Abdominal Radiology

Published online: 31 August 2018

Abdom Radiol (2019) 44:705–712
https://doi.org/10.1007/s00261-018-1751-5

Assessment of PI-RADS v2 categories ≥ 3 for diagnosis of clinically significant prostate cancer

Nayana U. Patel,[1] Kimberly E. Lind,[2,5] Kavita Garg,[1] David Crawford,[3] Priya N. Werahera,[4] and Sajal S. Pokharel[1,6]

[1]Department of Radiology, University of Colorado SOM, 12401 East 17th Avenue, Mail Stop L954, Aurora, CO 80045, USA
[2]Department of Health Systems, Management and Policy, Colorado School of Public Health, Aurora, CO 80045, USA
[3]Department of Urologic Oncology, University of Colorado, Denver, Mail Stop # F 710, PO Box # 6510, Aurora, CO 80045, USA
[4]Department of Pathology and Department of Bioengineering, University of Colorado Anschutz Medical Campus, Mail Stop 8104, P. O. Box 6511, Aurora, CO 80045, USA
[5]*Present address:* Australian Institute of Health Innovation, Macquarie University, Sydney, NSW 2109, Australia
[6]*Present address:* Diversified Radiology of Colorado, P.C., 1746 Cole Blvd. Suite 150, Lakewood, CO 80401, USA

Abstract

Purpose: To assess the diagnostic accuracy of PI-RADS v2 categories ≥ 3 to detect clinically significant prostate cancer (csPCa) against histopathology of Transperineal Mapping Biopsy (TPMB).

Materials and methods: IRB-approved retrospective cohort study included 47 men who had 3.0 T multi-parametric MRI (mpMRI) and TPMB of prostate. Two radiologists independently evaluated T2, DWI, ADC map, and DCE images using PI-RADS v2 categories. A third radiologist served as tie-breaker. PI-RADS v2 score (PS) ≥ 3 lesions were correlated with 3D model of TPMB (3DTPMB) results based on prostate sectors. Two groups of csPCa status were separately analyzed for accuracy measures at lesion and person levels: Group 1 with GS (Gleason Score) ≥ 7 and group 2 with tumor volume ≥ 0.5 cc. Inter-rater reliability for PS and MR lexicon was calculated.

Results: Forty-seven patients with 3DTPMB had at least one lesion with PS ≥ 3 on mpMRI. PS of 5 had high PPV and high specificity of 100% at the lesion and person levels. Sensitivity of a PS ≥ 3 was 68.27% for group 1 and was 48.39% for group 2. Specificity was 93.56% for group 1 and was 95.53% for group 2. At the person level, sensitivity of PS ≥ 3 was 81.25% for group 1 and was 82.35% for group 2. Specificity was 32.26% for group 1 and was 53.85% for group 2.

Conclusion: PI-RADS v2 category of 5 had high PPV and specificity; however, combined PS ≥ 3 had mixed performance in detection of csPCa.

Key words: Prostate MRI—Multi-parametric MRI—Prostate cancer—Clinically significant prostate cancer—PI-RADS v2

Prostate cancer (PCa) is the most common non-skin cancer and the third most common cause of cancer-related deaths in American men [1, 2].

Over the past several years, magnetic resonance imaging (MRI) technology has advanced, allowing high spatial resolution imaging of the prostate gland with multi-parametric MRI (mpMRI). More recently, due to variability in imaging, interpretation, and reporting of prostate MRI, attempts have been made to standardize prostate MRI scans and reports. One attempt at standardization was the Prostate Imaging Reporting and Data System (PI-RADS v1) in 2012 followed by the second version (PI-RADS v2) in 2015 to assess the risk of prostate cancer in lesions detected by mpMRI. PI-RADS v2 simplified rules for reporting, modified imaging sequences, and defined clinically significant cancer [3, 4]. Per PI-RADS v2, clinically significant cancer is defined as Gleason score ≥ 7 on pathology/histology, and/or volume ≥ 0.5 cc, and/or extra prostatic extension (EPE). A recent AUA (American Urological Association) policy statement strongly encourages radiologists to use the PI-RADS v2 to report prostate mpMRI findings [5]. The

Correspondence to: Nayana U. Patel; *email:* Nayana.patel@ucdenver.edu

results from a survey of specialty societies found strong preference for PI-RADS v2 for prostate MRI reporting by radiologists and urologists [6]. The ultimate success of the PI-RADS v2 system rests heavily on its ability to accurately detect csPCa and show reproducibility in PI-RADS suspicion scores among different readers.

Several studies have been performed to validate PI-RADS v2 since 2015. These studies have used mpMRI technique as a part of the standard MRI work-up, but no standard mode of biopsy for pathology correlation has been employed. Many of the studies used targeted biopsies including MRI guidance, MRI-US fusion [7–10], or random transrectal ultrasound guidance [11] or combination of targeted and random biopsy [12, 13] as the reference standard.

This study was designed to assess the diagnostic accuracy of PI-RADS v2 categories ≥ 3 to detect clinically significant cancer using 3DTPMB as the reference standard. TPMB technique has been validated and has shown excellent correlation with whole-mount prostatectomy specimens [14–16], and hence is a reasonable alternative to whole-mount prostatectomy as a reference standard. Unlike targeted and non-targeted random TRUS biopsy methods, TPMB technique is a less biased reference standard as it samples the entire prostate gland at 5-mm interval. Besides providing Gleason grade and Gleason score information from histopathology, TPMB also allowed tumor volume calculation to better estimate clinically significant PCa as defined by PI-RADS v2. Prostate cancer-positive biopsy locations from TPMB were compiled onto 3D model of each patient's prostate gland (3DTPMB), providing better mapping of PCa tumor within the gland to minimize location error during MRI correlation.

Materials and methods

Study design

A HIPAA-compliant, IRB-approved retrospective cohort study was conducted at an academic hospital. Informed consent was waived by the institutional review board.

Eligibility criteria

Patients who underwent prostate MRI and subsequently had 3DTPMB of the prostate between September 2010 and June 2013 were included in the study sample. None of these patients had prior intervention for prostate cancer. TPMB was performed after the mpMRI (multiparametric MRI) examination to avoid misinterpretation caused by hemorrhage. Histopathology from TPMB was used as a reference standard.

Imaging protocol and interpretation

The mpMRI study was performed on a 3 Tesla scanner (GE Signa HDxt—Fairfield, CT, USA) utilizing an endorectal coil (Medrad Prostate eCoil—Warrendale, PA, USA) and 8 channel pelvis phased array surface coil. MRI protocol included large field-of-view images of the pelvis and the following high-resolution images of the prostate: small field-of-view (FOV) tri-planar high resolution T2-weighted, diffusion weighted with ADC maps (*b* value of 0, 600 and 1000), and dynamic post-contrast. mpMRIs were reviewed independently by two sub-specialty trained radiologists (reader 1 with 3 years and reader 2 with 10 years of experience), blinded to MRI reports and clinical-pathologic outcomes. Both radiologists reviewed each participant's mpMRI independently for specific MRI features and for overall score for each lesion in PZ (peripheral zone) and TZ (Transitional zone) using PI-RADS v2 lexicon. Each MRI lesion was evaluated on small FOV T2, diffusion-weighted images (ADC map and DWI), and dynamic contrast-enhanced (DCE) images. Homogenous low signal intensity on T2, circumscribed margins, invasive behavior, restricted diffusion on ADC/DWI (low signal intensity on ADC and corresponding high signal intensity on DWI with *b* value of 1000), and positive enhancement on DCE were recorded for PZ lesions. Lenticular shape, non-circumscribed margins, heterogeneous T2 low signal intensity, T2 low signal rim/encapsulation, invasive behavior, restricted diffusion on ADC/DWI (low signal intensity on ADC and corresponding high signal intensity on DWI with *b* value of 1000), and positive enhancement on DCE were recorded for TZ lesions. In any instances of discordance between the two radiologists, a third radiologist (15 years of experience) served as the tie-breaker. MRI lesions with overall PI-RADS v2 suspicion score of ≥ 3 were recorded in the secured database.

Histopathology

All of these patients had systematic transperineal mapping biopsy (TPMB) of prostate following mpMRI. TPMB is routinely offered to the patients at our institution who wish to proceed with prostate biopsy procedures. TPMB consisted of systematic transperineal template-guided prostate biopsy using real-time transrectal ultrasound guidance to sample the entire prostate gland from apex to base at 5-mm increment with between 55 and 108 cores per patient depending on the size of the prostate gland. TPMB was performed by urologist using a Civco-type brachytherapy stepping unit to collect transverse ultrasound images of the prostate in 5-mm interval from apex to base for 3D reconstruction. Prostate biopsy was originally reviewed and reported by the uropathologist. Prostate cancer-positive biopsy foci were then compiled onto 3D models of each patient's prostate gland utilizing proprietary software ProView (Applied Coherent Technology, Herndon, Virginia) yielding a 3D model of prostate gland with PCa tumor foci (3DTPMB) which were detected by the TPMB (Fig. 2).

MRI-pathology co-relation

mpMRI lesions were correlated to the 3DTPMB PCa tumor foci (Figs. 1, 2) based on location by dividing the prostate gland into 12 sectors (apex, mid, base; anterior, posterior; left, right). Biopsy data including location, Gleason score (GS), and a binary indicator for csPCa were entered into the secure database independently and blindly of mpMRI data. Clinically significant prostate cancer was defined as GS of ≥ 7 and/or tumor volume ≥ 0.5 cc [4, 17]. Approximate tumor volume was determined using the cross-sectional area and length of the lesion. Cross-sectional area can be calculated using 5-mm transperineal grid points used for mapping biopsy and length of the core sample was given in the pathology report. In any instances of more than one lesion in a specific region, the lesions were matched using the additional criteria of maximum Gleason score and maximum PI-RADS score (i.e., we assumed that the highest Gleason score would be the highest PI-RADS score).

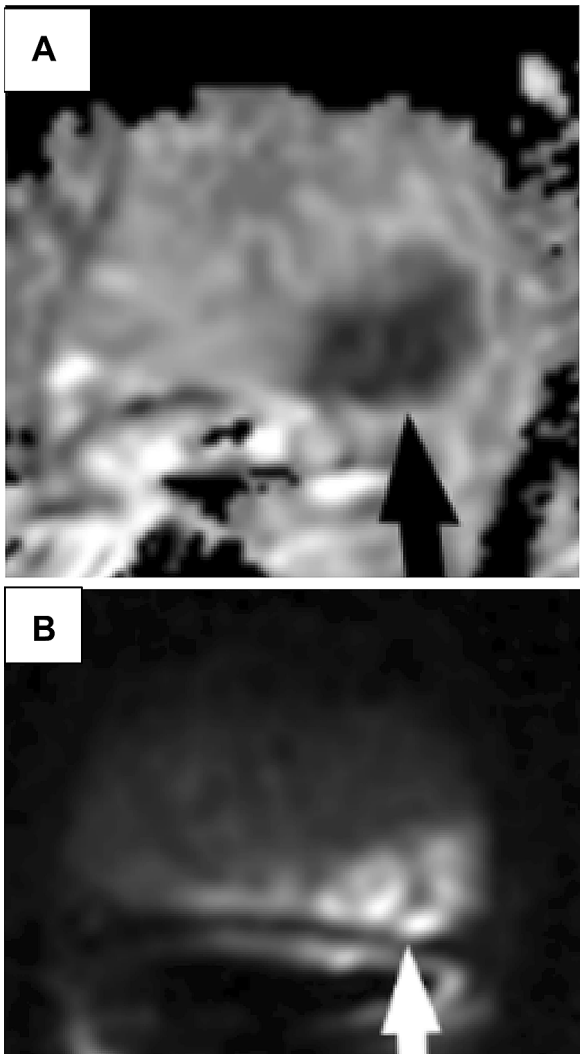

Fig. 1. A 78-year-old man with prostate adenocarcinoma. Axial MRI images demonstrates 2.1-cm lesion (arrow) in the left posterolateral peripheral zone in mid-gland, which is markedly hypointense on ADC map (**A**), markedly hyperintense on DWI with *b* value of 1000 (**B**). The lesion was moderately hypointense on T2-weighted image, and showed early enhancement on dynamic contrast-enhanced sequences. This was a PI-RADSv2 category 5 lesion that corresponded to multiple adjacent 3D transperineal mapping biopsy cores that were positive for adenocarcinoma (Fig. 2).

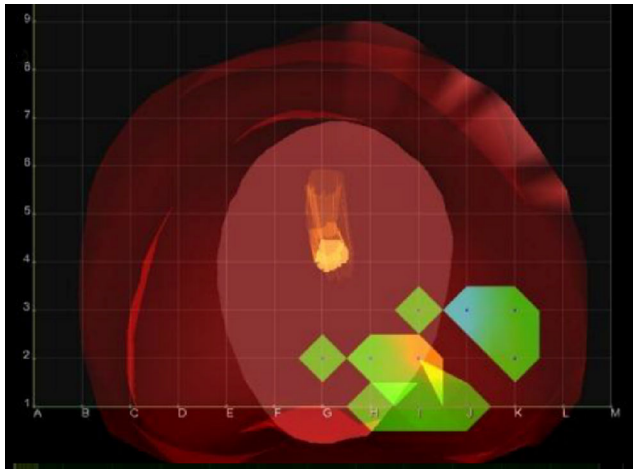

Fig. 2. 3D model of the cancer-positive results from transperineal mapping biopsy of prostate gland (3DTPMB) in this patient demonstrates the colored diamonds correspond to biopsy cores that were positive for cancer on pathology. Green diamonds represent Gleason 4 + 3 = 7, orange diamond was Gleason 3 + 4 = 7, and blue diamond was Gleason 6. The location of the tumor foci in the left posterior-lateral mid-gland corresponds to the lesion seen on mpMRI in Fig. 1.

Statistical methods

Descriptive statistics were calculated for all patient characteristics of interest including mean, median, standard deviation (SD), and interquartile range (IQR) for continuous variables, and frequencies and percents for categorical variables. Prostate volume was calculated from the three dimension prostate size in cm using ellipsoid formula [length × width × height × (pi/6)].

To evaluate inter-rater reliability for lesion characteristics assessed by the two radiologists, we calculated intra-class correlation for variables that were either continuous or ordinal, and Kappa for binary variables.

To assess the diagnostic accuracy of the PI-RADS v2 scoring system, we compared positive MRI with lesion's overall PS ≥ 3 to pathology results from 3DTPMB using generalized estimating equations (GEE) regression for lesion-level data. This regression approach accounts for correlation between lesions occurring in the same patient, generating corrected standard errors and confidence intervals. We evaluated two sets of results for csPCa:

group1 for PCa with GS (Gleason Score) ≥ 7 and group 2 for PCa with tumor volume ≥ 0.5 cc. We estimated sensitivity and positive predictive value (PPV). We estimated the binary proportion and corresponding 95% confidence intervals for each accuracy measure.

Any sector in prostate that was not observed on MRI and did not have a positive biopsy result on 3DTPMB was recorded as negative MRI and having a cancer negative result. This sector-level dataset was then used to calculate specificity and negative predictive value (NPV). Since there were multiple records per person in this dataset, we once again used GEE regression to calculate the proportion and corresponding 95% CIs as we did with the lesion-level data.

Finally, we generated a person-level dataset with indicators for group 1, group 2, and any lesion scored as PI-RADS 3, 4, or 5. Accuracy measures were calculated as simple proportions, and we calculated exact binomial confidence intervals for each proportion.

We used a type I error rate of 0.05 for all statistical tests. All analyses and data processing were completed using SAS 9.4 (SAS Institute, Cary, NC).

Results

A total of 47 patients with 3DTPMB and mpMRI were included in the final sample. The mean patient age was 63.2 years (Table 1). These 47 patients had 50 lesions with PS ≥ 3 observed on mpMRI (after lesion presence was tie-broken by the third radiologist); 23 lesions were rated as PI-RADS 3, 21 lesions were rated as PI-RADS 4, and 6 lesions were rated as PI-RADS 5 (Table 2). 56 clinically significant cancer lesions were found by biopsy: 25 were GS ≥ 7 and 31 were GS 6 with tumor volume > 0.5 cc. 29 out of 56 clinically significant lesions were missed on MRI: 10 lesions had GS ≥ 7 cancer and 19 lesions had GS 6 cancer with volume > 0.5 cc. Accuracy measures for lesion-level, sector-level, and person-level data are presented in Table 3 for the group 1 and group 2 csPCa. Six MRI lesions were assigned score of 5 based on PI-RADS v2 overall assessment categories; all PI-RADS 5 lesions had GS ≥ 7 cancer. PI-RADS v2 score of 5 had high PPV and specificity (both 100%) at the lesion level and person level.

Table 1. Sample characteristics (n = 47 patients)

Variable	Frequency (%)
Positive PSA test (> 4.0 ng/mL)	38 (80.85%)
Positive digital rectal exam	7 (14.89%)
Number of lesions on mpMRI	
0	14.00 (28.57)
1	22.00 (44.90)
2	10.00 (20.41)
3	2.00 (4.08)
4	1.00 (2.04)

	Mean (SD)	Median	IQR
Age (years)	63.23 (6.84)	63	60–69
PSA	6.57 (5.12)	5.73	3.8–7.8
Prostate volume (cm^3)	35.83 (15.97)	32.78	27.17–41.30

Table 2. PI-RADS v2 scores by lesion outcomes (n = 122 biopsy samples)

	Group 1		Group 2	
	Positive	Negative	Positive	Negative
PI-RADS 3	5	18	9	14
PI-RADS 4	4	17	12	9
PI-RADS 5	6	0	6	0
Negative MRI	10	62	29	43
Total	25	97	56	66

Group 1: prostate cancer with Gleason score ≥ 7
Group 2: prostate cancer with tumor volume ≥ 0.5 cc

Measures of inter-rater reliability (IRR) are presented in Table 4. PI-RADS v2 scores had moderate reliability (ICC = 0.58). For PZ lesions (Table 5), reliability was substantial for restricted diffusion on ADC (κ = 0.71), positive enhancement on DCE (κ = 0.77), and reliability was moderate for circumscribed lesions (κ = 0.54). For TZ lesions (Table 6), reliability was substantial for homogeneous lesions (κ = 0.75), lack of T2 dark rim (κ = 0.65), and restricted diffusion on ADC (κ = 0.71); reliability was moderate for non-circumscribed lesions (κ = 0.45) and positive enhancement on DCE (κ = 0.59).

Other characteristics were either in the fair, slight, or no reliability range, including assessment for signal abnormality on DWI. Some Kappa values were negative due to never having agreement on the presence of a given lesion characteristic; for DWI and lenticular shape, there was no agreement. For invasive behavior in TZ lesions, Kappa could not be calculated since this characteristic was only called as present by one radiologist for one lesion (i.e., the distribution of the data did not allow for Kappa to be calculated).

Discussion

Using 3DTPMB as a reference standard, PI-RADS v2 score of 5 had perfect PPV and specificity (100%) at the lesion level and person level. All patients with an mpMRI lesion of PI-RADS v2 score of 5 had csPCa and patients who had no cancerous results did not have any MRI detectable lesion with PI-RADS score of 5. Sensitivity of a PI-RADS v2 score ≥ 3 was 68.27% to detect group 1 csPCa and was only 48.39% to detect group 2 csPCa. Specificity of PI-RADs v2 assessment was high: over 90% for both group 1 and group 2 csPCa.

At the person level, sensitivity of a PI-RADS v2 score ≥ 3 was over 80% for both group 1 and group 2 csPCa results. However, specificity of PI-RADS v2

Table 3. PI-RADS v.2 accuracy measures (lesion level, sector level, and person level)

	Group 1		Group 2	
	Estimate (%)	95% CI	Estimate (%)	95% CI
Lesion level				
Sensitivity				
PI-RADS 3	19.58	9.24–36.80%	16.88	9.26–28.77%
PI-RADS 4	16.97	6.35–38.12%	23.21	13.41–37.10%
PI-RADS 5	22.92	11.42–40.68%	11.15	5.60–20.97%
PI-RADS 3, 4, or 5	68.27	44.62–85.17%	48.39	35.55–61.46%
PPV				
PI-RADS 3	22.76	11.63–39.74%	40.06	23.67–59.03%
PI-RADS 4	20.00	8.66–39.74%	56.67	39.79–72.12%
PI-RADS 5	100		100	
PI-RADS 3, 4, or 5	29.61	18.26–44.19%	53.46	39.9–66.52%
Sector level				
Specificity				
PI-RADS 3	96.68	94.99–97.81%	97.26	95.62–98.29%
PI-RADS 4	96.89	95.31–97.94%	98.27	96.87–99.04%
PI-RADS 5	100		100	
PI-RADS 3, 4, or 5	93.56	91.47–95.16%	95.53	93.52–96.93%
NPV				
PI-RADS 3	96.37	94.11–97.79%	91.55	89.1–93.25%
PI-RADS 4	96.22	93.7–97.76%	92.21	89.11–94.23%
PI-RADS 5	96.64	94.40–98.00%	91.33	88.69–93.11%
PI-RADS 3, 4, or 5	98.14	95.85–99.18%	94.55	91.99–96.18%
Person level				
Sensitivity				
PI-RADS 3	31.25	11.02–58.66%	38.24	22.17–56.44%
PI-RADS 4	43.75	19.75–70.12%	52.94	35.13–70.22%
PI-RADS 5	37.50	15.2–64.57%	17.65	6.76–34.53%
PI-RADS 3, 4, or 5	81.25	54.35–95.95%	82.35	65.47–93.24%
Specificity				
PI-RADS 3	54.84	36.03–72.68%	53.85	25.13–80.78%
PI-RADS 4	64.52	45.37–80.77%	100	
PI-RADS 5	100		100	
PI-RADS 3, 4, or 5	32.26	16.68–51.37%	53.85	25.13–80.78%
PPV				
PI-RADS 3	26.32	9.15–51.20%	68.42	43.45–87.42%
PI-RADS 4	38.89	17.3–64.25%	100	
PI-RADS 5	100		100	
PI-RADS 3, 4, or 5	38.24	22.17–56.44%	82.35	65.47–93.24%
NPV				
PI-RADS 3	60.71	40.58–78.50%	25.00	10.69–44.87%
PI-RADS 4	68.97	49.17–84.72%	44.83	26.45–64.31%
PI-RADS 5	75.61	59.70–87.64%	31.71	18.08–48.09%
PI-RADS 3, 4, or 5	76.92	46.19–94.96%	53.85	25.13–80.78%

PPV, positive predictive value; NPV, negative predictive value
Group 1: prostate cancer with Gleason score ≥ 7
Group 2: prostate cancer with tumor volume ≥ 0.5 cc

Table 4. Inter-rater reliability

Variable	ICC
ADC	0.71
Lesion size	0.72
PI-RADS v2 score ≥ 3	0.58

Reader 1 (3 years of experience)	Reader 2 (10 years of experience)			
	PI-RADS 3	PI-RADS 4	PI-RADS 5	Lesion not called
PI-RADS 3	9	5	0	2
PI-RADS 4	5	11	1	1
PI-RADS 5	0	1	5	0
Lesion not called	4	6	0	72

Table 5. Inter-rater reliability for peripheral zone lesions

Peripheral zone (PZ) lesion characteristics	Reader 1	Reader 2			
		Present	Absent	Kappa	95% CI
Homogeneous low signal on T2	Present	14	4	0.34	0.06–0.63
	Absent	9	12		
Circumscribed margins	Present	14	7	0.54	0.29–0.80
	Absent	2	16		
Invasive behavior	Present	0	3	− 0.04	− 0.10–0.02
	Absent	1	35		
Bright DWI	Present	5	11	0.30	0.03–0.56
	Absent	1	22		
Dark ADC	Present	28	3	0.71	0.45–0.97
	Absent	1	7		
Positive on dynamic post-contrast image	Present	18	2	0.77	0.52–1.00
	Absent	1	8		

Table 6. Inter-rater reliability for transitional zone lesions

Transitional zone (TZ) lesion characteristics	Reader 1	Reader 2			
		Present	Absent	Kappa	95% CI
Homogeneous low signal on T2	Present	6	1	0.75	0.49–1.00
	Absent	2	30		
Lack of T2 dark rim around the nodule	Present	5	2	0.65	0.34–0.97
	Absent	2	30		
Invasive behavior*	Present	0	1		
	Absent	0	38		
Bright DWI	Present	0	3	− 0.04	− 0.10–0.02
	Absent	1	35		
Dark ADC	Present	7	1	0.71	0.45–0.97
	Absent	3	28		
Non-circumscribed margins	Present	2	3	0.45	0.00–0.90
	Absent	1	33		
Lenticular shape	Present	0	3	− 0.04	− 0.10–0.02
	Absent	1	35		
Positive on dynamic post-contrast image	Present	5	1	0.59	0.27–0.91
	Absent	4	29		

*Kappa could not be calculated due to distribution of cells

assessment was low, 32.26% for group 1 and 53.85% for group 2, highlighting the high number of false positives that come as a consequence of maximizing sensitivity. As the approaches to prostate cancer screening and diagnosis continue to evolve, clinicians should carefully consider these accuracy limitations of PI-RADS v2 categories in order to limit the number of unnecessary biopsies and procedures.

Our study results were consistent with respect to some accuracy results from prior studies, but also inconsistent with other accuracy results. The differences may be attributable to use of different reference standards and different analyses for larger or smaller areas of the prostate. Grey and colleagues [18] used transperineal sector biopsy (TPSB) as a reference standard instead of 3DTPMB, performed mpMRIs on 1.5T scanner (without endo-coil), and reported with a Likert-type scale as opposed to 3T MRI and reporting using PI-RADS v2 categories in our study. Furthermore, csPCa in that study was defined as any Gleason pattern of 4 or ≥ 6 mm core length (approximate tumor volume of ≥ 0.5 mL) of Gleason pattern 3. That study generated PPV and NPV similar to our results, but sensitivity and specificity differ notably from our estimates. Arumainayagam and colleagues [19] used TPMB as a reference standard and MRI performed on 1.5 T with suspicion score on the Likert scale, again as opposed to 3T mpMRI and PI-RADS v2 scoring in our study. In that study, the prostate gland was divided in 4 sectors versus 12 sectors in our study. For csPCa defined as GS ≥ 7 and/or cancer core length of ≥ 6 mm, sensitivity was 64–81%, specificity was 68–80%, NPV was 91–95%, and PPV was 35–45%. These reported values are somewhat closer to our findings than the Grey study, which we attribute to the biopsy technique being more similar. A meta-analysis by Zhang et al. [20] analyzed 13 studies (2049 total patients) for overall diagnostic accuracy of PI-RADS v2 in diagnosing prostate cancer. This analysis estimated a

pooled sensitivity of 0.85 (0.78–0.91), pooled specificity of 0.71 (0.60–0.80), PPV from 0.54 to 0.97, and NPV from 0.26 to 0.92. The results of this meta-analysis showed significant heterogeneity among the studies. Inter-reader agreement was reported good to excellent for studies in which 2 or more readers provided separate results of MRI interpretations. A study by Seo et al. [21] included patients with biopsy proven GS ≥ 6 PCa who underwent MRI and radical prostatectomy. csPCa was surgically defined as GS ≥ 7 or a tumor volume of ≥ 0.5 cm^3, or tumor category ≥ T3. For the experienced readers, the proportions of csPCa were significantly higher in a group with PS ≥ 4 than in a group with a PS < 4 (< 0.001). For inexperienced reader, PI-RADS v2 scores were predictive of GS ≥ 7 and category ≥ T3 , but not of tumor volume ≥ 0.5 cm^3 or presence of csPCa.

Vargas et al. [22] used a whole-mount pathology as a reference standard and found correct classification of PZ cancer (94%) and TZ cancer (95%) ≥ 0.5 mL on pathology, However, majority of GS ≥ 7(4 + 3) PCa with volumes < 0.5 mL on pathology were not detectable on T2WI, DW-MRI, and DCE-MRI, and PI-RADS v2 was only able to identify 9/37 (24%) of PCa, which is similar to our study.

We found that inter-rater reliability was varied across the lesion characteristics used in PI-RADS v2 lexicon in our study. There was substantial to moderate IRR for ADC and T2 features, supporting the emphasis of these parameters in the scoring of mpMRI lesions by PI-RADS v2. However, there was no agreement on DWI signal abnormality. A possible factor for this lack of agreement is the subjectivity in assessing signal abnormality on DWI images. It is our experience that a factor as simple as the display contrast level/width setting can dramatically alter the perceived hyperintensity of lesions. Clinicians should be cautious interpreting the presence or absence of features in PI-RADS v2 that have low inter-rater reliability.

A study by Rosenkrantz [23] of "interobserver reproducibility among 6 experience prostate radiologists of the PI-RADS v2 lexicon" showed that the reproducibility for features related to DWI was substantial for focal shape (κ = 0.619) and moderate for marked high-*b*-value or ADC map signal abnormality (κ = 0.535–0.562) for PZ lesions. Reproducibility was fair to moderate for features related to DCE (κ = 0.266–0.439) for PZ lesions. The reproducibility for features related to lesion texture and margins at T2-weighted imaging ranged from 0.136 (moderately hypointense) to 0.529 (encapsulation) for TZ lesions. Reproducibility related to DWI was lower in the TZ (κ = 0.343–0.465) than in the PZ.

Our study had a number of strengths including the technical parameters for mpMRI similar to PI-RADS v2 guidelines, use of biopsy-verified cancer results by sampling of entire prostate gland, use of 3D prostate model (3DTPMB) of cancer foci detected by TPMB to correlate with mpMRI (use of a reference standard that minimized location error), use of multiple readers (sub-specialty trained radiologists) for the mpMRIs, and the blinded entry of biopsy data. Further, our regression approach for lesion-level outcomes corrected for correlation within patients who had more than one lesion observed.

One technical parameter in this study significantly differed from PI-RADS v2 guidelines in that the highest *b* value for DWI was 1000 s/$mm^{2.}$ and not ≥ 1400 s/$mm^{2.}$as prescribed in PI-RADS v2. This use of a relatively lower *b* value may have contributed to the low specificity that was observed at the person level.

Our greatest limitation is generalizability; this was a single-site small sample retrospective study and our patients and providers may differ from those at other sites, although we do not suspect that this is a substantial bias and expect that our patients and providers are similar to those at other academic hospitals in the US. Finally, in our study we did not evaluate the MRI lesions with PI-RADS v2 categories of ≤ 2.

In conclusion, PI-RADS v2 category of 5 had high PPV and specificity at the lesion level and person level; however, combined PI-RADS v2 categories of ≥ 3 had mixed performance in detection of clinically significant prostate cancer. Clinicians using PI-RADS v2 should be attentive to its limitations with respect to limited accuracy for scores 3 and 4, and for its low inter-rater reliability for certain features. The strengths and weaknesses of PI-RADS v2 identified in this study should be carefully considered as approaches to prostate cancer screening and diagnosis evolve.

Acknowledgments The authors gratefully acknowledge Dr. Nelson N. Stone, MD (Professor of Urology and Radiation Oncology, The Icahn School of Medicine at Mount Sinai New York), Alexandria Jensen, Dr. Francisco La Rosa, Emma Murugaverl, Paul Arangua, and Francisco G. La Rosa (all with the University of Colorado School of Medicine) for their insight and assistance with this project.

Compliance with ethical standards

Funding Dr. Sajal Pokharel had received a pilot Grant from the department of radiology at University of Colorado. No other grant funding.

Conflict of interest Dr. Nayana Patel, Dr. Kimberly Lind, Dr. Kavita Garg, Dr. David Crawford, and Dr. Priya Werahera declare that they have no conflict of interest.

References

1. Key Statistics in Prostate Cancer. https://www.cancer.org/cancer/prostate-cancer/about/key-statistics.html. Accessed 15 Jan 2017
2. Cancer of the Prostate - Cancer Stat Facts [Internet]. https://seer.cancer.gov/statfacts/html/prost.html. Accessed 13 Sept 2017
3. Weinreb JC, Barentsz JO, Choyke PL, et al. (2016) PI-RADS prostate imaging - reporting and data system: 2015, version 2. Eur Urol. 69(1):16–40
4. American College of Radiology. MR Prostate Imaging Reporting and Data System version 2.0. http://www.acr.org/Quality-Safety/Resources/PIRADS/. Accessed 15 Jan 2017

5. Fulgham PF, Rukstalis DB, Turkbey IB, et al. (2017) AUA policy statement on the use of multiparametric magnetic resonance imaging in the diagnosis, staging and management of prostate cancer. J Urol. 198(4):832–838
6. Spilseth B, Ghai S, Patel NU, et al. (2017) A comparison of radiologists' and urologists' opinions regarding prostate MRI reporting: results from a Survey of Specialty Societies. American Journal of Roentgenology. https://doi.org/10.2214/AJR.17.18241
7. Polanec S, Helbich TH, Bickel H, et al. (2016) Head-to-head comparison of PI-RADS v2 and PI-RADS v1. Eur J Radiol. 85(6):1125–1131
8. Cash H, Maxeiner A, Stephan C, et al. (2016) The detection of significant prostate cancer is correlated with the Prostate Imaging Reporting and Data System (PI-RADS) in MRI/transrectal ultrasound fusion biopsy. World J Urol. 34(4):525–532
9. Muller BG, Shih JH, Sankineni S, et al. (2015) Prostate cancer: interobserver agreement and accuracy with the revised Prostate Imaging Reporting and Data System at multiparametric MR imaging. Radiology. 277(3):741–750
10. Purysko AS, Bittencourt LK, Bullen JA, et al. (2017) Accuracy and interobserver agreement for prostate imaging Reporting and Data System, version 2, for the characterization of lesions identified on multiparametric MRI of the prostate. AJR Am J Roentgenol. 209(2):339–349
11. Baldisserotto M, Neto EJD, Carvalhal G, et al. (2016) Validation of PI-RADS v. 2 for prostate cancer diagnosis with MRI at 3T using an external phased-array coil. J Magn Reson Imaging. 44(5):1354–1359
12. Cash H, Günzel K, Maxeiner A, et al. (2016) Prostate cancer detection on transrectal ultrasonography-guided random biopsy despite negative real-time magnetic resonance imaging/ultrasonography fusion-guided targeted biopsy: reasons for targeted biopsy failure. BJU Int. 118(1):35–43
13. Hansen NL, Kesch C, Barrett T, et al. (2017) Multicentre evaluation of targeted and systematic biopsies using magnetic resonance and ultrasound image-fusion guided transperineal prostate biopsy in patients with a previous negative biopsy. BJU Int. 120(5):631–638
14. Crawford ED, Rove KO, Barqawi AB, et al. (2013) Clinical-pathologic correlation between transperineal mapping biopsies of the prostate and three-dimensional reconstruction of prostatectomy specimens. Prostate. 73(7):778–787
15. Dimmen M, Vlatkovic L, Hole K-H, et al. (2012) Transperineal prostate biopsy detects significant cancer in patients with elevated prostate-specific antigen (PSA) levels and previous negative transrectal biopsies. BJU Int. 110(2 Pt 2):E69–E75
16. Barqawi AB, et al. (2011) The role of 3-dimensional mapping biopsy in decision making for treatment of apparent early stage prostate cancer. J Urol 186(1):80–85
17. Ploussard G, Epstein JI, Montironi R (2011) The contemporary concept of significant versus insignificant prostate cancer. Eur Urol. 60(2):291–303. https://doi.org/10.1016/j.eururo.2011.05.006
18. Grey ADR, Chana MS, Popert R, et al. (2015) Diagnostic accuracy of magnetic resonance imaging (MRI) prostate imaging reporting and data system (PI-RADS) scoring in a transperineal prostate biopsy setting. BJU Int. 115(5):728–735
19. Arumainayagam N, Ahmed HU, Moore CM, et al. (2013) Multiparametric MR imaging for detection of clinically significant prostate cancer: a validation Cohort Study with transperineal template prostate mapping as the reference standard. Radiology 268(3):761–769
20. Zhang L, Tang M, Chen S, et al. (2017) A meta-analysis of use of Prostate Imaging Reporting and Data System Version 2 (PI-RADS V2) with multiparametric MR imaging for the detection of prostate cancer. Eur Radiol. 27:5204
21. Seo JW, Shin S-J, Taik OhY, et al. (2017) PI-RADS version 2: detection of clinically significant cancer in patients with biopsy gleason score 6 prostate cancer. AJR Am J Roentgenol. 209(1):W1–W9
22. 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. https://doi.org/10.1007/s00330-015-4015-6
23. Rosenkrantz AB, Ginocchio LA, Cornfeld D, et al. (2016) Interobserver reproducibility of the PI-RADS version 2 lexicon: A Multicenter Study of six experienced prostate radiologists. Radiology. 280(3):793–804