions in the PZ. This improvement stresses the importance of appropriate lexicon descriptors [24]. Fittingly, in our study there was no inter-reader disagreement with a difference of ≥ 2 score points for lesions harboring csPca in the PZ, whereas in the TZ 5% of lesions harboring csPca were rated with a difference of ≥ 2 score points using PI-RADS v2.1.

The assignment of a PI-RADS score of 1 or 2 indicates a low likelihood of csPca, and, on the basis of these mpMRI findings, biopsy is not performed in most cases. A PI-RADS score of 4 or 5 indicates a high likelihood of csPca, and targeted biopsy should be included in the further management recommendations [25]. In contrast, an indeterminate PI-RADS score of 3 might augment un-

certainly regarding management recommendations for biopsy [26]. Consequently, a low rate of PI-RADS 3 assignments is favorable. In our study, only nine percent of all annotated lesions were assigned PI-RADS score 3 with use of both PI-RADS v2 and v2.1, indicating a high diagnostic certainty with both PI-RADS versions. Even though the number of csPca lesions is low, one has to bear in mind that, in the clinical routine, there is a large proportion of PI-RADS scores 1, 2, and 3. Patients with these scores have a low to indeterminate likelihood of csPca based on the mpMRI and therefore might not undergo prostate biopsy. Nevertheless, these ratings contribute to the overall inter-reader agreement using PI-RADS and need to be evaluated.

Several recently published studies have compared the inter-reader agreement of PI-RADS v2 and v2.1. First, Tamada et al.



► **Fig. 3** Examples of improved inter-reader agreement (Examples A and B) and reduced inter-reader agreement (Examples C and D) using PI-RADS v2.1, 3-Tesla mpMRI images: axial and coronal T2 W, axial DWI/ADC, and axial unenhanced T1 W and DCE. * = lesion, R1–3 = reader 1 to 3. **A** Diffuse T2-hypointense area in the PZ of the left midgland mildly hypointense on ADC map scored as follows: score of 3 by R1 and R3, score of 2 by R2 using PI-RADS v2 and score of 2 by all three readers using PI-RADS v2.1. **B** Wedge-shaped T2-hypointense area in the PZ of the right posterior apex mildly hypointense on ADC map scored as follows: score of 2 by R1 and score of 3 by R2 and R3 according to PI-RADS v2 and score of 2 by all three readers according to PI-RADS v2.1. **C** Images showing two mostly encapsulated nodules in the TZ of the midgland. The atypical nodule on the right side contains a small area with moderate diffusion restriction and was assigned a score of 2 by all three readers according to PI-RADS v2 versus a score of 2 by R1 and R3 and a score of 3 by R2 according to PI-RADS v2.1. The other nodule on the left side shows no areas of restricted diffusion and was assigned a score of 2 by all three readers according to PI-RADS v2 and a score of 1 by R1 and R2 and a score of 2 by R3 according to PI-RADS v2.1. **D** Images showing two mostly encapsulated nodules in the TZ of the midgland. The atypical nodule on the left side contains a small area with mild diffusion restriction and was assigned a score of 2 by all three readers according to PI-RADS v2 and a score of 2 by R1 and R3 and a score of 3 by R2 according to PI-RADS v2.1. The other nodule on the right side shows no areas of restricted diffusion and was assigned a score of 2 by all three readers according to PI-RADS v2 and a score of 1 by R1 and R2 and a score of 2 by R3 according to PI-RADS v2.1.

► **Table 5** Number of cases with a difference of ≥ 2 score points separated for benign/normal prostate tissue, inflammation, cnsPCa, and csPCa in the PZ and TZ. TZ = transition zone, PZ = peripheral zone, csPCa = clinically significant prostate cancer, cnsPCa = clinically non-significant prostate cancer.

Histology	Total	Benign/normal		Inflammation		cnsPCa		csPCa	
		PZ	TZ	PZ	TZ	PZ	TZ	PZ	TZ
Zone	PZ/TZ	PZ	TZ	PZ	TZ	PZ	TZ	PZ	TZ
Number	17/14	8	6	1	0	8	3	0	5

report better inter-reader agreement for TZ lesions scored using PI-RADS v2.1 for the TZ, but the cohort was comparably small [27]. Brembilla et al. describe substantial inter-reader agreement for index lesions [28]. On the other hand, Bhayana et al. report an improved inter-reader agreement using PI-RADS v2.1 in the PZ but not in the TZ, which is consistent with our results [29]. None of these earlier investigators addressed possible underlying causes of the observed differences in inter-reader agreement.

We attribute the lower inter-reader agreement in the TZ to a notable discrepancy in the assignment of BPH nodules to PI-RADS scores 1 and 2, as confirmed by the improved inter-reader agreement in the TZ when PI-RADS v2.1 scores 1 and 2 were combined. Additionally, in accordance with other comparative studies we have demonstrated improved diagnostic performance of PI-RADS v2.1 in the TZ, whereas the diagnostic performance remained unchanged in the PZ [28, 30].

Whereas the first revision of PI-RADS introduced major changes by defining dominant MRI sequences for the PZ and the TZ, the next update v2.1 presented only fairly minor modifications. Indeed, PI-RADS v2 quickly gained recognition from urologists and radiologists, although initial inter-reader studies comparing PI-RADS v1 and v2 were controversial, too. Especially the new zone-specific definition of dominant MRI sequences and the downgrading of the DCE sequences initially raised skepticism [31, 32]. The PI-RADS classification is an ongoing “work-in-progress”, and further refinement of lesion descriptions, as seen in the PZ in PI-RADS v2.1, will help improve future versions of the PI-RADS scoring system.

Limitations

The study is limited by the use of a retrospective dataset. Moreover, owing to the greater level of longer experience with the earlier PI-RADS classification system, readers might have been biased and rated the annotated lesions more correctly according to PI-RADS v2 compared to PI-RADS v2.1. Our study design using predefined annotated lesions reduces bias but differs from clinical practice, which might decrease the overall generalizability of our results.

Conclusion

PI-RADS v2.1 improves inter-reader agreement in the PZ. Differences in inter-reader agreement were mainly the result of the assignment of PI-RADS v2.1 scores 1 and 2 to lesions in the TZ, which improved when combining scores 1 and 2. These findings suggest that more refined definitions of these scores may be needed.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

- [1] Rosenkrantz AB, Verma S, Choyke P et al. Prostate Magnetic Resonance Imaging and Magnetic Resonance Imaging Targeted Biopsy in Patients with a Prior Negative Biopsy: A Consensus Statement by AUA and SAR. *The Journal of urology* 2016; 196: 1613–1618. doi:10.1016/j.juro.2016.06.079
- [2] Team NGU. National Institute for Health and Care Excellence: Clinical Guidelines. In: Prostate cancer: diagnosis and management. London: National Institute for Health and Care Excellence (UK). Copyright (c) NICE 2019.; 2019:
- [3] Mottet N, van den Bergh RCN, Briers E et al. EAU-EANM-ESTRO-ESUR-SIOG Guidelines on Prostate Cancer-2020 Update. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. *Eur Urol* 2020. doi:10.1016/j.eururo.2020.09.042
- [4] Ahmed HU, El-Shater Bosaily A, Brown LC et al. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. *Lancet (London, England)* 2017; 389: 815–822. doi:10.1016/s0140-6736(16)32401-1
- [5] Kasisvisvanathan V, Rannikko AS, Borghi M et al. MRI-Targeted or Standard Biopsy for Prostate-Cancer Diagnosis. *The New England journal of medicine* 2018; 378: 1767–1777. doi:10.1056/NEJMoa1801993
- [6] Volkin D, Turkbey B, Hoang AN et al. Multiparametric magnetic resonance imaging (MRI) and subsequent MRI/ultrasonography fusion-guided biopsy increase the detection of anteriorly located prostate cancers. *BJU international* 2014; 114: E43–E49. doi:10.1111/bju.12670
- [7] Komai Y, Numao N, Yoshida S et al. High diagnostic ability of multiparametric magnetic resonance imaging to detect anterior prostate cancer missed by transrectal 12-core biopsy. *The Journal of urology* 2013; 190: 867–873. doi:10.1016/j.juro.2013.03.078
- [8] Padhani AR, Weinreb J, Rosenkrantz AB et al. Prostate Imaging-Reporting and Data System Steering Committee: PI-RADS v2 Status Update and Future Directions. *Eur Urol* 2018. doi:10.1016/j.eururo.2018.05.035
- [9] Le JD, Stephenson S, Brugger M et al. Magnetic resonance imaging-ultrasound fusion biopsy for prediction of final prostate pathology. *The Journal of urology* 2014; 192: 1367–1373. doi:10.1016/j.juro.2014.04.094
- [10] Barentsz JO, Richenberg J, Clements R et al. ESUR prostate MR guidelines 2012. *European radiology* 2012; 22: 746–757. doi:10.1007/s00330-011-2377-y
- [11] Vache T, Bratan F, Mege-Lechevallier F et al. Characterization of prostate lesions as benign or malignant at multiparametric MR imaging: comparison of three scoring systems in patients treated with radical prostatectomy. *Radiology* 2014; 272: 446–455. doi:10.1148/radiol.14131584
- [12] Vargas HA, Hotker AM, Goldman DA et al. Updated prostate imaging reporting and data system (PI-RADS v2) recommendations for the detection of clinically significant prostate cancer using multiparametric MRI: critical evaluation using whole-mount pathology as standard of reference. *European radiology* 2016; 26: 1606–1612. doi:10.1007/s00330-015-4015-6
- [13] Greer MD, Shih JH, Lay N et al. Validation of the Dominant Sequence Paradigm and Role of Dynamic Contrast-enhanced Imaging in PI-RADS Version 2. *Radiology* 2017; 285: 859–869. doi:10.1148/radiol.2017161316
- [14] Padhani AR, Weinreb J, Rosenkrantz AB et al. Prostate Imaging-Reporting and Data System Steering Committee: PI-RADS v2 Status Update and Future Directions. *Eur Urol* 2019; 75: 385–396. doi:10.1016/j.eururo.2018.05.035
- [15] Weinreb JC, Barentsz JO, Choyke PL et al. PI-RADS Prostate Imaging – Reporting and Data System: 2015, Version 2. *Eur Urol* 2016; 69: 16–40. doi:10.1016/j.eururo.2015.08.052
- [16] Turkbey B, Rosenkrantz AB, Haider MA et al. Prostate Imaging Reporting and Data System Version 2.1: 2019 Update of Prostate Imaging Reporting and Data System Version 2. *Eur Urol* 2019; 76: 340–351. doi:10.1016/j.eururo.2019.02.033
- [17] Baur AD, Maxeiner A, Franiel T et al. Evaluation of the prostate imaging reporting and data system for the detection of prostate cancer by the results of targeted biopsy of the prostate. *Invest Radiol* 2014; 49: 411–420. doi:10.1097/rli.0000000000000030
- [18] Meng X, Rosenkrantz AB, Mendhiratta N et al. Relationship Between Prebiopsy Multiparametric Magnetic Resonance Imaging (MRI), Biopsy Indication, and MRI-ultrasound Fusion-targeted Prostate Biopsy Outcomes. *Eur Urol* 2016; 69: 512–517. doi:10.1016/j.eururo.2015.06.005
- [19] Epstein JI, Egevad L, Amin MB et al. The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma: Definition of Grading Patterns and Proposal for a New Grading System. *Am J Surg Pathol* 2016; 40: 244–252. doi:10.1097/pas.0000000000000530
- [20] Donin NM, Laze J, Zhou M et al. Gleason 6 prostate tumors diagnosed in the PSA era do not demonstrate the capacity for metastatic spread at the time of radical prostatectomy. *Urology* 2013; 82: 148–152. doi:10.1016/j.urology.2013.03.054
- [21] Nickel JC, Speakman M. Should we really consider Gleason 6 prostate cancer? *BJU international* 2012; 109: 645–646. doi:10.1111/j.1464-410X.2011.10854.x

- [22] de Rooij M, Israël B, Tummers M et al. ESUR/ESUI consensus statements on multi-parametric MRI for the detection of clinically significant prostate cancer: quality requirements for image acquisition, interpretation and radiologists' training. *European radiology* 2020; 30: 5404–5416. doi:10.1007/s00330-020-06929-z
- [23] Konietzschke F, Brunner E, Hrsg. Nonparametric analysis of clustered data in diagnostic trials: Estimation problems in small sample sizes. 2009
- [24] Rudolph MM, Baur ADJ, Haas M et al. Validation of the PI-RADS language: predictive values of PI-RADS lexicon descriptors for detection of prostate cancer. *European radiology* 2020; 30: 4262–4271. doi:10.1007/s00330-020-06773-1
- [25] Rosenkrantz AB, Babb JS, Taneja SS et al. Proposed Adjustments to PI-RADS Version 2 Decision Rules: Impact on Prostate Cancer Detection. *Radiology* 2017; 283: 119–129. doi:10.1148/radiol.2016161124
- [26] Hansen NL, Koo BC, Warren AY et al. Sub-differentiating equivocal PI-RADS-3 lesions in multiparametric magnetic resonance imaging of the prostate to improve cancer detection. *Eur J Radiol* 2017; 95: 307–313. doi:10.1016/j.ejrad.2017.08.017
- [27] Tamada T, Kido A, Takeuchi M et al. Comparison of PI-RADS version 2 and PI-RADS version 2.1 for the detection of transition zone prostate cancer. *Eur J Radiol* 2019; 121: 108704. doi:10.1016/j.ejrad.2019.108704
- [28] Brembilla G, Dell'Oglio P, Stabile A et al. Interreader variability in prostate MRI reporting using Prostate Imaging Reporting and Data System version 2.1. *European radiology* 2020; 30: 3383–3392. doi:10.1007/s00330-019-06654-2
- [29] Bhayana R, O'Shea A, Anderson MA et al. PI-RADS versions 2 and 2.1: Interobserver Agreement and Diagnostic Performance in Peripheral and Transition Zone Lesions Among Six Radiologists. *Am J Roentgenol* 2020. doi:10.2214/ajr.20.24199
- [30] Rudolph MM, Baur ADJ, Cash H et al. Diagnostic performance of PI-RADS version 2.1 compared to version 2.0 for detection of peripheral and transition zone prostate cancer. *Sci Rep* 2020; 10: 15982. doi:10.1038/s41598-020-72544-z
- [31] Polanec S, Helbich TH, Bickel H et al. Head-to-head comparison of PI-RADS v2 and PI-RADS v1. *Eur J Radiol* 2016; 85: 1125–1131. doi:10.1016/j.ejrad.2016.03.025
- [32] Schaudinn A, Gawlitzka J, Mucha S et al. Comparison of PI-RADS v1 and v2 for multiparametric MRI detection of prostate cancer with whole-mount histological workup as reference standard. *Eur J Radiol* 2019; 116: 180–185. doi:10.1016/j.ejrad.2019.04.012