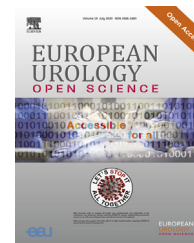


available at www.sciencedirect.com
journal homepage: www.eu-openscience.europeanurology.com



Prostate Cancer

Accuracy of Prostate Magnetic Resonance Imaging: Reader Experience Matters

Hyunseon C. Kang^{a,*}, Nahyun Jo^a, Anas Saeed Bamashmos^a, Mona Ahmed^a, Jia Sun^b, John F. Ward^c, Haesun Choi^a

^a Department of Abdominal Imaging, The University of Texas MD Anderson Cancer Center, Houston, TX, USA; ^b Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA; ^c Department of Urology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Article info

Article history:

Accepted March 6, 2021

Associate Editor:

Guillaume Ploussard

Keywords:

Likert scoring system
Magnetic resonance imaging/
ultrasonography fusion biopsy
Prostate magnetic resonance
imaging
Prostate cancer
Reader experience, Double read

Abstract

Background: Prostate magnetic resonance imaging (MRI) is increasingly used in the detection, image-guided biopsy, and active surveillance of prostate cancer. The accuracy of prostate MRI may differ based on factors including imaging technique, patient population, and reader experience.

Objective: To determine whether the accuracy of prostate MRI varies with reader experience.

Design, setting, and participants: We rescored regions of interest from 194 consecutive patients who had undergone MRI/ultrasonography fusion biopsy. Original prostate MRI scans had been interpreted by one of 33 abdominal radiologists (AR group). More than 14 mo later, rescoring was performed by two blinded, prostate MRI radiologists (PR group). Likert scoring was used for both original MRI reports and rescoring.

Outcome measurements and statistical analysis: Test performance (sensitivity, specificity, positive predictive value [PPV], and negative predictive value [NPV]) of prostate MRI was defined for the AR and PR groups. A Likert score of 4–5 was considered test positive and clinically significant prostate carcinoma (csPCa; Gleason grade group [GGG] ≥ 2) was considered outcome positive.

Results and limitations: MRI-positive lesions (Likert 4–5) scored by the PR group resulted in csPCa more frequently than those scored by the AR group (64.9% vs 39.3%). MRI-negative lesions (Likert 2–3) were more likely to result in a clinically insignificant biopsy (benign pathology or GGG 1) when scored by the PR versus the AR group (91.8% vs 76.6%). Sensitivity and specificity of MRI to detect csPCa were higher for the PR group than for the AR group (sensitivity 85.9% vs 70.7%; specificity 77.3% vs 46.8%). Overall diagnostic accuracy was higher for the PR group than for the AR group (80.1% vs 54.6%).

* Corresponding author. Department of Abdominal Imaging, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd., Unit 1473, Houston, TX 77030, USA. Tel. +1-713-794-4100; Fax: +1-713-794-4379.

E-mail address: hkang2@mdanderson.org (H.C. Kang).



Conclusions: Sensitivity, specificity, PPV, and NPV of prostate MRI were higher for the PR group than for the AR group.

Patient summary: We examined the accuracy of prostate magnetic resonance imaging (MRI) in two groups of radiologists. Experienced radiologists were more likely to detect clinically significant prostate cancer on MRI.

© 2021 The Authors. Published by Elsevier B.V. on behalf of European Association of Urology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

The traditional pathway of prostate cancer (PCa) diagnosis involves a random, systematic biopsy of the prostate using transrectal ultrasound (TRUS) guidance. The sampling error with this technique results in false negative findings in men with clinically significant PCa (csPCa), imprecise risk stratification, and overdiagnosis of clinically insignificant PCa that can lead to patient anxiety and unnecessary treatment-related morbidity [1].

Multiparametric magnetic resonance imaging (mpMRI) has emerged as a valuable tool for PCa detection and management, as it allows for accurate localization of the tumor and facilitates image-guided biopsy [2]. Magnetic resonance imaging (MRI)-guided biopsy may be performed within the bore of the MRI scanner or, more commonly, under ultrasound guidance using visual estimation or software fusion [3]. In particular, developments in image fusion software capable of overlaying MRI findings onto real-time ultrasound images have led to a paradigm shift in PCa diagnosis, as suspicious findings on prostate MRI can directly be targeted during TRUS biopsy [4]. Several well-performed diagnostic accuracy studies and randomized controlled trials involving mpMRI-targeted biopsies have demonstrated that this approach outperforms systematic TRUS biopsies [5,6].

As with other modalities, there is a learning curve for the interpretation of mpMRI [7,8]. This is particularly true since PCa can overlap in imaging appearance with several common benign entities and, furthermore, may have indistinct margins [9]. The degree of experience needed to interpret prostate MRI accurately, however, is not known. On the one hand, consensus panels have recommended a minimum of 50–100 interpretations to be considered qualified for interpreting prostate MRI and at least 1000 cases to be considered an expert [10–12]. On the other hand, a recent study found no benefit in diagnostic accuracy for radiologists who had read >500 mpMRI scans compared with those who had read <500 scans [13].

As prostate MRI continues to be adopted more widely, it is important to consider the degree of reader experience needed for diagnostic accuracy. In our practice, mpMRI is interpreted during the clinical workflow by one of 33 abdominal radiologists (AR group). If MRI/TRUS fusion biopsy is needed, one of two dedicated prostate MR radiologists (PR group) performs prostate segmentation and annotation. Thus, all cases referred for MRI/TRUS fusion biopsy are reviewed twice, once during the clinical workflow and once during segmentation/annotation. This allows

us to compare diagnostic accuracy (positive predictive value [PPV], negative predictive value [NPV], sensitivity, and specificity) in the two groups of radiologists, using the same cohort of cases. We hypothesized that accuracy measures would be higher for the dedicated prostate radiologists.

2. Patients and methods

This retrospective study was approved by our institutional review board, with waiver of informed consent. Consecutive patients ($n = 199$) who had MRI/TRUS fusion biopsy from January 2015 through March 2016 were included. Five patients were excluded due to incomplete imaging ($n = 1$), loss of the original segmentation report ($n = 2$), or referral for extraprostatic mass ($n = 1$) or metastasis ($n = 1$). For reference, 1926 prostate MRI scans were performed at our institution during this period, the vast majority of which were for preoperative staging; thus, approximately 10% of prostate MRI patients were referred for fusion biopsy.

2.1. Data acquisition

Patients were imaged at our institution on a 1.5 or 3 T GE HealthCare Signa HDx MR scanner (GE Healthcare, Waukesha, WI, USA) or a 1.5 T Siemens MR scanner (Siemens, Erlangen, Germany) using an eight-channel abdominal array coil and, for all except one, an endorectal coil (MR Innerva; Medrad, Pittsburgh, PA, USA). Acquisition parameters evolved during the study period. A typical MR protocol included smaller field-of-view axial, sagittal, and coronal fast spin-echo T2-weighted imaging; diffusion weighted imaging with b values of 50–100 and 700–800 s/mm^2 and apparent diffusion coefficient (ADC) reconstruction; and dynamic-contrast-enhanced imaging (DCE), as well as whole-pelvis T1-weighted imaging and diffusion-weighted imaging with ADC reconstruction. DCE MRI was performed after an intravenous injection of gadopentetate dimeglumine (Gadavist; Bayer HealthCare Pharmaceuticals) at 0.1 mmol/kg of body weight at a rate of 3 ml/s via a power injector, with <14 s temporal resolution. Thirteen patients with outside institution MRI scans were included, with MRI performed without an endorectal coil in 12 of them.

MRI scans were prospectively read at our institution by one of 33 abdominal radiologists (AR group) who rotated through the body MR service, with 1–10 yr (median 8) of experience. For outside institution scans, the original report ($n = 10$) or reinterpretation ($n = 3$) was available. Most reports utilized a Likert scale to assign a suspicion for the presence of PCa (Table 1). For unstructured reports ($n = 65$), the intended level of suspicion was inferred based on the terminology in the report, such as “indeterminate, equivocal for carcinoma” (Likert 3); “suspicious, probable tumor” (Likert 4); or “highly likely, compatible with tumor, extraprostatic extension” (Likert 5).

2.2. MRI/TRUS fusion biopsy

Patients were referred for fusion biopsy based on MRI findings and/or clinical suspicion (indications listed in Table 2), and reviewed by a

Table 1 – Likert scoring system

Likert score	Description
1	Clinically significant cancer is highly unlikely to be present
2	Clinically significant cancer is unlikely to be present
3	Clinically significant cancer is equivocal
4	Clinically significant cancer is likely to be present
5	Clinically significant cancer is highly likely to be present

Table 2 – Patient demographics (n=194)

Characteristic	Median	Range
Age (yr)	65	40–88
PSA (ng/ml)	6.6	
Prostate volume on TRUS (ml)	47	13–188
Number of lesions on MRI	3	1–6
	No.	%
Ethnicity		
Caucasian	157	80.9
African American	19	9.8
Asian	7	3.6
Latin American (Hispanic)	8	4.1
Unknown	3	1.5
Prior evaluation		
First biopsy	20	10.3
Prior biopsy	174	89.7
Negative	109	56.2
Positive	65	33.5
Indication ^a		
Elevated PSA	148	75.9
Active surveillance	79	40.5
Prior negative biopsy	76	39.0
Rising PSA	79	40.5
Radiation therapy with rising PSA	11	5.6
Cryoblation with rising PSA	7	3.6
HIFU with rising PSA	1	0.5
Suspicious DRE	8	4.1
Elevated PCA3	8	4.1
Elevated 4K score	1	0.5
Abnormal MRI	2	1.0

DRE = digital rectal examination; HIFU = high-intensity focused ultrasonography; MRI = magnetic resonance imaging; PCA3 = prostate cancer antigen 3; PSA = prostate-specific antigen; TRUS = transrectal ultrasonography.

^a A single patient may have more than one indication for biopsy.

dedicated PR group radiologist (H.C. or H.C.K. with 9 and 3 yr of experience, respectively, in mpMRI interpretation and >250 cases of prostate segmentation/annotation with pathological correlation prior to study initiation). For those patients who decided to proceed with biopsy, targeted biopsy was performed for Likert ≥ 2 lesions detected by the AR reader (mentioned in the MRI report), and any additional Likert ≥ 3 lesions detected by the PR group radiologist at the time of annotation. Profuse software (Eigen, Grass Valley, CA, USA) was used for prostate segmentation and annotation.

Transrectal biopsy was performed with the Artemis system (Eigen); two to three cores were taken from each region of interest (ROI) marked on the MRI images. Systematic biopsy cores acquired during the same session were not included in the analysis, as they could not be assigned

accurately to particular MRI findings. All biopsies were performed by a single urologist with 3 yr fusion biopsy experience prior to the study period.

Biopsies were reported by one of six specialized uropathologists. The final Gleason grade group (GGG) was used for this study, with GGG ≥ 2 (ie, Gleason score $\geq 3 + 4$) considered csPCa [14].

2.3. Rescoring of biopsied lesions

When the PR group performs segmentation/annotation, they have access to the MRI report (by the AR group), so they may be perceived to have an advantage in detecting lesions. In order to remove potential bias, a prolonged washout period of 14 mo was used, after which PR group radiologists independently rescored all biopsied ROIs using the Likert scale. Screenshots from the original annotation were provided to denote ROI location, as only biopsied lesions could be reanalyzed. Although the location of the ROI was available, PR group radiologists were not aware of its Likert score, or whether the ROI had been detected by the AR group, PR group, or both. They were blinded to the original MRI report, histopathology, and clinical information. All sequences of the mpMRI were available. Lesions with a discrepant score were graded in consensus (28 out of 302).

2.4. Statistical analysis

Using GGG as the gold standard, the PPV of each Likert score to predict csPCa was calculated for the PR and AR groups. Only lesions graded by both groups were included in this analysis ($n = 302$). The generalized estimation equation method, which takes into account correlations of readings from the same lesion, was used to calculate p values [15].

Subsequently, the test performance of a “positive” or “negative” MRI scan was calculated using 2×2 contingency tables for the PR and AR groups. For this analysis, Likert 4–5 was considered test positive, while Likert 2–3 was considered test negative, consistent with prior recommendations [16,17]. A biopsy result of GGG ≥ 2 was considered condition positive, while a biopsy result of GGG ≤ 1 (ie, Gleason score ≤ 6) was considered condition negative. Sensitivity, specificity, PPV, and NPV were calculated using McNemar’s test for diagnostic accuracy. A p value of <0.05 was considered statistically significant.

Cohen’s kappa statistic was used to assess inter-reader agreement between the two radiologists in the PR group, and between each radiologist in the PR group and the reading radiologist in the AR group. Agreement was defined as substantial ($\kappa \geq 0.61$), moderate ($\kappa = 0.41–0.60$), fair ($\kappa = 0.21–0.40$), and poor ($\kappa \leq 0.20$). Statistical analyses were performed using SAS version 9.4 (SAS Institute, Cary, NC, USA).

3. Results

Table 2 shows baseline demographics of the 194 patients. The median age at biopsy was 65 yr, and the prebiopsy median prostate-specific antigen (PSA) level was 6.6 ng/ml. An elevated PSA level was the most common indication for fusion biopsy ($n = 148$), followed by active surveillance ($n = 79$) and rising PSA ($n = 79$).

Figure 1 and Table 3 show the distribution of Likert scores for the 302 lesions identified in the MRI reports. The AR group scored more than twice as many Likert 4 lesions as did the PR group. Conversely, a higher number of Likert 2, 3, and 5 lesions were scored by the PR group.

The probability of detecting csPCa increased with Likert scores in both the AR and the PR group, but the increase was more marked for the PR group (Fig. 2). The probability of

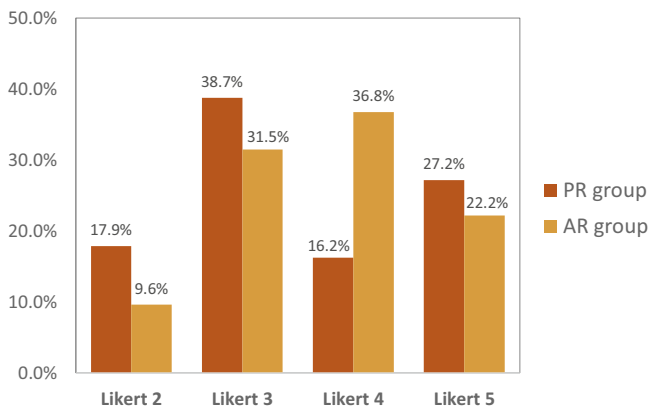


Fig. 1 – Distribution of Likert scores for the two dedicated prostate MRI radiologists (PR group) and the 33 abdominal imaging radiologists (AR group). The AR group scored a higher percentage of lesions as Likert 4, compared with the PR group. Conversely, the PR group scored higher numbers of Likert 2, 3, and 5 lesions. AR = abdominal radiologist; MRI = magnetic resonance imaging; PR = prostate MRI radiologist.

detecting csPCa in a Likert 5 lesion was significantly higher for the PR group than for the AR group (Fig. 2 and Table 3). The probability of detecting csPCa was lower for Likert 2 or 3 lesions assigned by the PR group versus the AR group (Fig. 2 and Table 3). There was no significant difference in the detection of csPCa between Likert 4 lesions scored by the two groups. Examples of discrepant scoring between the AR and PR groups are shown in Figures 3 and 4.

An MRI-positive lesion had a higher PPV for csPCa when scored by the PR group than when scored by the AR group (64.9% vs 39.3%, $p > 0.0001$). An MRI-negative lesion scored by the PR group had a higher NPV to result in a clinically insignificant biopsy (either benign pathology or Gleason 6) than that scored by the AR group (91.8% vs 76.6%, $p < 0.0001$; Table 4).

Sensitivity for the detection of csPCa was higher for the PR group than for the AR group (85.9% vs 70.7%; $p = 0.0027$). Specificity was also higher for the PR group than for the AR group (77.3% vs 46.8%, $p < 0.0001$). Overall diagnostic accuracy was significantly higher for the PR group than for the AR group (80.1% vs 54.6%, $p < 0.0001$; Table 4).

Inter-reader agreement for the two PR group radiologists was substantial ($\kappa = 0.7$, $p < 0.001$). Inter-reader agreement between each radiologist in the PR group and the original AR group reader was only fair ($\kappa = 0.26$ and 0.24 , respectively; both $p < 0.001$).

A total of 272 additional lesions were annotated by the PR group radiologist at the time of segmentation. Of these, 34 lesions (12.5%) in 19 patients resulted in csPCa.

4. Discussion

This study shows that the utility of prostate MRI to detect csPCa depends on reader expertise. The two radiologists who routinely reviewed/annotated mpMRI and had frequent histopathological feedback on a per-lesion basis had higher accuracy and reproducibility than the group of radiologists who interpreted scans without being involved in fusion biopsy. The PR group detected csPCa in 64.9% of MRI-positive lesions, while the AR group detected csPCa in 39.3% of MRI-positive lesions. All radiologists were familiar with prostate MRI, which had been performed in our department for 7 yr prior to the study period.

Interpretation of prostate MRI requires a reliable, easily understood method of communicating the level of suspicion for PCa based on MRI findings. The Likert scale has been recommended for national implementation in the UK by expert consensus [10,16], while Prostate Imaging Reporting and Data System (PI-RADS) has been proposed by the European Society of Urogenital Radiology and American College of Radiology [18].

However, considerable inter-reader variability has been reported in assignment of both Likert and PI-RADS scores and subsequent detection of csPCa, with experienced readers showing greater interobserver agreement and higher accuracy [11,19–24]. For example, Shin et al [19] found that detection rates of csPCa varied from 47.3% for more experienced to 28.6% for less experienced readers. Likewise, Gaziev et al [20] found that the same two radiologists experienced an increase in cancer detection rate from 42% to 81%, as they gained experience.

Conversely, studies by Pickersgill et al [13] and Di Campli et al [25] have found no significant differences in diagnostic accuracy for more versus less experienced radiologists.

Table 3 – Distribution of Likert scores and positive predictive value of each Likert score to detect clinically significant (GGG ≥ 2) prostate carcinoma between dedicated prostate MRI radiologists (PR group) and abdominal imaging radiologists (AR group)

Likert score	Group	No. of lesions % (n/N)	Pathology		p value ^a
			Benign or GGG = 1 % (n/N)	GGG ≥ 2 % (n/N)	
Likert 5	PR	27.2 (82/302)	18.3 (15/82)	81.7 (67/82)	<0.0001
	AR	22.2 (67/302)	44.8 (30/67)	55.2 (37/67)	
Likert 4	PR	16.2 (49/302)	63.3 (31/49)	36.7 (18/49)	0.304
	AR	36.8 (111/302)	70.3 (78/111)	29.7 (33/111)	
Likert 3	PR	38.7 (117/302)	88.9 (104/117)	11.1 (13/117)	0.003
	AR	31.5 (95/302)	76.8 (73/95)	23.2 (22/95)	
Likert 2	PR	17.9 (54/302)	98.2 (53/54)	1.9 (1/54)	0.007
	AR	9.6 (29/302)	75.9 (22/29)	24.1 (7/29)	

GGG = Gleason grade group.

^a A p value refers to the significance of the difference in positive predictive values between the PR and AR groups for each Likert score.

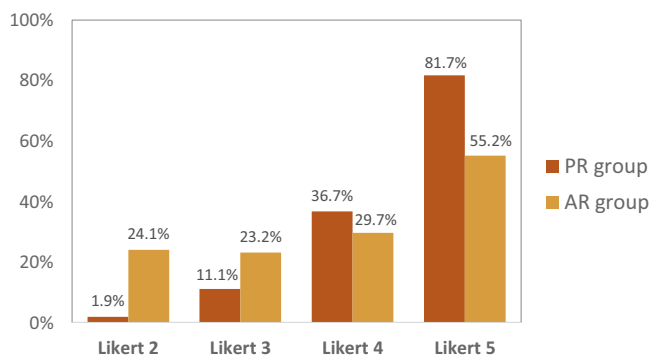


Fig. 2 – The probability of detecting clinically significant prostate carcinoma per Likert score assigned by the two dedicated prostate MRI radiologists (PR group) and the 33 abdominal imaging radiologists (AR group). AR = abdominal radiologist; MRI = magnetic resonance imaging; PR = prostate MRI radiologist.

Different definitions of what constitutes experience may explain these conflicting results. In the study by Shin et al [19], experienced readers interpreted >80 mpMRI scans with subsequent fusion biopsy during the study period (2 yr 10 mo), while less experienced radiologists read <50 scans during the same period. In the study by Gaziev et al [20], two radiologists reviewed and segmented 200 cases for MRI/TRUS fusion biopsy (~100 cases per radiologist) between the first and last cohorts, while their cancer detection rate increased. In contrast, Pickersgill et al [13] considered radiologists who had clinically interpreted >500 mpMRI scans to be experienced. However, during the study period when fusion biopsy was performed, 469 scans were divided among nine radiologists. This likely led to similar degrees of experience with MRI/TRUS fusion and histopathological correlation, possibly accounting for the lack of

difference in accuracy measures for radiologists with more versus less reading experience.

The aforementioned studies demonstrated higher cancer detection rates for radiologists who had experience with fusion biopsy and thus histopathological correlation. Indeed, dedicated training including pathological correlation has been found to improve significantly the accuracy of prostate MRI interpretation for less experienced radiologists [8,26].

Radiology-pathology feedback has been shown to provide valuable learning opportunities for radiologists in other areas as well [27]. A study examining the accuracy of screening mammogram interpretation found that radiologists who followed up cases in which they recommended further workup had higher specificity, while the interpretation of a higher volume of mammograms was not associated with better performance [28]. Similarly, diagnosis of extramural venous invasion in rectal cancer improved significantly following targeted training which included histological correlation, while there was no improvement in diagnostic accuracy for radiologists who performed clinical work without the targeted training [29].

Findings from the current study support the notion that both the volume of cases interpreted and histopathological feedback play a role in the development of expertise. The average volume of clinical prostate MRI interpretations was not significantly different between the PR and AR groups. However, radiologists in the PR group had the benefit of ROI annotation and per-lesion histopathological correlation, which likely led to improved accuracy and reproducibility. These results may help plan interventions to increase reader accuracy and decrease variability of prostate MRI interpretation. An educational curriculum focused on correlating MRI findings with histopathology could provide an avenue for rapidly increasing reader expertise.

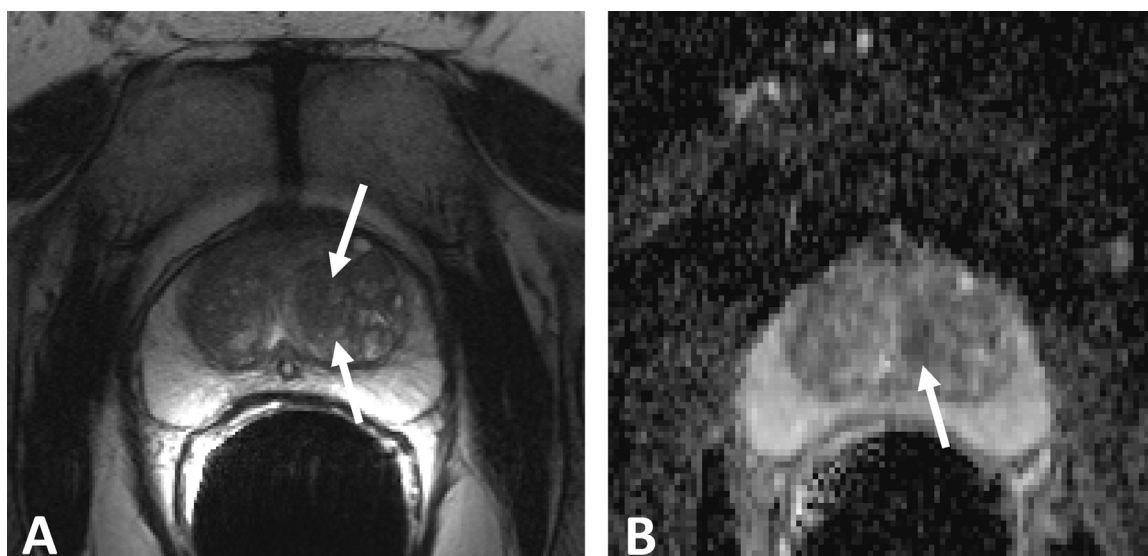


Fig. 3 – Discrepant scoring between the AR and PR groups. A 65-yr-old male on active surveillance for GGG 1 presented for an elevated PSA level of 27.3 ng/ml. An 8-mm focus (arrows) of (A) mild T2 hypointensity and (B) restricted diffusion in the left transition zone was assigned a Likert score of 5 on the MRI report. Both PR group radiologists graded this as Likert 2. Fusion biopsy showed focal atrophy without tumor. Review of the medical record revealed that the patient had been training for a cycling race prior to the elevated PSA sample. Two weeks after cessation of training, a repeat PSA value was 11.4 ng/ml. AR = abdominal radiologist; GGG = Gleason grade group; MRI = magnetic resonance imaging; PR = prostate MRI radiologist; PSA = prostate-specific antigen.

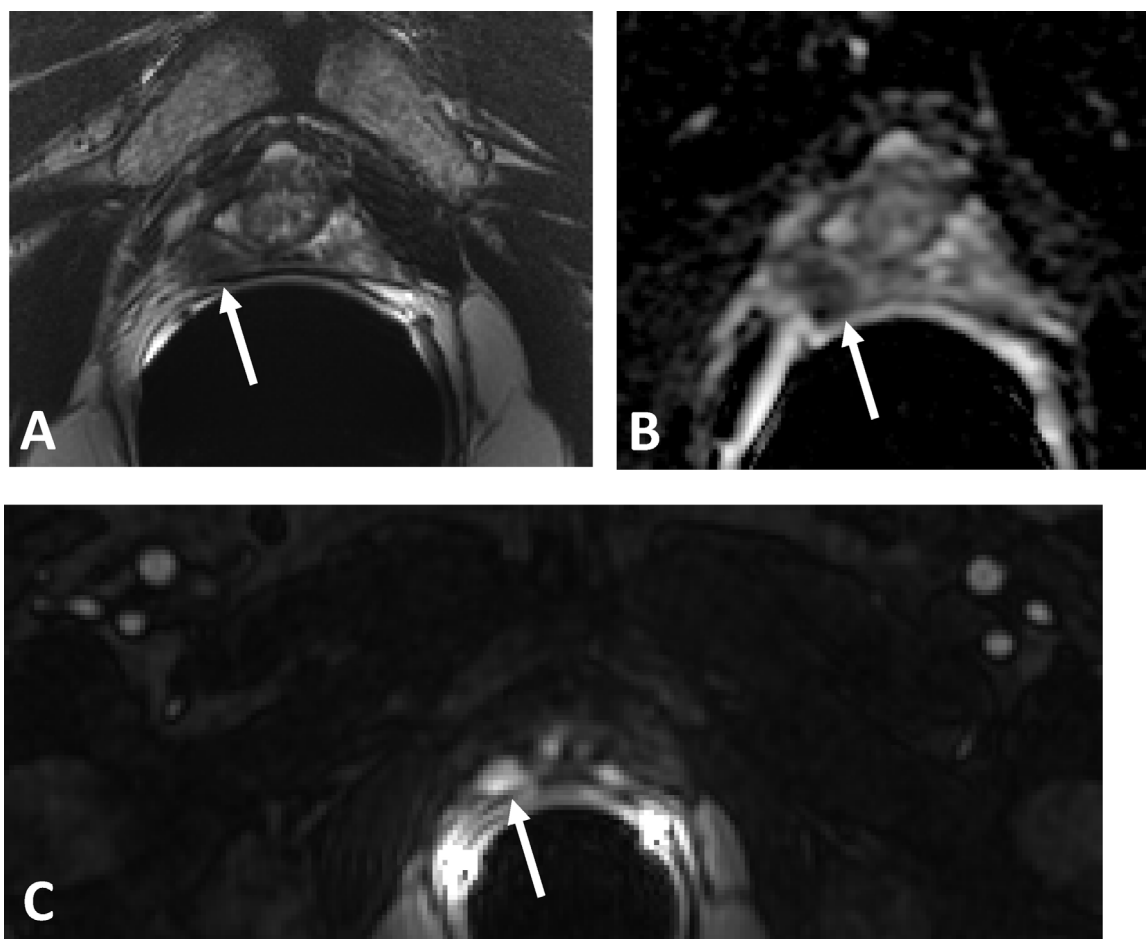


Fig. 4 – Discrepant scoring between the AR and PR groups. A 69-yr-old male presented with a rising PSA level following cryoablation 4 yr ago. A 1-cm focus of (A) moderate T2 hypointensity and (B) restricted diffusion in the right lateral midgland demonstrates early arterial enhancement on (C) postcontrast images (arrows). This was assigned a Likert score of 3 on the MRI report. Both PR group radiologists graded this as Likert 5. Fusion biopsy showed GGG 4. AR = abdominal radiologist; GGG = Gleason grade group; MRI = magnetic resonance imaging; PR = prostate MRI radiologist; PSA = prostate-specific antigen.

Our study is limited by its single-center retrospective design. In order to remove bias from the original reports, the PR group rescored lesions in a blinded fashion after a prolonged washout period. Despite our efforts, an additional bias may have been introduced since locations of prior biopsies were provided. Additionally, significant cancers may have been missed because they were not targeted or sampled at systematic biopsy. Although a prostatectomy

cohort would allow more definitive histopathology, this would not allow inclusion of patients without cancer. Furthermore, many patients with equivocal lesions do not undergo prostatectomy. Our study mirrors the utilization of MRI in current practice, as a significant proportion of patients imaged with mpMRI do not have follow-up biopsies.

A limitation of our technique is that high b -value imaging was not part of the prostate MR protocol during the acquisition period. Subsequently, after implementing a high b -value ($b = 1400$) sequence, our radiologists' feedback is that high b -value images rarely provide additional information beyond what is already seen on routine multiple b -value diffusion-weighted images with ADC reconstruction. Therefore, we do not feel that the quality of MR scans was diagnostically inferior to our current standard.

Another limitation is variability of the MRI technique, as we included MRI scans performed on both 1.5 and 3 T scanners, without and with an endorectal coil, and 13 outside institution scans were also included (three of which were reinterpreted by the AR group). This is a

Table 4 – Diagnostic accuracy, sensitivity, specificity, positive predictive value, and negative predictive value for the two dedicated prostate MRI radiologists (PR group) and the 33 abdominal imaging radiologists (AR group)

Performance measure	PR group	AR group	p value
PPV	64.9	39.3	<0.0001
NPV	91.8	76.6	<0.0001
Sensitivity	85.9	70.7	0.0027
Specificity	77.3	46.8	<0.0001
Diagnostic accuracy	80.1	54.6	<0.0001

NPV = negative predictive value; PPV = positive predictive value.

pragmatic reflection of the current clinical practice [20]. In addition, our use of the Likert scale allowed us to include previously treated patients, as the Likert scale allows assessment of the whole gland and can be used after focal therapy, whereas PI-RADS is lesion based and was developed for treatment-naïve patients. Studies have shown that, for experienced radiologists, the Likert scale has accuracy comparable with or greater than that of PI-RADS [23,30].

5. Conclusions

The sensitivity, specificity, PPV, and NPV of prostate MRI were higher for the PR group than for the AR group. Using only targeted biopsy results, csPCa would have been missed in 9.8% of patients without review by a prostate radiologist. It is important to be aware that the accuracy of prostate MRI may depend on the radiologist involved as well as on the institutional patient mix and biopsy preferences. Histopathological feedback likely plays a key role in the development of expertise.

Author contributions: Hyunseon C. Kang and Haesun Choi had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Kang, Choi.

Acquisition of data: Jo, Saeed Bamashmos, Ahmed, Ward, Choi, Kang.

Analysis and interpretation of data: Sun, Kang, Choi, Ward, Jo.

Drafting of the manuscript: Kang.

Critical revision of the manuscript for important intellectual content: Choi, Ward, Sun, Kang.

Statistical analysis: Sun.

Obtaining funding: None.

Administrative, technical, or material support: Jo, Ahmed, Saeed Bamashmos.

Supervision: Kang.

Other: None.

Financial disclosures: Hyunseon C. Kang certifies that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (eg, employment/affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received, or pending), are the following: None.

Funding/Support and role of the sponsor: This research was supported in part by the NIH/NCI under award number P30CA016672 and used the CCSG Biostatistics Resource Group at MD Anderson Cancer Center.

Acknowledgments: Editorial support was provided by Bryan Tutt in Editing Services, Research Medical Library, The University of Texas MD Anderson Cancer Center.

CRedit authorship contribution statement

Hyunseon C. Kang: Conceptualization, Data curation, Investigation, Methodology, Project administration, Supervision, Writing - original draft. **Nahyun Jo:** Data curation, Resources, Investigation. **Anas Saeed Bamashmos:** Data curation, Resources, Investigation. **Mona Ahmed:** Data curation, Resources, Investigation. **Jia Sun:** Formal analysis,

Data curation, Writing - review & editing. **John F. Ward:** Resources, Investigation, Writing - review & editing. **Haesun Choi:** Conceptualization, Methodology, Investigation, Writing - review & editing.

References

- [1] Loeb S, Bjurlin MA, Nicholson J, et al. Overdiagnosis and overtreatment of prostate cancer. *Eur Urol* 2014;65:1046–55.
- [2] Johnson LM, Turkbey B, Figg WD, Choyke PL. Multiparametric MRI in prostate cancer management. *Nat Rev Clin Oncol* 2014;11:346–53.
- [3] Overduin CG, Futterer JJ, Barentsz JO. MRI-guided biopsy for prostate cancer detection: a systematic review of current clinical results. *Curr Urol Rep* 2013;14:209–13.
- [4] Stabile A, Giganti F, Rosenkrantz AB, et al. Multiparametric MRI for prostate cancer diagnosis: current status and future directions. *Nat Rev Urol* 2020;17:41–61.
- [5] Kasivisvanathan V, Rannikko AS, Borghi M, et al. MRI-targeted or standard biopsy for prostate-cancer diagnosis. *N Engl J Med* 2018;378:1767–77.
- [6] Drost FH, Osses DF, Nieboer D, et al. Prostate MRI, with or without MRI-targeted biopsy, and systematic biopsy for detecting prostate cancer. *Cochrane Database Syst Rev* 2019;4:Cd012663.
- [7] Latchamsetty KC, Borden Jr LS, Porter CR, et al. Experience improves staging accuracy of endorectal magnetic resonance imaging in prostate cancer: what is the learning curve? *Can J Urol* 2007;14:3429–34.
- [8] Akin O, Riedl CC, Ishill NM, Moskowitz CS, Zhang J, Hricak H. Interactive dedicated training curriculum improves accuracy in the interpretation of MR imaging of prostate cancer. *Eur Radiol* 2010;20:995–1002.
- [9] Panebianco V, Giganti F, Kitzing YX, et al. An update of pitfalls in prostate mpMRI: a practical approach through the lens of PI-RADS v. 2 guidelines. *Insights Imaging* 2018;9:87–101.
- [10] Kirkham AP, Haslam P, Keanie JY, et al. Prostate MRI: who, when, and how? Report from a UK consensus meeting. *Clin Radiol* 2013;68:1016–23.
- [11] Muller BG, Shih JH, Sankineni S, et al. Prostate cancer: interobserver agreement and accuracy with the revised prostate imaging reporting and data system at multiparametric MR imaging. *Radiology* 2015;277:741–50.
- [12] 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. *Eur Radiol* 2020;30:5404–16.
- [13] Pickersgill NA, Vetter JM, Raval NS, et al. The accuracy of prostate magnetic resonance imaging interpretation: impact of the individual radiologist and clinical factors. *Urology* 2019;127:68–73.
- [14] Epstein JI, Zelefsky MJ, Sjoberg DD, et al. A contemporary prostate cancer grading system: a validated alternative to the Gleason score. *Eur Urol* 2016;69:428–35.
- [15] Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986;73:13–22.
- [16] Brizmohun Appayya M, Adshead J, Ahmed HU, et al. National implementation of multi-parametric magnetic resonance imaging for prostate cancer detection - recommendations from a UK consensus meeting. *BJU Int* 2018;122:13–25.
- [17] Alberts AR, Roobol MJ, Drost FH, et al. Risk-stratification based on magnetic resonance imaging and prostate-specific antigen density may reduce unnecessary follow-up biopsy procedures in men on active surveillance for low-risk prostate cancer. *BJU Int* 2017;120:511–9.

- [18] 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.
- [19] Shin T, Smyth TB, Ukimura O, et al. Diagnostic accuracy of a five-point Likert scoring system for magnetic resonance imaging (MRI) evaluated according to results of MRI/ultrasonography image-fusion targeted biopsy of the prostate. *BJU Int* 2018;121:77–83.
- [20] Gaziev G, Wadhwa K, Barrett T, et al. 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* 2016;117:80–6.
- [21] Westphalen AC, McCulloch CE, Anaokar JM, et al. Variability of the positive predictive value of PI-RADS for prostate MRI across 26 centers: experience of the Society of Abdominal Radiology Prostate Cancer Disease-focused Panel. *Radiology* 2020;296:76–84.
- [22] Sonn GA, Fan RE, Ghanouni P, et al. Prostate magnetic resonance imaging interpretation varies substantially across radiologists. *Eur Urol Focus* 2019;5:592–9.
- [23] Rosenkrantz AB, Lim RP, Haghghi M, Somberg MB, Babb JS, Taneja SS. Comparison of interreader reproducibility of the prostate imaging reporting and data system and Likert scales for evaluation of multiparametric prostate MRI. *AJR Am J Roentgenol* 2013;201:W612–8.
- [24] Mussi TC, Yamauchi FI, Tridente CF, et al. Interobserver agreement of PI-RADS v. 2 lexicon among radiologists with different levels of experience. *J Magn Reson Imaging* 2020;51:593–602.
- [25] Di Campli E, Delli Pizzi A, Seccia B, et al. Diagnostic accuracy of biparametric vs multiparametric MRI in clinically significant prostate cancer: comparison between readers with different experience. *Eur J Radiol* 2018;101:17–23.
- [26] Garcia-Reyes K, Passoni NM, Palmeri ML, et al. Detection of prostate cancer with multiparametric MRI (mpMRI): effect of dedicated reader education on accuracy and confidence of index and anterior cancer diagnosis. *Abdom Imaging* 2015;40:134–42.
- [27] Doshi AM, Huang C, Melamud K, et al. Utility of an automated radiology-pathology feedback tool. *J Am Coll Radiol* 2019;16:1211–7.
- [28] Molins E, Macia F, Ferrer F, Maristany MT, Castells X. Association between radiologists' experience and accuracy in interpreting screening mammograms. *BMC Health Serv Res* 2008;8:91.
- [29] Wang S, Li XT, Zhang XY, et al. MRI evaluation of extramural vascular invasion by inexperienced radiologists. *Br J Radiol* 2019;92:20181055.
- [30] Khoo CC, Eldred-Evans D, Peters M, et al. Likert vs. PI-RADS v2: a comparison of two radiological scoring systems for detection of clinically significant prostate cancer. *BJU Int* 2020;125:49–55.