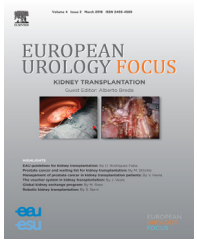
available at www.sciencedirect.com
journal homepage: www.europeanurology.com/eufocus



Prostate Cancer

Accuracy and Variability of Prostate Multiparametric Magnetic Resonance Imaging Interpretation Using the Prostate Imaging Reporting and Data System: A Blinded Comparison of Radiologists

Nicholas A. Pickersgill[a], *Joel M. Vetter*[a], *Gerald L. Andriole*[a], *Anup S. Shetty*[b], *Kathryn J. Fowler*[b], *Aaron J. Mintz*[b], *Cary L. Siegel*[b], *Eric H. Kim*[a,*]

[a] *Division of Urology, Washington University School of Medicine, St. Louis, MO, USA;* [b] *Department of Radiology, Washington University School of Medicine, St. Louis, MO, USA*

Article info

Article history:
Accepted October 8, 2018

Associate Editor:
Derya Tilki

Keywords:
Prostate cancer
Radiologist variability
Accuracy
Magnetic resonance imaging
Biopsy
Diagnosis

Abstract

Background: Multiparametric (mp) magnetic resonance imaging (MRI) has become an important tool for the detection of clinically significant prostate cancer. However, diagnostic accuracy is affected by variability between radiologists.
Objective: To determine the accuracy and variability in prostate mpMRI interpretation among radiologists, both individually and in teams, in a blinded fashion.
Design, setting, and participants: A study cohort ($n = 32$) was created from our prospective registry of patients who received prostate mpMRI with subsequent biopsy. The cohort was then independently reviewed by four radiologists of varying levels of experience, who assigned a Prostate Imaging Reporting and Data System (PI-RADS) classification, blinded to all clinical information. Consensus interpretation by teams of two radiologists was evaluated after a 12-wk wash-out period. Interpretive accuracy was calculated with various cutoffs for PI-RADS classification and Gleason score. Variability among individual radiologists and teams was calculated using the Fleiss kappa and intraclass correlation coefficient (ICC).
Results and limitations: Using PI-RADS 3+/Gleason 7+ ($p < 0.01$) and PI-RADS 4+/Gleason 6+ ($p = 0.02$) as cutoffs, significant differences in accuracy among the four radiologists were noted. At no cutoff for PI-RADS classification or Gleason score did a team read achieve higher accuracy than the most accurate radiologist. The kappa and ICC ranged from 0.22 to 0.29 for the individuals and from 0.16 to 0.21 for the teams (poor agreement). A larger sample size may be needed to adequately power differences in accuracy among individual radiologists.
Conclusions: At various cutoffs for PI-RADS classification and Gleason score, we find significant differences in individual radiologist accuracy, as well as a poor agreement among individual radiologists. Consensus interpretations—as teams of two radiologists—did not improve accuracy or reduce variability.
Patient summary: This study investigated radiologist variability and differences in accuracy using multiparametric magnetic resonance imaging for the diagnosis of prostate cancer. Despite attempts to standardize interpretation within the field, we found substantial variability and significant differences in accuracy among individual radiologists.

* Corresponding author. Washington University School of Medicine, 4960 Children's Place, Campus Box 8242, St. Louis, MO 63110, USA. Tel.: +1-314-362-8510; Fax: +1-314-454-5244.
E-mail address: ehkim@wustl.edu (E.H. Kim).

https://doi.org/10.1016/j.euf.2018.10.008
2405-4569/

Please cite this article in press as: Pickersgill NA, et al. Accuracy and Variability of Prostate Multiparametric Magnetic Resonance Imaging Interpretation Using the Prostate Imaging Reporting and Data System: A Blinded Comparison of Radiologists. Eur Urol Focus (2018), https://doi.org/10.1016/j.euf.2018.10.008

1. Introduction

Multiparametric (mp) magnetic resonance imaging (MRI) of the prostate, with subsequent MRI–targeted biopsy (MRITB), has emerged as an important tool for the detection of clinically significant prostate cancer (PCa) [1–5]. As the use of prostate mpMRI has become more widespread, the Prostate Imaging Reporting and Data System (PI-RADS) was created for the standardization of study interpretations. Since then, PI-RADS has repeatedly been validated and updated to its second iteration (version 2) [6–9].

Although some PCa foci remain undetectable by prostate mpMRI [10–12], one important determinant of PCa detection is interpretive accuracy of the reader. Single and multicenter studies have found variability in interpretation of the same mpMRI among different readers when using PI-RADS [13–15]. Given the known variability between readers, differences in accuracy for predicting clinically significant PCa are likely to exist. Unfortunately, prior studies of radiologist variability and accuracy are limited by important elements of study design, including image interpretation methodology (eg, including circled mpMRI lesions and screen captures of mpMRI lesions) and use of radical prostatectomy (RP) as the pathology reference, which eliminates false positive interpretations [14–19].

As prostate mpMRI becomes increasingly important to PCa diagnosis and management, defining the variability and differences in interpretive accuracy among radiologists becomes important. We aim to evaluate variability as well as compare accuracy of individual radiologists using PI-RADS to predict clinically significant PCa, in a blinded fashion. Additionally, we aim to evaluate consensus reads—teams of two radiologists—as a method to improve accuracy and reduce variability.

2. Patients and methods

2.1. Study cohort creation

With Institutional Review Board approval, we developed a study cohort of prostate mpMRI from our prospective database of patients who underwent prostate mpMRI at our institution followed by a biopsy—standard template transrectal ultrasound (TRUS) for PI-RADS classification 1 or 2 lesions and software-fusion MRITB with standard template TRUS for PI-RADS classification 3–5 lesions—from September 2014 to December 2015. The Uronav-platform (Invivo Corporation, Gainesville, FL, USA) was used to perform software-fusion MRITB. We excluded patients whose mpMRI findings were clinically interpreted by one of the study radiologists. Cases were selected in a stratified random (stratified by original clinical PI-RADS classification), nonconsecutive fashion, and evaluated for adequate image quality by a third party prior to inclusion in the study cohort. The cohort size ($n = 32$) was created based on a power analysis described below and designed to be proportionally representative of the distribution of clinical mpMRI interpretations at our institution (Table 1). Patient clinical information and biopsy results were recorded. A Gleason score of >7 on biopsy pathology was considered clinically significant PCa.

2.2. Study design

Four radiologists were enrolled in our study and are included in authorship (A.S., K.F., A.M., and C.S.). The level of experience varied widely from a chief resident in radiology (A.M., zero independent prostate mpMRI cases) to an assistant professors in radiology (A.S., 216 cases) to an associate professor in radiology (K.F., 548 cases) to a full professor in radiology (C.S., 185 cases). In our primary analysis, the four radiologists independently reviewed the entire study cohort of prostate mpMRI, blinded to the original mpMRI interpretation and all patient clinical information. The images were loaded in a picture archiving and communication system (PACS) workstation by an independent operator, in a randomized order, for each radiologist. The study PI-RADS interpretations were then submitted to a blinded third party who combined the observed data for statistical analysis. For our secondary analysis, after a 12-wk wash-out period, the radiologists were divided into teams of two, and the study cohort of prostate mpMRI was reviewed in a fashion similar to that described above by each team. In all cases, the clinically obtained biopsy cores served as the pathology reference. As the clinical and research mpMRI reads differed in some cases, the highest Gleason scores from targeted or systematic biopsy cores were considered for all cases.

Table 1 – Distribution of PI-RADS classification on a per-patient basis.

PI-RADS classification	Proportion of the institutional database (%)	Study cohort ($n = 32$)
1 or 2	9	3
3	19	6
4	44	14
5	28	9

PI-RADS = Prostate Imaging Reporting and Data System.

2.3. MRI technique

Specific description of our institutional prostate mpMRI technique has been published previously [11,20]. Briefly, all patients underwent 3-Tesla mpMRI using a pelvic phased-array coil on Siemens Trio and Skyra platforms (Siemens Healthcare, Erlangen, Germany). High-resolution turbo spin echo T2-weighted images consisted of 3-mm slice thickness, and small field of view imaging (160 mm) with a matrix of 512 × 512 for axial imaging and 640 × 640 for coronal imaging. The small field of view diffusion-weighted imaging sequence consisted of 3-mm slice thickness, with b values of 50 and 800, and a calculated b value of 1400 with a matrix of 128 × 128. Apparent diffusion coefficient maps were generated by the scanner. Dynamic contrast enhancement sequences consisted of three-dimensional gradient recall echo T1-weighted imagines with a temporal resolution of 6–8 s, imaged over 2 min, with a matrix of 256 × 256 and 3-mm slice thickness.

2.4. MRI Interpretation

All prostate mpMRI findings were interpreted using the DynaCAD software platform (Invivo Corporation) on a PACS workstation. No identifying information or regions of interest were provided. Original clinical interpretation was recorded, but, as stated above, it was not available to the radiologists during the study. Each radiologist provided PI-RADS version 2 interpretation and location for up to three suspicious lesions per prostate mpMR image.

2.5. Statistical analysis

The size of the study cohort was determined by a power analysis using the observed maximum difference in interpretive accuracy among individual radiologists at our institution over the last 2 yr. At our institution, prostate mpMRI findings are read as part of the clinical workflow of the

Please cite this article in press as: Pickersgill NA, et al. Accuracy and Variability of Prostate Multiparametric Magnetic Resonance Imaging Interpretation Using the Prostate Imaging Reporting and Data System: A Blinded Comparison of Radiologists. Eur Urol Focus (2018), https://doi.org/10.1016/j.euf.2018.10.008

abdominal imaging section of diagnostic radiology. Nine fellowship trained attending radiologists interpret prostate mpMRI reads, supervising fellows and residents. In our retrospective review of our institutional experience, we found that the maximum difference in interpretive accuracy—defined as PI-RADS classification 4 or 5 corresponding to Gleason >7 PCa and PI-RADS classification 1–3 corresponding to benign or Gleason 6 PCa—was approximately 45% between the most and the least accurate radiologists. Using this observed difference, alpha 0.05, and beta 0.80, we arrived at a sample size estimate of 32 cases.

For our primary analysis, the interpretive accuracy for clinically significant PCa was calculated for individual radiologists and compared. The analysis was performed on a per-patient basis, and as stated above, the pathologic reference standard was the highest Gleason score on systematic or targeted (for PI-RADS 3–5 cases) biopsy. Using various cutoffs for PI-RADS classification (>3 or >4) as test positive and biopsy Gleason score (>6 or >7) as condition positive, accuracy was defined as the sum of true positives and true negatives divided by the entire cohort (n = 32).

For our secondary analysis, interpretive accuracy of the teams was compared with each other as well as with individual radiologists. Additionally, the variability among individual radiologists and the teams of radiologists were calculated. Fleiss kappa was used for variability when a binary comparison was performed (PI-RADS 1–3 vs 4–5), and the intraclass correlation coefficient (ICC) was used for variability when the comparison was of multiple groupings (PI-RADS 1–2 vs 3 vs 4–5, or PI-RADS 1 vs 2 vs 3 vs 4 vs 5).

The interpretive accuracy of individual and team radiologists was compared using chi-square analysis. Multivariate logistic regression was performed to determine the effect of clinical variables on accuracy and variability, by pooling the observations (n = 192). All p values <0.05 were considered statistically significant. All analyses were completed with R version 3.2.2.

3. Results

The mean age of patients in our study cohort was 64.8 ± 6.4 yr, mean prostate-specific antigen level was 6.4 ± 4.4 ng/ml, and digital rectal examination was normal in all patients. Approximately half of our study cohort underwent prior biopsy (15/32 = 47%), while the remaining patients were biopsy naïve. Distribution of their original clinical mpMRI interpretation is included in Table 1.

Table 2 summarizes individual PI-RADS interpretations by each radiologist. Original clinical PI-RADS classification and biopsy results were not available to the radiologists during the study. When PI-RADS 3+/Gleason 6+ and PI-RADS 4+/Gleason 7+ were used as cutoffs, no significant differences in accuracy were noted (p = 0.31 and p = 0.41, respectively). However, when PI-RADS 3+/Gleason 7+ and PI-RADS 4+/Gleason 6+ were used as cutoffs, significant differences in accuracy were noted ($p < 0.01$ and p = 0.02, respectively). No significant differences in team accuracy were noted at any cutoffs. At no cutoff for PI-RADS or Gleason score did a team read achieve higher accuracy than the most accurate radiologist.

Multivariate logistic regression analysis found the presence of clinically significant PCa on biopsy to be a significant predictor of improved accuracy (odds ratio [OR] 3.07, 95% confidence interval [CI] 1.11–9.64), controlling for radiologist, Prostate Cancer Prevention Trial risk estimate, prostate volume, presence of clinically significant PCa, or location of mpMRI lesion.

Table 3 summarizes the variability of the individual and team reads, with various PI-RADS classification groupings. Fleiss kappa values were 0.29 and 0.16 for individual and team interpretations, respectively, when creating a binary division between PI-RADS classifications (1–3 vs 4–5). ICC values were, respectively, 0.27 and 0.19 for individual and team interpretations, when creating three PI-RADS classification groups (1–2 vs 3 vs 4–5). ICC values were 0.22 and 0.21 for individual and team interpretations, respectively, when using each PI-RADS classification alone (1 vs 2 vs 3 vs 4 vs 5).

Multivariate logistic regression analysis found that increasing prostate volume was a significant predictor of increasing dispersion (OR 1.6, 95% CI 1.07–2.50), controlling for radiologist, Prostate Cancer Prevention Trial risk estimate, prostate volume, presence of clinically significant PCa, or location of mpMRI lesion.

4. Discussion

In our blinded study, we found significant differences in interpretative accuracy based on various cutoffs for PI-RADS classification and Gleason score. Using PI-RADS classification 3–5, interpretative accuracy for clinically significant PCa (Gleason >7) was significantly different among individual radiologists (63%, 72%, 75%, and 47%; $p < 0.01$). Additionally, using PI-RADS classification 4–5, interpretive accuracy for any PCa (Gleason >6) was significantly different among individual radiologists (63%, 72%, 56%, and 59%; p = 0.02). Team reads did not substantially improve

Table 2 – Summary of radiologist and team accuracy based on various cutoffs for PI-RADS classification and Gleason score.

Cutoffs	Radiologist #1	Radiologist #2	Radiologist #3	Radiologist #4	p value
PI-RADS 3+, Gleason 6+	59.4%	75.0%	71.9%	56.3%	0.31
PI-RADS 3+, Gleason 7+	62.5%	71.9%	75.0%	46.9%	<0.01
PI-RADS 4+, Gleason 6+	62.5%	71.9%	56.3%	59.4%	0.02
PI-RADS 4+, Gleason 7+	59.4%	75.0%	65.6%	56.3%	0.41
	Team #1	Team #2			
PI-RADS 3+, Gleason 6+	68.8%	68.8%			1.0
PI-RADS 3+, Gleason 7+	59.4%	65.6%			0.61
PI-RADS 4+, Gleason 6+	71.9%	71.9%			1.0
PI-RADS 4+, Gleason 7+	62.5%	75.0%			0.28

PI-RADS = Prostate Imaging Reporting and Data System.

Please cite this article in press as: Pickersgill NA, et al. Accuracy and Variability of Prostate Multiparametric Magnetic Resonance Imaging Interpretation Using the Prostate Imaging Reporting and Data System: A Blinded Comparison of Radiologists. Eur Urol Focus (2018), https://doi.org/10.1016/j.euf.2018.10.008

Table 3 – Measures of variability among the individual and team prostate mpMRI interpretations using various PI-RADS groupings.

	Fleiss kappa	ICC	ICC
PI-RADS groupings	(1, 2, 3) vs (4, 5)	(1, 2) vs (3) vs (4, 5)	1 vs 2 vs 3 vs 4 vs 5
Individual reads	0.291	0.272	0.216
Team reads	0.162	0.194	0.207

ICC = intraclass correlation coefficient; mpMRI = multiparametric magnetic resonance imaging; PI-RADS = Prostate Imaging Reporting and Data System.

accuracy compared with individual radiologists, and interestingly, resulted in worse inter-reader agreement with various PI-RADS classification groupings.

Most previous studies of radiologists using prostate mpMRI for PCa diagnosis demonstrate similar interpretive accuracy to our study, but less variability (eg, greater agreement). These differences are attributable to study design. Previous studies, their results, and important elements of their study design are summarized in Table 4.

RP specimens were used as the reference standard by Garcia-Reyes et al [18] and Greer et al [19]. The use of a cohort restricted to RP enriches the study for patients with PCa and biases the cohort toward those with clinically significant PCa. One would then expect improvement in interpretive accuracy and reduction in variability with this study design. In fact, our multivariate analysis found that the presence of clinically significant PCa improves the interpretive accuracy of mpMRI. Expectedly, interpretive accuracy is higher in the study by Garcia-Reyes et al [18] than ours (74% vs 66%), and variability between radiologists is low (kappa = 0.72) in the study by Greer et al [19].

A recent retrospective study by Sonn et al [15] also found considerable variability in the diagnostic accuracy of prostate MRI when MRITB histopathology was used as the reference standard. Although MRITB specimens were used as the reference standard by Rosenkrantz et al [14], Muller et al [16], and Schimmoller et al [17], mpMR images were evaluated in a manner distinct from our study. Rather than allowing study radiologists full access to all mpMRI sequences on a PACS workstation (how mpMRI interpretation is routinely performed clinically), the study authors provided screen captures of suspicious mpMRI lesions or mpMRI sequences on a PACS workstation with the lesion circled. Naturally, highlighting the lesion in question biases the result to reduce variability and improve accuracy. As expected, interpretive accuracy is higher in the studies by Schimmoller et al [17] and Muller et al [16] than in ours (76%, 78%, and 66%, respectively). Likewise, variability among radiologists was much lower than in our study (kappa = 0.65 and 0.46, respectively, vs 0.29). Although Rosenkrantz et al [14] found lower variability compared with our study (kappa = 0.56 vs 0.29), they found that interpretive accuracy was lower than that in our study (56% vs 66%). As theirs is the only multicenter study, institutional differences in prostate mpMRI interpretation may have affected interpretive accuracy. Additionally, the reference standard of MRITB was used for 63 of the 120 mpMRI included in their study cohort, with systematic biopsy used as the reference standard for the other patients. In those cases, assessment of interpretive accuracy of the radiologist may be inadequate as the mpMRI lesion may not have been adequately sampled.

Compared with the previously published studies on accuracy and variability among radiologists, our study design has some advantages that may provide greater generalizability. First, in the application of prostate mpMRI in a diagnostic setting, the reference histopathology that the radiologists' interpretation is compared against is the biopsy, as not all patients who receive prostate mpMRI will go on to receive RP (eg, patients may not harbor PCa and patients may decide to receive radiation treatment rather than RP). Second, clinical prostate mpMRI interpretation is performed with the images available on a PACS workstation (not captured images) and without previous annotations. We believe these elements to be more representative of diagnostic radiology interpretation (eg, biopsy as the pathology reference and access to a PACS workstation with no prior annotation); thus, we believe that our study may better represent the current state of prostate mpMRI interpretation.

Several findings were notable from our multivariate analyses. First, we found no differences in interpretive accuracy among the study radiologists, despite significant differences in experience (ranging from a chief resident to a

Table 4 – Summary of studies examining radiologists' accuracy and variability.

Study	Patients	Radiologists	Image reviewed	Pathology reference	Scoring system	Accuracy (%)	Kappa
Rosenkrantz et al [14]	120	6	Screen capture	MRI/US fusion biopsy	PI-RADS v 2	56	0.56
Sonn et al [15]	409	9	PACS access	MRI fusion biopsy	PI-RADS v 1, 2	73	
Muller et al [16]	94	5	Screen capture	MRI/US fusion biopsy	PI-RADS v 2	78	0.46
Schimmoller et al [17]	67	3	Circled lesion	In-bore MRI biopsy	PI-RADS v 1	76	0.65
Garcia-Reyes et al [18]	31	5	PACS access	Prostatectomy	Gleason 6 vs. 7+	55–74	
Greer et al [19]	34	5	PACS access	Prostatectomy	PI-RADS v 2		0.72
Current study	32	4	PACS access	MRI/US fusion biopsy	PI-RADS v 2	66	0.29

MRI = magnetic resonance imaging; PACS = picture archiving and communication system; PI-RADS = Prostate Imaging Reporting and Data System; US = ultrasound.

Please cite this article in press as: Pickersgill NA, et al. Accuracy and Variability of Prostate Multiparametric Magnetic Resonance Imaging Interpretation Using the Prostate Imaging Reporting and Data System: A Blinded Comparison of Radiologists. Eur Urol Focus (2018), https://doi.org/10.1016/j.euf.2018.10.008

full professor with over 20 yr of clinical experience). In earlier studies of prostate mpMRI, radiologist experience appeared to influence accuracy. Ruprecht et al [21] found that the sensitivity and specificity for correct staging with prostate mpMRI using RP as the reference standard were, respectively, 78% and 93% for an experienced reader and 33% and 71% for a less experienced reader. In a more contemporary study discussed above, Greer et al [19] report similar positive predictive values and sensitivities for radiologists of differing levels of experience using PI-RADS. As the use of prostate mpMRI is relatively novel, particularly outside of a research setting, the additional years in practice may not have provided a benefit in interpreting these images accurately. Additionally, the standardization provided by PI-RADS may attenuate the advantages of experience, as Schieda et al [22] found that more experienced radiologists performed significantly better when using non-standardized reporting in predicting extraprostatic extension at prostatectomy, but performed similarly to less experienced radiologists when using PI-RADS.

Second, we found that the presence of clinically significant PCa was a predictor of interpretive accuracy. As many of the early studies demonstrating the diagnostic ability of prostate mpMRI were based on RP specimens, which necessarily harbor PCa and are more likely to harbor clinically significant PCa, the improved accuracy with increasing disease burden is intuitive [23–26].

Third, we found that the dispersion of PI-RADS interpretation was reduced when the prostate volume was reduced. As increasing prostate volume is typically associated with enlargement of the transition zone due to benign prostatic hypertrophy, which is a known confounder of prostate mpMRI interpretation for PCa [27], this result is expected. A recent study by Starobinets et al [28] demonstrated that 5-alpha reductase inhibitor use (known to reduce prostate volume) was associated with improved mpMRI discrimination and reduced variability.

Our study is not without limitations. Using MRITB as the reference histopathology has intrinsic limitations due to potential undersampling or technical miss during biopsy; however, as discussed above, the use of RP introduces a selection bias that likely improves accuracy and reduces variability. Although we found significant differences between radiologists with certain PI-RADS and Gleason cutoffs, we did not demonstrate a significant difference between individual radiologists at all cutoffs in this study. Some of these observed differences were not statistically significant but may be considered clinically significant. Thus, a larger sample size may be needed to adequately power differences in accuracy among individual radiologists; however, the time-intensive nature of individual and independent reviews of prostate mpMRI did not allow for a substantially larger study cohort of prostate mpMRI. Although a greater number of radiologists would provide additional data, we anticipate that the interpretive accuracy of any additional individual radiologists would likely fall within the range of this study, which would not significantly impact overall accuracy or inter-reader agreement.

5. Conclusions

In a blinded study of radiologists of varying levels of experience, we find significant differences in interpretive accuracy based on various PI-RADS classifications and Gleason score cutoffs. Furthermore, using a clinically representative study design (eg, radiologist access to PACS workstation and biopsy as the pathology reference), we find that the inter-reader agreement was poor, with variability much higher than previously reported. Consensus interpretation by a team of two radiologists did not significantly improve accuracy or reduce variability.

Author contributions: Eric H. Kim 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: Kim, Shetty, Fowler, Mintz, Siegel, Andriole.
Acquisition of data: Kim, Shetty, Fowler, Mintz, Siegel, Vetter.
Analysis and interpretation of data: Pickersgill, Vetter, Y.M., Kim.
Drafting of the manuscript: Pickersgill, Y.M., Andriole, Kim.
Critical revision of the manuscript for important intellectual content: Pickersgill, Y.M., Kim, Andriole.
Statistical analysis: Vetter, Y.M., Kim, Pickersgill.
Obtaining funding: None.
Administrative, technical, or material support: Vetter, Kim, Andriole.
Supervision: Kim, Andriole.
Other: None.

Financial disclosures: Eric H. Kim 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:*** None.

References

[1] Moore CM, Robertson NL, Arsanious N, et al. Image-guided prostate biopsy using magnetic resonance imaging-derived targets: a systematic review. Eur Urol 2013;63:125–40.
[2] Siddiqui MM, Rais-Bahrami S, Truong H, et al. Magnetic resonance imaging/ultrasound-fusion biopsy significantly upgrades prostate cancer versus systematic 12-core transrectal ultrasound biopsy. Eur Urol 2013;64:713–9.
[3] Rastinehad AR, Turkbey B, Salami SS, et al. Improving detection of clinically significant prostate cancer: magnetic resonance imaging/transrectal ultrasound fusion guided prostate biopsy. J Urol 2014;191:1749–54.
[4] Tompson JE, Moses D, Shnier R, et al. Multiparametric magnetic resonance imaging guided diagnostic biopsy detects significant prostate cancer and could reduce unnecessary biopsies and over detection: a prospective study. J Urol 2014;192:67–74.
[5] Futterer JJ, Briganti A, De Visschere P, et al. Can clinically significant prostate cancer be detected with multiparametric magnetic resonance imaging? A systematic review of the literature. Eur Urol 2015;68:1045–53.
[6] Roethke MC, Kuru TH, Schultze S, et al. Evaluation of the ESUR PI-RADS scoring system for multiparametric MRI of the prostate with targeted MR/TRUS fusion-guided biopsy at 3.0 Tesla. Eur Radiol 2014;24:344–52.

Please cite this article in press as: Pickersgill NA, et al. Accuracy and Variability of Prostate Multiparametric Magnetic Resonance Imaging Interpretation Using the Prostate Imaging Reporting and Data System: A Blinded Comparison of Radiologists. Eur Urol Focus (2018), https://doi.org/10.1016/j.euf.2018.10.008

[7] Hamoen EHJ, de Rooij M, Witjes JA, Barentsz JO, Rovers MM. Use of the Prostate Imaging Reporting and Data System (PI-RADS) for prostate cancer detection with multiparametric magnetic resonance imaging: a diagnostic meta-analysis. Eur Urol 2015;67:1112–21.
[8] Weinreb JC, Barentsz JO, Choyke PL, et al. Prostate Imaging—Reporting and Data System: 2015, version 2. Eur Urol 2016;69:16–40.
[9] Vargas HA, Hotker AM, Goldman DA, et al. 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 2016;26:1606–12.
[10] Nassiri N, Natarajan S, Margolis DJ, Marks LS. Targeted prostate biopsy: lessons learned midst the evolution of a disruptive technology. Urology 2015;86:432–8.
[11] Wang RS, Kim EH, Vetter JM, et al. Determination of the role of negative magnetic resonance imaging of the prostate in clinical practice: is biopsy still necessary? Urology 2017;102:190–7.
[12] Le JD, Tan N, Shkolyar E, et al. Multifocality and prostate cancer detection by multiparametric magnetic resonance imaging: correlation with whole-mount histopathology. Eur Urol 2015;67:569–76.
[13] Rosenkrantz AB, Lim RP, Haghighi 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. Am J Roentgenol 2013;201:W612–8.
[14] Rosenkrantz AB, Ginocchio LA, Cornfeld D, et al. Inter-observer reproducibility of the PI-RADS version 2 lexicon: a multi-center study of six experienced prostate radiologists. Radiology 2016;280:793–804.
[15] Sonn GA, Fan RE, Ghanouni P, et al. Prostate magnetic resonance imaging interpretation varies substantially among radiologists. Eur Urol Focus 2017;17:30266–73.
[16] Muller BG, Shih JH, Sankineni S, et al. Prostate cancer: intraobserver agreement and accuracy with the revised prostate imaging reporting and data system at multiparametric MR imaging. Radiology 2015;277:741–50.
[17] Schimmoller L, Quentin M, Arsov C, et al. Inter-reader agreement of the ESUR score for prostate MRI using in-bore MRI-guided biopsies as the reference standard. Eur Radiol 2013;23:3185–90.
[18] 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;40:124–42.
[19] Greer MD, Brown AM, Shih JH, et al. Accuracy and agreement of PIRADSv2 for prostate cancer mpMRI: a multireader study. J Magn Reson Imaging 2017;45:579–85.
[20] Kim EH, Weaver JK, Shetty AS, Vetter JM, Andriole GL, Strope SA. Magnetic resonance imaging provides added value to the prostate cancer prevention trial risk calculator for patients with estimated risk of high grade prostate cancer less than or equal to 10%. Urology 2017;102:183–9.
[21] Ruprecht O, Weisser P, Bodelle B, Ackermann H, Vogl TJ. MRI of the prostate: intraobserver agreement compared with histopathologic outcome after radical prostatectomy. Eur J of Radiol 2012;81:456–60.
[22] Schieda N, Quon JS, Lim C, et al. Evaluation of the European Society of Urogenital Radiology (ESUR) PI-RADS scoring system for assessment of extra-prostatic extension in prostatic carcinoma. Eur J Radiol 2015;84:1843–8.
[23] Turkbey B, Mani H, Shah V, et al. Multiparametric 3T prostate magnetic resonance imaging to detect cancer: histopathological correlation using prostatectomy specimens processed in customized magnetic resonance imaging based molds. J Urol 2011;186:1818–24.
[24] Bartolozzi C, Menchi I, Lencioni R, et al. Local staging of prostate carcinoma with endorectal coil MRI: correlation with whole-mount radical prostatectomy specimens. Eur Radiol 1996;6:339–45.
[25] Isebaert S, Van den Bergh L, Haustermans K, et al. Multiparametric MRI for prostate cancer localization in correlation to whole-mount histopathology. J Magn Reson Imaging 2013;37:1392–401.
[26] Haider MA, Van Der Kwast TH, Tanguay J, et al. Combined T2-weighted and diffusion-weighted MRI for localization of prostate cancer. Am J Roentgenol 2007;189:323–8.
[27] Rosenkrantz AB, Taneja SS. Radiologist, be aware: ten pitfalls that confound the interpretation of multiparametric prostate MRI. Am J Roentgenol 2014;202:109–20.
[28] Starobinets O, Kurhanewicz J, Noworolski SM. Improved multiparametric MRI discrimination between low-risk prostate cancer and benign tissues in a small cohort of 5alpha-reductase inhibitor treated individuals as compared with an untreated cohort. NMR Biomed 2017;30:1.

Please cite this article in press as: Pickersgill NA, et al. Accuracy and Variability of Prostate Multiparametric Magnetic Resonance Imaging Interpretation Using the Prostate Imaging Reporting and Data System: A Blinded Comparison of Radiologists. Eur Urol Focus (2018), https://doi.org/10.1016/j.euf.2018.10.008