Prostate cancer detection with biparametric magnetic resonance imaging (bpMRI) by readers with different experience: performance and comparison with multiparametric (mpMRI)

Marco Gatti[1] · Riccardo Faletti[1] · Giorgio Calleris[2] · Jacopo Giglio[1] · Claudio Berzovini[1] · Francesco Gentile[1] · Giancarlo Marra[2] · Francesca Misischi[1] · Luca Molinaro[3] · Laura Bergamasco[4] · Paolo Gontero[2] · Mauro Papotti[3] · Paolo Fonio[1]

Abstract

Purpose To study the detection of clinically significant prostate cancer (PCa) by readers with different experience, comparing performance with biparametric magnetic resonance imaging (bmMRI) and with the reference multiparametric (mpMRI).

Methods Retrospective analysis of 68 patients with mpMRI of the prostate at 1.5 Tesla using a 32 phased-array coil. Forty-five patients (cases) underwent radical prostatectomy, whereas 23 (controls) had a negative prostate biopsy and ≥ 2.5 years of negative follow-up. Six observers (two with 1000 cases interpreted, two with 300, two with 100) performed the analysis first with bpMRI including diffusion-weighted imaging (DWI), apparent diffusion coefficient (ADC) maps and T2-weighted (T2W) imaging in three planes and, after 1 month, with mpMRI, adding dynamic contrast enhancement (DCE). The performance was quantified by sensitivity (SNS), specificity (SPC) and area under the curve (AUC) of the ROC (Receiver Operating Characteristics) procedure.

Results Concordance within observers of equivalent experience was good (weighted Cohen's $k \approx 0.7$). The two expert readers performed as well in bpMRI as in mpMRI (SNS = 0.91–0.96, AUC = 0.86–0.93; $p \geq 0.10$); readers with 300 cases performed well in mpMRI, but significantly worse in bpMR: SNS = 0.58 versus 0.91 ($p < 0.0001$) and AUC = 0.73 versus 0.86 ($p = 0.01$); the limited experience of readers with 100 cases showed in mpMRI (SNS = 0.71; AUC = 0.77) and even more in bpMRI (SNS = 0.50; AUC = 0.68).

Conclusion The study revealed the impact of the readers' experience when using bpMRI. The bpMRI without contrast media was a valid alternative for expert readers, whereas less experienced ones needed DCE to significantly boost SNS and AUC. Results indicate 700–800 cases as threshold for reliable interpretation with bpMRI.

Keywords Prostate imaging reporting and data system version 2 (PI-RADS v2) · Prostate cancer (PCa) · Biparametric magnetic resonance imaging (bpMRI) · Dynamic contrast-enhanced imaging (DCE) · Diffusion-weighted imaging

✉ Marco Gatti
marcogatti17@gmail.com

1   Radiology Unit, Department of Surgical Sciences, University of Turin, Via Genova 3, 10126 Turin, Italy

2   Urology Unit, Department of Surgical Sciences, University of Turin, Turin, Italy

3   Pathology Unit, Department of Medical Sciences, University of Turin, Turin, Italy

4   Department of Surgical Sciences, University of Turin, Turin, Italy

Introduction

Multiparametric magnetic resonance imaging (mpMRI) is increasingly being used in prostate cancer (PCa) detection, staging and treatment planning [1–3]. A recent, large, multi-center trial [1] showed that mpMRI as a triage test in men with elevated Prostate Specific Antigen (PSA) could improve the detection of clinically significant cancers while reducing the number of unnecessary biopsies in men with low risk of harboring such cancers.

The European Society of Urogenital Radiology (ESUR) guidelines [4] recommend the acquisition of T1- (T1W) and T2-weighted (T2W) imaging, diffusion-weighted imaging

(DWI) and dynamic contrast material–enhanced (DCE) MR imaging pulse sequences for all prostate MR examinations. As a matter of fact, the added value of DCE sequences to the combination of T2W and DWI, in terms of diagnosis and staging, is still debated. Several retrospective studies [5–7] showed that DCE improved cancer detection or underlined the role of bpMRI in addition to the standard digital rectal examination and prostate specific contrast antigen-based screening [8, 9]; others reported no added value from DCE use when comparing biparametric MRI (bpMRI) with the standard mpMRI [10–16] and the latest revision of PI-RADSv2 attributed a minor role to DCE sequences [4].

Since the indications to prostate MRI in PCa are likely to become more and more widespread in the near future, improvement in patient comfort and minimization of MRI duration and adverse effects are key factors to be considered for further optimizing imaging protocols. Gadolinium administration has some drawbacks, such as longer scanning and preparation time, high costs and potential adverse events. Indeed, if contrast usage were proven to provide no additional benefits, significant advantages would derive from its avoidance; moreover, T2W images alone or with DWI are sufficient for MRI-sonography fusion to direct biopsy needles under transrectal ultrasound guidance.

In most institutions, including ours, the standard protocol is mpMRI. The existence of a learning curve in mpMRI interpretation is well acknowledged [17, 18], even if not yet precisely described. The unknown is how the experience accrued in interpreting images with this complete protocol prepares readers to interpret images without the information given by the DCE sequences, in view of eventual examinations without the use of contrast media.

To address this issue, we performed a retrospective study on the detection of the index lesions in men with suspicion of prostate cancer (PCa) by radiologists with different records of mpMR images interpreted. Specifically, our goal was to find an answer to the following questions:

(i)   What is the difference in diagnostic accuracy between bpMRI and the reference mpMRI? i.e., what is the actual contribution of the DCE sequences?

(ii) Is this difference influenced by the experience acquired with mpMRI? If yes, which level of experience is necessary to reach an acceptable diagnostic accuracy also in bpMRI?

Materials and methods

Study design and patient characteristics

We designed and implemented a retrospective study on the images of patients who underwent mpMRI in our Institution

for PCa suspicion. The same scans were to be interpreted twice, first according to the biparametric protocol, and subsequently according to the multiparametric protocol, to be considered as reference.

The initial population of PCa cases was composed by 57 consecutive patients with diagnosis of PCa who underwent mpMRI and subsequent radical prostatectomy (RP) at our Institution from September 2014 to September 2016. This choice was made to minimize possible memories of images already examined in the routine diagnostic work. Exclusion criteria were (a) biopsy-MRI interval shorter than 6 weeks, to reduce post-biopsy changes that may hinder the interpretation of prostate MRIs [4] and (b) an MRI-surgery-interval longer than 6 months. The number of PCa cases included amounted to 45.

To allow a blinded review of MRI scans, reducing interpretation biases, our institution's database was searched backward to add to the 45 PCa cases a control sample of 23 men with no diagnosis of PCa (ratio of one control every two case patients) and at least a 2.5-year follow-up negative for PCa, including at least another MRI scan and/or another negative prostate biopsy plus a stable PSA. All controls had a negative 12-core transrectal ultrasound (TRUS)-guided prostate biopsy.

Table 1 reports the patient/tumor characteristics of the 68 patients constituting our study population.

According to our Institution's procedure, before the MR examination, all patients were informed about the possible use of their data for study purposes and gave their consent. As required by the Ethics Committee, patients' information was anonymized prior to the analysis. The study is a retrospective trial without any study-related clinical intervention and conforms to the Helsinki Declaration.

MRI protocol

The MR examination was performed at 1.5 T (Achieva, version 2.6, Philips Medical Systems, Eindhoven, The Netherlands) using a 32 phased-array coil. The MR protocol with its detailed sequences and its technical parameters is shown in Table 2. Apparent diffusion coefficient (ADC) maps were automatically constructed using the manufacturer's software (b-values of 50 and 1000 s/mm^2). A minimal preparation enema was administered to the patient in the hours prior to the examination. If tolerated, a 1 mg Hyoscinbutylbromide (Buscopan, BoehringerIngelheim, Ingelheim am Rhein, Germany) intravenous injection was given to the patient to reduce peristaltic motion.

Imaging analysis

Image evaluation was performed using a medical imaging workstation (MR Extended Work Space 2.6.3.2

Table 1 Demographical, clinical and pathological characteristics of the patients

Clinical and pathological characteristics		Cases (n=45)	Controls (n=23)	p
Age (years)		65 (59;71)	62 (58.5;67.5)	0.26
PSA (ng/ml)		7.5 (5.1;9.3)	6.35 (4.6;8.4)	0.37
Prostate volume (ml)		61.3 (52.5;77.3)	47.7 (35.3;67.4)	0.08
PSA density (ng/ml^2)		0.11 (0.07;0.18)	0.125 (0.09;0.14)	0.45
Transurethral resection prostate (n°)		11	4	0.50
MRI-surgery interval (months)		2.7 (2.2;5.2)	n.a.*	–
Negative follow-up (y)		n.a.	4 (3.1;4.3)	–
Index lesion diameter (mm)		18 (13;22)	n.a.	–
Index lesion location	Peripheral zone	43 (95.6%)	n.a.	–
	Transitional zone	2 (4.4%)	n.a.	–
ISUP 2014 grading	Group 2 (Gleason 3+4)	24 (5%)	n.a.	–
	Group 3 (Gleason 4+3)	15 (33%)	n.a.	–
	Group 4 (Gleason 8)	5 (11%)	n.a.	–
	Group 5 (Gleason ≥ 9)	1 (2%)	n.a.	–
Staging *(%)*	pT2a	7 (16%)	n.a.	–
	pT2b	3 (7%)	n.a.	–
	pT2c	19 (42%)	n.a.	–
	pT3a	14 (31%)	n.a.	–
	pT3b	1 (2%)	n.a.	–
	pT4	1 (2%)	n.a.	–

Continuous variables expressed as median (25th percentile; 75th percentile)

n.a. not applicable

Table 2 MRI acquisition protocol

Parameters	T2W axial	T2W sagittal	T2W coronal	DWI	DCE
Section thickness/gap (mm)	3/0	3/0	3/0	3/0	3/0
Phase-encoding direction	Right to left	Foot to head	Right to left	Anteroposterior	Anteroposterior
Repetition time (ms)	4157	5214	4615	4061	4.7
Echo time (ms)	100	100	100	74	2.3
Field of view	180×180×66	180×86×180	180×180×76	180×243×69	231×80×80
Acquisition matrix	256×198	256×204	256×198	80×94	100×129
Acquired voxel size	0.7×0.88	0.7×0.88	0.7×0.88	2.25×2.5	1.8×2.0×1.8
Reconstructed voxel size	0.469	0.469	0.469	1.41	1
b values (s/mm^2)	n.a.	n.a.	n.a.	50, 1000	n.a.
No. ff repetition	4	3	4	6	1
Turbo factor	18				
Acquisition duration (min)	4	3	3	5	3
Flip angle (degrees)	90	90	90	90	10

2009 software, Philips Medical Systems, Eindhoven, the Netherlands).

The MRI scans were interpreted by three groups of readers: Group A (two senior radiologists with a record of about 1000 cases analyzed), Group B (two junior radiologists with about 300 cases) and Group C (two residents, with about 100 cases). The inclusion of two readers in each group was done for avoiding biases due to singularities in the individual performing capabilities.

The readers were blinded to clinical and pathological features: each of them performed two readings of the same image, first with the biparametric protocol and, after at least 1 month, with the multiparametric one. The order of the sessions was chosen to prevent possible biases in the biparametric reading due to possible (and not controllable) remembrances of the mpMRI images.

The biparametric protocol included axial DWI images (the $b = 1000$ s/mm^2 images were used for visual

assessment with the corresponding ADC maps) and axial, sagittal and coronal T2-weighted images. The multiparametric protocol included T2-weighted images, DWI and the corresponding ADC maps and DCE-MRI. In each session, the sequence of cases was established by a random number generator.

In every session, for each case, the following data were recorded: (1) lesion location (sector map), (2) lesion diameters and (3) probability of malignancy for each finding using 1–5 PI-RADSv2 categories [4]. For the biparametric protocol, we used the alternative overall assessment category proposed by De Visschere et al. [14]: similar to PI-RADSv2, but using T2W in the peripheral zone, instead of DCE, to resolve a DWI score 3. As reported in [19] a positive DCE MRI lesion was considered a lesion where the enhancement is focal, earlier or contemporaneous with enhancement of adjacent normal prostatic tissues, and corresponds to a finding on T2W and/or DWI; differently a negative DCE MRI lesion was one that either does not enhance early compared to surrounding prostate or enhances diffusely so that the margins of the enhancing area do not correspond to a finding on T2W and/or DWI. Scores 4 or 5 were considered as positive findings, whereas PI-RADSv2 scores 1, 2 and 3 were considered negative [4, 19].

Pathology

Whole-mount slices were used for histopathology review. Index tumor was defined as the tumor with the highest volume at pathology [20, 21], as calculated with the ellipsoid formula. A dedicated uropathologist, blinded to clinical and radiological data, graded, measured and located all PCa index lesions on a sector-map, identical to the one used by MRI readers. The ISUP 2014 system, which combines updated Gleason patterns into 5 grading groups, was used for lesion grading, as indicated by EAU guidelines [22].

Subsequently, a radiologist with experience on more than 2000 prostate cases and with access to all available data from MRI, pathology and follow-up, crossed the results of the 6 readers against definitive histology, which was considered the gold standard. To be considered MRI-detected, lesions had to be identified in the same sector on the axial MRI and whole-mount slice, and in the appropriate segment. Positive MRI corresponding to histologically confirmed tumors were considered true positives (TP), whereas those without corresponding histopathologic correlates were considered false positives (FP). Negative MRI without histological confirmation were considered true negatives (TN), whereas those corresponding to histologically confirmed tumors were considered missed (false negative, FN).

Statistical analysis

Continuous variables did not pass the Shapiro–Wilk W test for normality and are thus reported as median and Inter Quartile range (25th percentile–75th percentile), and compared with the nonparametric Mann–Whitney test. Concordance between readers in the PI-RADSv2 attribution was estimated by Cohen's coefficient k with linear weighting and its standard error SE.

The primary outcome variable of interest was the correct individuation of PCa index lesions by each reader in the two MRI protocols, measured by Sensitivity (SNS), Specificity (SPC) and by the Area under the curve (AUC) of the Receiver Operating Characteristics (ROC) curve. The predictor was the number of cases examined by readers in the three groups.

For comparison between the two protocols, Sensitivity and Specificity were compared with McNemar's test for correlated proportions, whereas, for comparison between two different reader groups, they were compared with Chi square test or, when more appropriate, with Fisher's exact test. The relevance of the ROC curve lies in the fact that its shape is sensitive not only to TP, TN, FP and FN, but also to the ratio of the number of PI-RADSv2 classified 3 to the number of those classified 1 and 2: a high ratio pushes the curve toward the upper left corner, (SNS = SPC = 1), thus increasing AUC. Since an assignment of PI-RADS 3 means that "the presence of clinically significant cancer is equivocal" [4], the ROC curve adds information not present in SNS and SPC. Differences between two ROC curves were assessed according to Hanley and McNeil [23].

Statistical significance was set at two-tails $p < 0.05$. The analysis was performed with StatPlus:Mac v.6 (AnalysisSoft. Walnut. CA. USA).

Results

The concordance within each group on the PI-RADSv2 attribution to the 68 image sets was satisfying, with Cohen's kappa coefficient with linear weighting ranged from $k = 0.67$ with SE = 0.12 to $k = 0.71$ with SE = 0.07. This allowed pooling the PI-RADSv2 distributions of the two readers to obtain the diagnostic parameters characterizing the whole group over the basin of $68 \times 2 = 136$ patients (90 cases and 46 controls).

The analysis yielded 6 sets of diagnostic parameters (SNS, SPC, AUC) characterizing each group in each protocol. Table 3 shows the values of TP, SNS and AUC obtained by the three groups when working with the biparametric protocol compared with those obtained with the multiparametric protocol. Specificity was always around 0.80 irrespective of the protocol and the readers'

Table 3 True positives, sensitivity and area under the curve (AUC)

	Reading group	bpMRI	mpMRI	*p*
True positives/90	A (1000 cases)	82	86	0.13
	B (300)	52	82	*<0.0001*
	p (A vs B)	*<0.0001*	0.13	
	C (100)	45	64	*0.006*
	p (B vs C)	*0.02*	*0.0001*	
Sensitivity	A (1000 cases)	0.91	0.96	0.25
	B (300)	0.58	0.91	*<0.0001*
	p (A vs B)	*<0.0001*	0.23	
	C (100)	0.50	0.71	*0.04*
	p (B vs C)	*0.02*	*<0.0001*	
AUC	A (1000 cases)	0.86	0.93	0.10
	B (300)	0.73	0.86	*0.01*
	C (100)	0.68	0.77	0.14
	p (A vs B)	*0.03*	0.13	
	p (B vs C)	0.40	0.07	

Significant differences were reported in italics

experience (range 0.79–0.84). Significant differences were instead found for Sensitivity and AUC, both related to strong variations in the number of true positives (TP). For the bpMRI protocol, the decreasing experience reduced significantly TP (and consequently SNS and AUC) from Group A to Group B ($p < 0.0001$ for SNS and $p = 0.03$ for AUC) and to a lesser extent from Group B to Group C ($p = 0.02$ for SNS and $p = 0.40$ for AUC), whereas for the mpMRI protocol, the only significant reductions in performance are from Group B to Group C.

The comparison of each reader's log for the 68 images was used to investigate the differences in the evaluation of the same image when using one or the other protocol. With the mpMRI protocol, each reader of Group A reclassified two false negative patients into two true-positives, increasing slightly ($p = 0.13$) the number of TP from 82/90 to 86/90. As a consequence, the performance of Group A was essentially similar with or without DCE sequences: SNS 0.96 versus 0.91 ($p = 0.25$), SPC 0.84 versus 0.78 ($p = 0.5$) and AUC 0.93 versus 0.86 ($p = 0.10$).

Conversely, for Group B, the mpMRI reading mode witnessed an increase in the TP number by 30 units, from 52/90 in the previous session to 82/90 ($p < 0.0001$): the two readers' log showed that the increase was due to 13 upgrading of the PI-RADSv2 attribution from 3 to 4 and to the entry of 17 lesions which had escaped detection in the biparametric evaluation. Eight lesions remained undetected.

Figure 1 illustrates a typical case of PCa: a Gleason 4 + 4 PCa identified and rated as PIRADS 4 by all readers in both bpMRI and mpMRI. Figure 2 illustrates a case of upgrading: a Gleason 4 + 3 prostate cancer which was rated as PIRADS 3 (negative) before the addition of DCE and classified as 4 (positive) after the addition by of DCE by both Group B readers. Figure 3 illustrates a case of another Gleason 4 + 3 prostate cancer in which the lesion was instead missed by
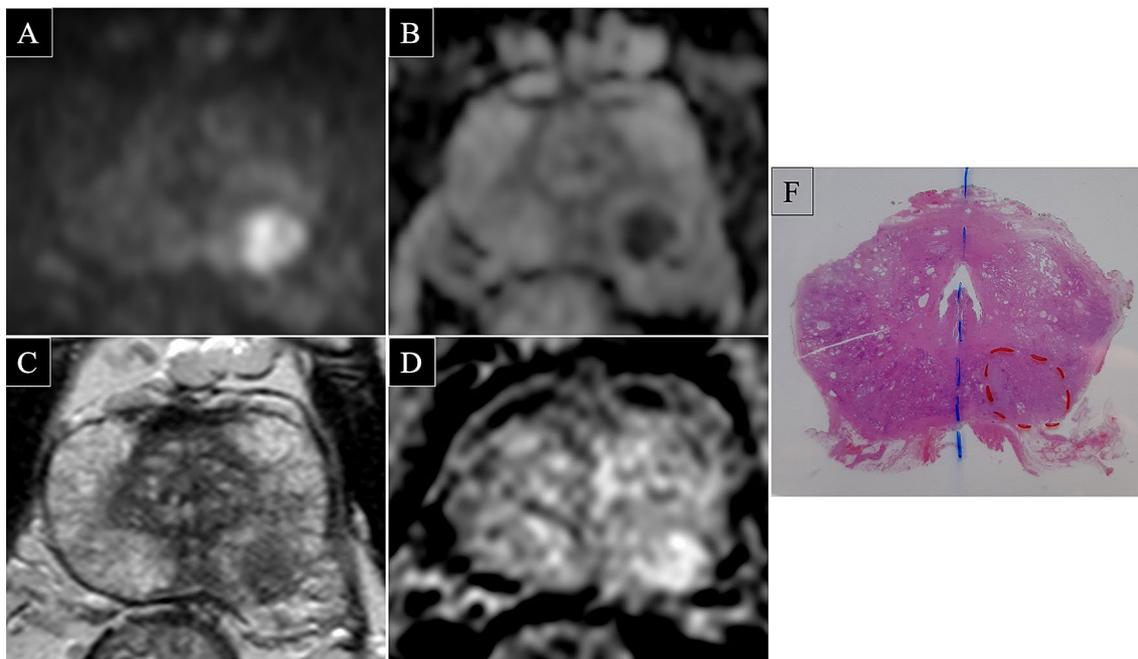

Fig. 1 High-*b*-value (1000 s/mm^2) DW image (panel **a**), ADC map (panel **b**), Axial T2-weighted MR image (panel **c**), DCE image (panel **d**) and the corresponding histopathology map (**f**) of a Gleason 4 + 4 prostate cancer (inked in red) in left peripheral zone. This lesion was rated as PIRADS 4 (positive) both in bpMRI and in mpMRI by all readers
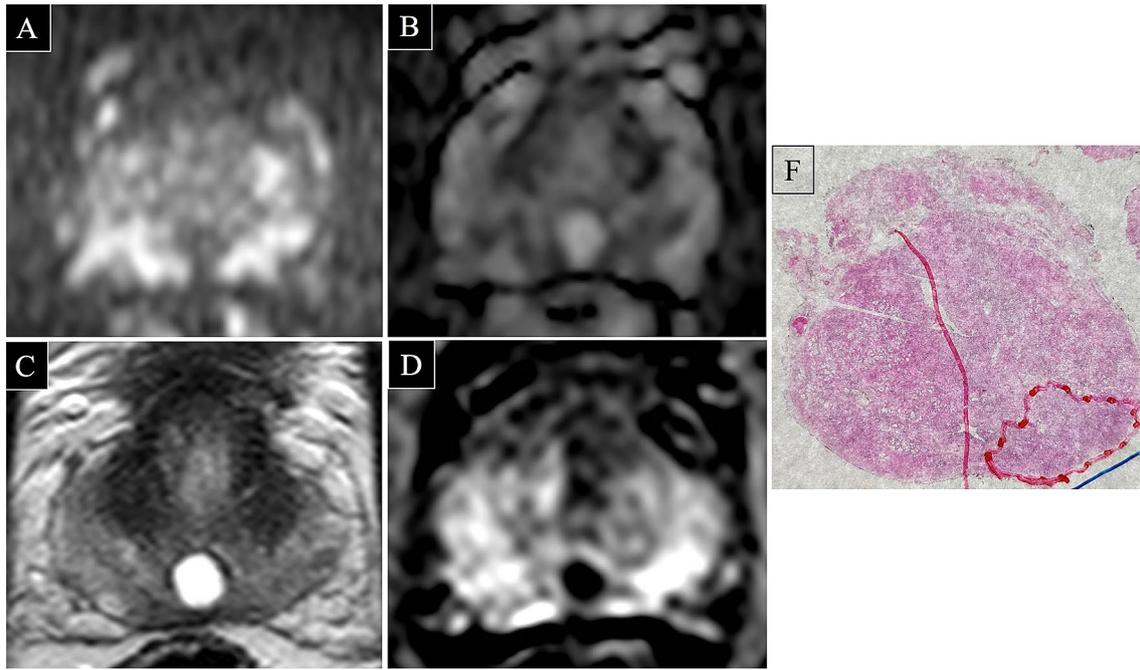
Fig. 2 High-*b*-value (1000 s/mm²) DW image (panel **a**), ADC map (panel **b**), Axial T2-weighted MR image (panel **c**), DCE image (panel **d**) and the corresponding histopathology map (**f**) of a Gleason 4+3 prostate cancer (inked in red) in left peripheral zone. This lesion was rated as PIRADS 3 (negative) before the addition of DCE and classified as 4 (positive) after the addition by both group B readers
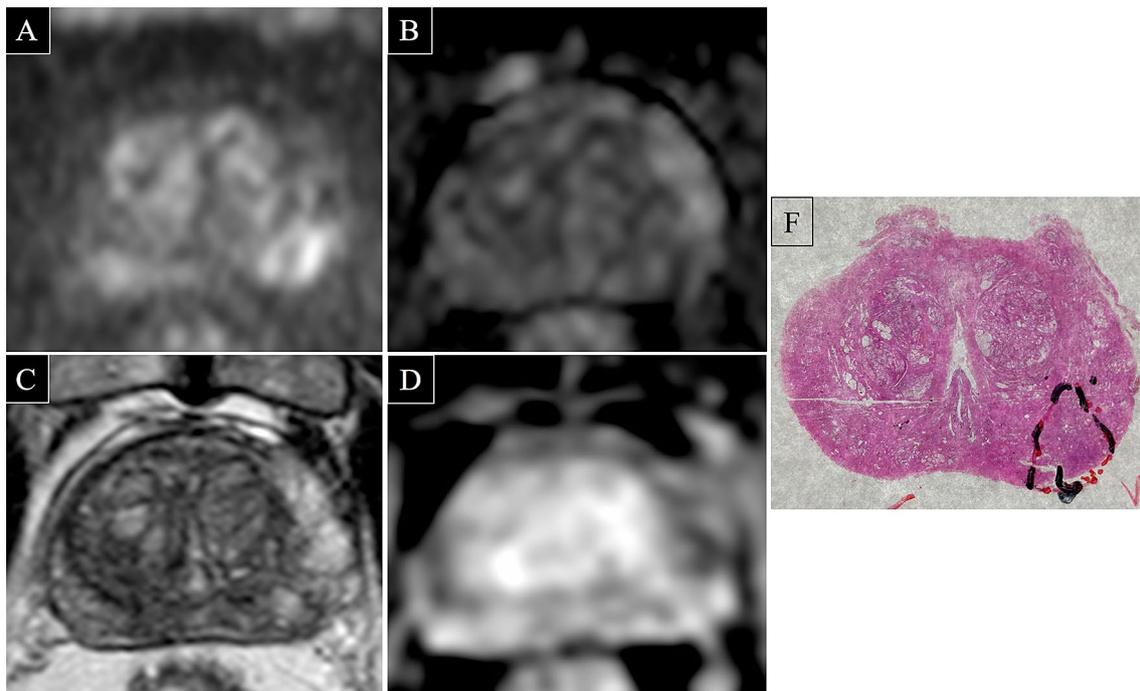

Fig. 3 High-*b*-value (1000 s/mm²) DW image (panel **a**), ADC map (panel **b**), Axial T2-weighted MR image (panel **c**), DCE image (panel **d**) and the corresponding histopathology map (**f**) of a Gleason 4+3 prostate cancer (inked in black and red) in left peripheral zone. This lesion was missed by both group B readers in bpMRI and in mpMRI

the two Group B readers both in bpMRI and in mpMRI. The significant increase in TP caused by the addition of DCE led to the increase in SNS from 0.58 to 0.91 ($p < 0.0001$), and to the change in shape of the ROC curve quantified by the increase of AUC from 0.73 to 0.86 ($p = 0.01$).

Finally, for Group C, the mpMRI reading mode witnessed an increase in the TP number from 45/90 in the previous session to 64/90 ($p = 0.006$): the two readers' log showed that the increase was due to 5 upgradings of the PI-RADSv2 attribution from 3 to 4 and to the entry of 14 lesions which had escaped detection in the biparametric evaluation. The sensitivity increased from 0.50 to 0.71 ($p = 0.04$), and so did AUC, even if not significantly ($p = 0.14$). Twenty-six lesions remained undetected.

All readers detected the two lesions in the transitional zone, with volumes of 1.1 and 1.2 cm^3.

The greatest differences in diagnostic parameters between the two protocols were seen for Group B. The performances of its two readers were thus used to understand which features of the prostate images impaired the detection of lesions when the DCE sequences were not considered. Nineteen lesions were detected by both readers (overall 38 TP) and 14 lesions by only one of them (14 TP), with 15 lesions missed by both (overall FN = 30) and 8 by either one of them (FN = 8). Table 4 compares the features of the 15 patients with index lesions missed by both readers with those of the 19 patients with lesions detected by both readers. The only variable significantly associated with the missing of a lesion in absence of DCE is a small prostate volume ($p = 0.02$).

Figure 4 illustrates the different dependence of Sensitivity (top panel) and AUC (bottom panel) on the level of experience of the readers, expressed as number of mpMRI cases. For the biparametric protocol, both SNS and AUC show a monotone almost-linear increase over the whole range of 900 cases: the slope is about 4×10^{-4} for SNS and 2×10^{-4} for AUC, i.e., every 100 additional cases increase SNS by 0.04 and AUC by 0.02. For the multiparametric protocol, SNS and AUC have an initial steep increase over the first 200 cases, with slope 10^{-3} for SNS and 4×10^{-4} for AUC, after which both parameters stabilize around 0.90. The two

plots give also a rough indication on the number of cases necessary to reach values ≥ 0.80 for Sensitivity and AUC: 150–200 for the multiparametric protocol, and 700–800, about 3–4 times as many, for the biparametric protocol at 1.5 T without the use of ERC.

Discussion

The motivation of our study was to investigate the extent to which the experience gained in the standard multiparametric mode could advantage the transition to a biparametric mode without penalizing diagnostic accuracy. This is an essential point to warrant the reliability of examinations performed without the help of contrast media. Our study evaluated the performance of three groups of readers in the detection of PCa index lesion by MRI at 1.5 T without endorectal coil, using the RP specimens as a reference standard. The three groups differed for the levels of expertise: Group A included two readers with experience on about 1000 mpMR prostate images, Group B two readers with experience on about 300 and Group C two readers with experience on about 100 cases. Sixty-eight images (45 cases and 23 controls) were interpreted first with MRI biparametric protocol, without DCE sequences, and 1 month later with MRI multiparametric protocol.

Our results on the readers of Group A, with a record of about 1000 cases, showed no differences in their performance between biparametric and multiparametric protocol: SNS 0.96 versus 0.91 ($p = 0.25$) and AUC 0.93 versus 0.86 ($p = 0.10$).

Quite different was the outcome for Group B, with a record of about 300 cases. Despite the fact that when working with mpMRI they achieved levels of performance comparable to those of the more expert readers of group A (SNS = 0.91 and AUC = 0.86), when they could not rely on the DCE addition, their performance dropped significantly (SNS = 0.58, AUC = 0.73; $p < 0.0001$). The two readers had to rely on the DCE addition in about 30% of the cases to reach a trustworthy lesion detection. The

Table 4 Biparametric protocol: comparison of features relative to the 15 patients missed by both readers of Group B (300 cases) with those relative to the 19 patients whose lesions were correctly detected by both readers

	Missed by both readers	Detected by both readers	p
N patients	15/45 (33%)	19/45 (42%)	
Age (years)	66 (59–72)	65.5 (59.3–72.8)	0.86
MR-PR interval (months)	2.7 (2.4–5.9)	2.9 (2.4–5.3)	0.75
PSA (ng/ml)	7.0 (3.0–9.4)	7.0 (5.2–8.2)	0.81
Prostate volume (ml)	50.0 (28.2–62.8)	69.1 (60.3–76.7)	*0.02*
Lesion max diameter (mm)	16.5 (8.3–21.8)	16.0 (13.0–22.0)	0.63
ISUP ≥ 3	7 (46.7%)	10 (52.6%)	> 0.99
multifocal	7 (46.7%)	10 (52.6%)	0.33
TURP	5 (33.3%)	3 (15.8%)	0.43

Significant difference is reported in italics
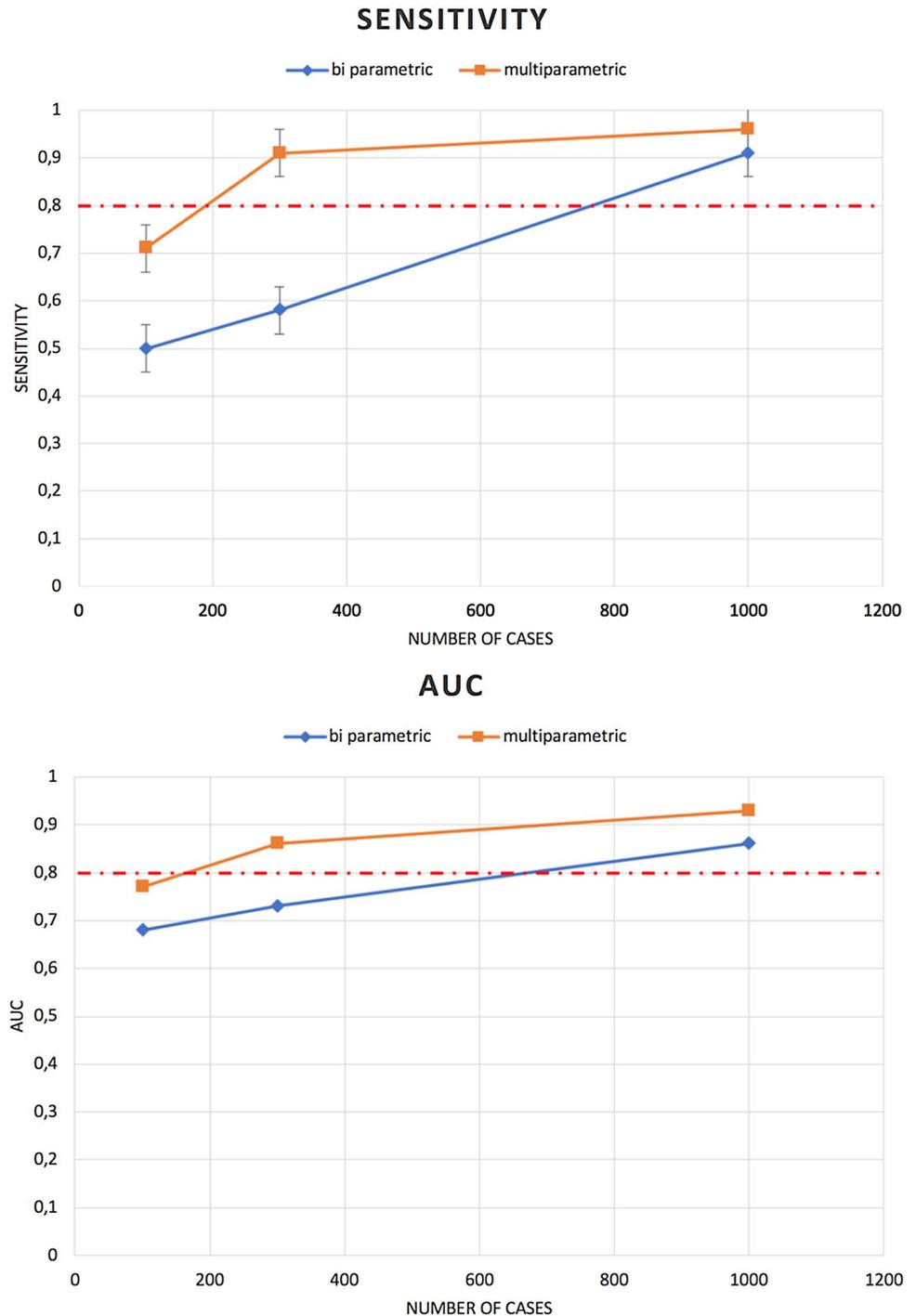
DCE sequences contribution to higher sensitivity had two different elements: 57% of the improvement was due to the detection of suspect areas that were missed during the biparametric evaluation, whereas 43% of it was represented by peripheral zone lesions that were upgraded from 3 to 4 for their typical contrast enhancement. This last finding is in agreement with Greer et al. [24], who reported a higher sensitivity for detecting index lesions

with threshold PIRADS 4 for experienced readers: less experienced readers were more likely to score a lesion as PIRADS 3. A possible explanation is that the criterion "mild-to-moderate" versus "marked" signal abnormality on high-b-value DWI [4] is inherently subjective and experience related [25]. Therefore, for less-experienced readers, DCE helped the choice of a final interpretation score between PIRADS 3 and 4 and, moreover, provided

useful information for the detection by drawing attention to foci which had been overlooked in the biparametric mode. Among the various patients' and PCa features, only a small prostate volume was significantly associated with a decreased cancer detection rate ($p = 0.02$). The heterogeneous appearance of smaller prostates, in some cases related to the TURP procedure, is likely to have hindered the assessment by the less-experienced readers. Lesion size and lesion grading, that were reported by other authors to be associated with detection by MRI [18, 21], showed no significant association.

The results on Group C, formed by young residents with about 100 cases, showed that they still had to increase their experience to reach an acceptable accuracy even with the multiparametric protocol: their diagnostic parameters are definitely inferior to those of the readers with 300 cases: SNS = 0.71 versus 0.91, $p < 0.0001$ and AUC = 0.77 versus 0.86, $p = 0.07$. This result confirms the findings of other authors. A recent retrospective study on novice radiologists showed an initial rapid improvement up to 40 mpMRIs, followed by a slowdown. According to the authors, the accuracy levels reached in this context, inferior to those of multi-year prostate-dedicated experienced radiologists, suggested the existence of an intermediate phase in which learning proceeds at much slower pace [18]. Also, Gaziev et al. [17] reported a nearly doubled detection rate after 2 years of training and about 330 bpMRI. The magnitude of the improvement was, however, partly tempered by the evolution in mpMRI indications during the study and the existence of a parallel urologic learning curve for target-biopsies [17].

Overall, our results indicate that for reaching the needed accuracy with the bpMRI protocol, 300 cases are not sufficient: based on the almost-linear increase of SNS and AUC with the number of cases (every 100 cases + 0.05 points for SNS and + 0.02 for AUC), a minimum of 700–800 cases seems to be necessary.

Because the difference in performance between bpMRI and mpMRI is dependent on the experience of the reader, the comparison with works lacking this piece of information is not straightforward. To our knowledge, there is only one study that directly compared the diagnostic accuracy of bpMRI to the one achieved with the use of mpMRI using the reference standard of RP specimens. Scialpi et al. [15] carried out a retrospective analysis of mpMRI at 3 T of 41 patients who underwent radical prostatectomy and found that the sensitivity for detecting index lesions was the same for bpMRI and mpMRI. They, however, give no information on the readers' background; by the context and by analogy with our unit, we presume they were radiologists routinely performing these examinations as staff of University and Hospital departments and thus with a good record of examined cases. If this is the case, our results pointing to equality

between the two protocols for 700–800 cases are consistent with theirs.

Agreement on equivalent performance with bpMRI and mpMRI when the reading is done by expert radiologists is reported also by authors [10–14, 16] who evaluated the cancer detection rate at 3T using the biopsy findings as reference standards. Thestrup et al. [12], studying a series of 204 PCa cases read by two radiologists specialized in prostate MR, reported that sensitivity and specificity were nearly the same when assessed by a bpMRI (including only the axial T2W and the DWI) or mpMRI (high sensitivity and low specificity). Stanzione et al. [11] reported no significant difference in the identification of true positives (26/29 patients with mpMRI compared to 24/29 with bpMRI) by two radiologists with a 10-years' experience. Kuhl et al. [16] and Barth et al. [13] studied the diagnostic accuracy for clinically significant prostate cancer achieved with abbreviated bpMRI, based only on single-plane (axial) T2W images and diffusion-weighted imaging at 0, 800, 1000, and 1400 mm^2/s acquired within 8 min and 45 s: they obtained a diagnostic accuracy and cancer detection rate equivalent to those of conventional full multiparametric contrast-enhanced MR imaging protocols. However, Barth et al. [13] recognized that such results could not be directly applicable to settings different from high-field mpMRI (3.0 T) with endorectal coil (ERC) and interpreted by highly experienced readers (6-year experience in prostate MR). De Visschere et al. [14] found that the added value of DCE over T2-WI and DWI was limited (DCE findings were redundant in 80.8% of patients) when using PI-RADSv2 for diagnosis of PCa in patients with elevated PSA before biopsy, but there is no mention of the reader's experience.

In the present study, all prostate MRI examinations were performed using a 1.5 T scanner, which are generally more widespread than 3 T scanners, and without the use of ERC, which improves patients' acceptability. We used the classification proposed by De Visschere el al. for the bpMRI protocol, even if their analysis was validated at 3 T (T). Ullrich et al. [26], in fact, showed that a 1.5 T scanner has an objective image quality in T2WI comparable to that obtained with a 3 T one; furthermore, the PIRADS scoring and diagnostic performance were independent from the field strength and also from the usage of the endorectal coil. With these settings, our expert readers achieved equally good performing levels with bpMRI as with mpMRI, yielding values of sensitivity, specificity and AUC around 0.90, in agreement with the literature [11, 13, 14]. As the paper addresses the value of MRI in the perspective of a triage test, performance on non-index lesion (often smaller than 0.5 ml) was not presented, despite whole-mount section pathology review.

This study has some limitations. First, it was a retrospective, single-center study on a limited number of patients. Second, RP series tend to select patients with an aggressive

disease, as witnessed by the absence of any ISUP grade 1 (i.e., Gleason 6) index lesion, potentially improving sensitivity. Third, the standard reference adopted for controls is not infallible, although a long-term follow-up, a stable PSA and further histologic and/or radiologic confirmations were obtained for all of them.

In conclusion, we evidenced significant differences between readers of different experience when facing PCa detection using either mpMRI or bpMRI: the DCE sequences had an important role in boosting sensitivity and AUC in readers with experience of about 300 cases, while it could have been safely omitted for readers with a threefold experience. In practical terms, bpMRI seems to be a valid technique in PCa detection, as long as it is restricted to readers with a record of at least 700–800 cases. Readers with 100 cases are yet to achieve good performance already in mpMRI and even more so in bpMRI.

Compliance with ethical standards

References

1. Ahmed HU, El-Shater Bosaily A, Brown LC, et al (2017) Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. Lancet Lond Engl 389:815–822. https://doi.org/10.1016/S0140-6736(16)32401-1

2. Mottet N, Bellmunt J, Bolla M, et al (2017) EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. Eur Urol 71:618–629. https://doi.org/10.1016/j.eururo.2016.08.003

3. Marra G, Gontero P, Valerio M (2016) Changing the prostate cancer management pathway: why Focal Therapy is a step forward. Arch Esp Urol 69:271–280

4. Weinreb JC, Barentsz JO, Choyke PL, et al (2016) PI-RADS Prostate Imaging – Reporting and Data System: 2015, Version 2. Eur Urol 69:16–40. https://doi.org/10.1016/j.eururo.2015.08.052

5. Delongchamps NB, Beuvon F, Eiss D, et al (2011) Multiparametric MRI is helpful to predict tumor focality, stage, and size in patients diagnosed with unilateral low-risk prostate cancer. Prostate Cancer Prostatic Dis 14:232–237. https://doi.org/10.1038/pcan.2011.9

6. Tamada T, Sone T, Higashi H, et al (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 197:664–670. https://doi.org/10.2214/AJR.10.5923

7. Schimmöller L, Quentin M, Arsov C, et al (2014) MR-sequences for prostate cancer diagnostics: validation based on the PI-RADS scoring system and targeted MR-guided in-bore biopsy. Eur Radiol 24:2582–2589. https://doi.org/10.1007/s00330-014-3276-9

8. Rais-Bahrami S, Siddiqui MM, Vourganti S, et al (2015) Diagnostic value of biparametric magnetic resonance imaging (MRI) as an adjunct to prostate-specific antigen (PSA)-based detection of prostate cancer in men without prior biopsies. BJU Int 115:381–388. https://doi.org/10.1111/bju.12639

9. Fascelli M, Rais-Bahrami S, Sankineni S, et al (2016) Combined Biparametric Prostate Magnetic Resonance Imaging and Prostate-specific Antigen in the Detection of Prostate Cancer: A Validation Study in a Biopsy-naive Patient Population. Urology 88:125–134. https://doi.org/10.1016/j.urology.2015.09.035

10. Radtke JP, Boxler S, Kuru TH, et al (2015) Improved detection of anterior fibromuscular stroma and transition zone prostate cancer using biparametric and multiparametric MRI with MRI-targeted biopsy and MRI-US fusion guidance. Prostate Cancer Prostatic Dis 18:288–296. https://doi.org/10.1038/pcan.2015.29

11. Stanzione A, Imbriaco M, Cocozza S, et al (2016) Biparametric 3T Magentic Resonance Imaging for prostatic cancer detection in a biopsy-naïve patient population: a further improvement of PI-RADS v2? Eur J Radiol 85:2269–2274. https://doi.org/10.1016/j.ejrad.2016.10.009

12. Thestrup KCD, Logager V, Baslev I, et al (2016) Biparametric versus multiparametric MRI in the diagnosis of prostate cancer. Acta Radiol Open 5:2058460116663046. https://doi.org/10.1177/2058460116663046

13. Barth BK, De Visschere PJL, Cornelius A, et al (2017) Detection of Clinically Significant Prostate Cancer: Short Dual-Pulse Sequence versus Standard Multiparametric MR Imaging-A Multireader Study. Radiology 162020. https://doi.org/10.1148/radiol.2017162020

14. De Visschere P, Lumen N, Ost P, et al (2017) Dynamic contrast-enhanced imaging has limited added value over T2-weighted imaging and diffusion-weighted imaging when using PI-RADSv2 for diagnosis of clinically significant prostate cancer in patients with elevated PSA. Clin Radiol 72:23–32. https://doi.org/10.1016/j.crad.2016.09.011

15. Scialpi M, Prosperi E, D'Andrea A, et al (2017) Biparametric versus Multiparametric MRI with Non-endorectal Coil at 3T in the Detection and Localization of Prostate Cancer. Anticancer Res 37:1263–1271. https://doi.org/10.21873/anticanres.11443

16. Kuhl CK, Bruhn R, Krämer N, et al (2017) Abbreviated Biparametric Prostate MR Imaging in Men with Elevated Prostate-specific Antigen. Radiology 285:493–505. https://doi.org/10.1148/radiol.2017170129

17. Gaziev G, Wadhwa K, Barrett T, et al (2016) Defining the learning curve for multiparametric magnetic resonance imaging (MRI) of the prostate using MRI-transrectal ultrasonography (TRUS) fusion-guided transperineal prostate biopsies as a validation tool. BJU Int 117:80–86. https://doi.org/10.1111/bju.12892

18. Rosenkrantz AB, Ayoola A, Hoffman D, et al (2017) The Learning Curve in Prostate MRI Interpretation: Self-Directed Learning Versus Continual Reader Feedback. AJR Am J Roentgenol 208:W92–W100. https://doi.org/10.2214/AJR.16.16876

19. Vargas HA, Hötker AM, Goldman DA, et al (2016) Updated prostate imaging reporting and data system (PIRADS v2) recommendations for the detection of clinically significant prostate cancer using multiparametric MRI: critical evaluation using whole-mount pathology as standard of reference. Eur Radiol 26:1606–1612. https://doi.org/10.1007/s00330-015-4015-6

20. Ahmed HU (2009) The index lesion and the origin of prostate cancer. N Engl J Med 361:1704–1706. https://doi.org/10.1056/NEJMcibr0905562

21. Russo F, Regge D, Armando E, et al (2016) Detection of prostate cancer index lesions with multiparametric magnetic resonance imaging (mp-MRI) using whole-mount histological sections as the reference standard. BJU Int 118:84–94. https://doi.org/10.1111/bju.13234

22. Epstein JI, Amin MB, Reuter VE, Humphrey PA (2017) Contemporary Gleason Grading of Prostatic Carcinoma: An Update With Discussion on Practical Issues to Implement the 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma. Am J Surg Pathol 41:e1–e7. https://doi.org/10.1097/PAS.0000000000000820

23. Hanley JA, McNeil BJ (1983) A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 148:839–843. https://doi.org/10.1148/radiology.148.3.6878708

24. Greer MD, Brown AM, Shih JH, et al (2017) Accuracy and agreement of PIRADSv2 for prostate cancer mpMRI: A multireader study. J Magn Reson Imaging JMRI 45:579–585. https://doi.org/10.1002/jmri.25372

25. Rosenkrantz AB, Oto A, Turkbey B, Westphalen AC (2016) Prostate Imaging Reporting and Data System (PI-RADS), Version 2: A Critical Look. AJR Am J Roentgenol 206:1179–1183. https://doi.org/10.2214/AJR.15.15765

26. Ullrich T, Quentin M, Oelers C, et al (2017) Magnetic resonance imaging of the prostate at 1.5 versus 3.0T: A prospective comparison study of image quality. Eur J Radiol 90:192–197. https://doi.org/10.1016/j.ejrad.2017.02.044