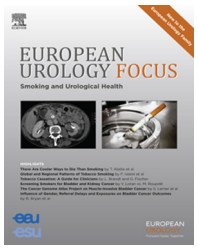
available at www.sciencedirect.com
journal homepage: www.europeanurology.com/eufocus



Prostate Cancer

Prostate Magnetic Resonance Imaging Interpretation Varies Substantially Across Radiologists

Geoffrey A. Sonn [a,b,*], *Richard E. Fan* [a], *Pejman Ghanouni* [b], *Nancy N. Wang* [a], *James D. Brooks* [a], *Andreas M. Loening* [b], *Bruce L. Daniel* [b], *Katherine J. To'o* [b,c], *Alan E. Thong* [a], *John T. Leppert* [a,c,d]

[a] *Department of Urology, Stanford University School of Medicine, Stanford, CA, USA;* [b] *Department of Radiology, Stanford University School of Medicine, Stanford, CA, USA;* [c] *Veterans Affairs, Palo Alto Health Care System, Palo Alto, CA;* [d] *Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA*

Article info

Article history:
Accepted November 28, 2017

Keywords:
Magnetic resonance imaging
Observer variation
Prostatic neoplasms
Radiologists

Abstract

Background: Multiparametric magnetic resonance imaging (mpMRI) interpreted by experts is a powerful tool for diagnosing prostate cancer. However, the generalizability of published results across radiologists of varying expertise has not been verified.
Objective: To assess variability in mpMRI reporting and diagnostic accuracy across radiologists of varying experience in routine clinical care.
Design, setting, and participants: Men who underwent mpMRI and MR-fusion biopsy between 2014–2016. Each MRI scan was read by one of nine radiologists using the Prostate Imaging Reporting and Data System (PIRADS) and was not re-read before biopsy. Biopsy histopathology was the reference standard.
Outcome measurements and statistical analysis: Outcomes were the PIRADS score distribution and diagnostic accuracy across nine radiologists. We evaluated the association between age, prostate-specific antigen, PIRADS score, and radiologist in predicting clinically significant cancer (Gleason $\geq$7) using multivariable logistic regression. We conducted sensitivity analyses for case volume and changes in accuracy over time.
Results and limitations: We analyzed data for 409 subjects with 503 MRI lesions. While the number of lesions (mean 1.2 lesions/patient) did not differ across radiologists, substantial variation existed in PIRADS distribution and cancer yield. The significant cancer detection rate was 3–27% for PIRADS 3 lesions, 23–65% for PIRADS 4, and 40–80% for PIRADS 5 across radiologists. Some 13–60% of men with a PIRADS score of $<$3 on MRI harbored clinically significant cancer. The area under the receiver operating characteristic curve varied from 0.69 to 0.81 for detection of clinically significant cancer. PIRADS score ($p < 0.0001$) and radiologist ($p = 0.042$) were independently associated with cancer in multivariable analysis. Neither individual radiologist volume nor study period impacted the results. MRI scans were not retrospectively re-read by all radiologists, precluding measurement of inter-observer agreement.
Conclusions: We observed considerable variability in PIRADS score assignment and significant cancer yield across radiologists. We advise internal evaluation of mpMRI accuracy before widespread adoption.
Patient summary: We evaluated the interpretation of multiparametric magnetic resonance imaging of the prostate in routine clinical care. Diagnostic accuracy depends on the Prostate Imaging Reporting and Data System score and the radiologist.
© 2017 Published by Elsevier B.V. on behalf of European Association of Urology.

* Corresponding author. Department of Urology, Stanford University School of Medicine, 300 Pasteur Drive, S-287, Stanford, CA 94305, USA. Tel.: +1 650 7935585; Fax: +1 650 4985346.
E-mail address: gsonn@stanford.edu (G.A. Sonn).

https://doi.org/10.1016/j.euf.2017.11.010
2405-4569/© 2017 Published by Elsevier B.V. on behalf of European Association of Urology.

Please cite this article in press as: Sonn GA, et al. Prostate Magnetic Resonance Imaging Interpretation Varies Substantially Across Radiologists. Eur Urol Focus (2017), https://doi.org/10.1016/j.euf.2017.11.010

1. Introduction

Multiparametric MRI (mpMRI) is being rapidly adopted for prostate cancer diagnosis and management. The PROMIS trial found that 27% of biopsy-naïve men with elevated prostate-specific antigen (PSA) and nonsuspicious mpMRI findings could avoid biopsy [1]. If confirmed, these findings could significantly reduce the cost and morbidity of prostate cancer diagnosis by reducing the number of men biopsied.

Before widespread adoption of mpMRI, the generalizability of published results should be rigorously evaluated. To date, most studies have come from expert centers with a small number of experienced radiologists interpreting all mpMRIs [2–4]. It is unknown if these results could be reproduced in practice settings with less experienced radiologists. The initial [5] and revised [6] Prostate Imaging Reporting and Data System (PIRADS) guidelines offered a standard to help radiologists diminish variation in the acquisition, interpretation, and reporting of prostate MRI. A multicenter, multireader study including six expert prostate radiologists found moderate reproducibility for PIRADS version 2, but did find "considerable inter-reader variation" [7]. The extent of this variation in routine clinical practice is unknown.

In our practice, nine radiologists of varying experience read all prostate mpMRI scans as part of routine clinical care. We hypothesized that differences might exist in mpMRI interpretation across radiologists that could impact clinical decision-making.

2. Patients and methods

2.1. Patient population

We identified consecutive study subjects who underwent MRI and MRI-ultrasound (US) fusion targeted biopsy by a single urologist (G.A.S.) from April 2014 to October 2016. We obtained consent for prospective data collection before biopsy under a protocol approved by the institutional review board. We included subjects undergoing initial or repeat biopsy. We excluded those who underwent external MRI, those who did not receive gadolinium contrast, and those who were previously treated for prostate cancer. For men who underwent multiple MRIs and targeted biopsies, we evaluated the most recent biopsy. The analytic cohort included 409 men. No patients were included in a prior publication. Reporting is in accordance with the START guidelines where applicable [8].

2.2. MRI protocol and interpretation

All mpMRI was performed using a 3-T scanner (MR750; GE Healthcare, Waukesha, WI, USA) and an external 32-channel body array coil in prostate mode (peripheral channels not used) without an endorectal coil. The imaging protocol included T2-weighted imaging (T2WI), diffusion-weighted imaging (DWI), and dynamic contrast-enhanced imaging (DCE). Supplementary Table 1 lists specifics of the imaging protocol.

All MRI scans were interpreted using PIRADS by an attending radiologist with expertise in body imaging as part of routine clinical care and were not re-read before biopsy. PIRADS v1 was used until department-wide adoption of PIRADS v2 in 2015. All radiologists were involved from the beginning of the study. Because the study objective was to evaluate prospective interpretations performed during routine clinical care, older studies were not re-evaluated using PIRADS v2. Radiologists could access clinical information including age, PSA, indication for biopsy, and any prior biopsy results. Radiologists varied in years of prostate MRI experience (median 6 yr, range 1–25). Four of the nine radiologists had specific prostate MRI training. No standardized training or performance feedback was conducted before or during the study period.

2.3. Targeted biopsy protocol

A single urologist (G.A.S.) performed all MRI-US fusion targeted prostate biopsies using a robotic biopsy device (Artemis, Eigen, Grass Valley, CA, USA) according to a standard protocol [9]. All biopsies included systematic sampling and targeted cores from any MRI-visible lesions (median 3 cores/target). When no lesions were identified on MRI ($n = 70$, 17.1%), only systematic sampling was performed. The biopsy device selected systematic core locations independent of MRI target locations. MRI target locations were hidden during systematic sampling.

Tissue cores were sent for histopathologic evaluation, and biopsy results were used as the standard for assessment of the presence of cancer. We defined clinically significant cancer as Gleason $\geq 3 + 4$.

2.4. Statistical analysis

The final analytic cohort included 409 men and 503 MRI lesions; all lesions were biopsied. We compared the patient and prostate cancer risk factors among the nine radiologists using the Mann-Whitney test for continuous variables and the χ^2 test for categorical variables. Using univariable and multivariable logistic regression models, we assessed whether each radiologist was associated with identification of prostate cancer or clinically significant prostate cancer. We also examined two-way interactions between radiologist and number of MRI scans read during the study, as well as the study period divided by quartiles. We adjusted for multiple comparisons in the model using p values adjusted for the false discovery rate. We illustrated the percentage of lesions with any cancer or clinically significant cancer by PIRADS score for each radiologist. We compared the performance of each radiologist's PIRADS score for detection of any cancer and clinically significant prostate cancer using the area under the receiver operator characteristic curve (AUC). We conducted all statistical analyses using SAS v9.4 (SAS Institute, Cary, NC, USA) and figures were generated using JMP Pro v13 (SAS Institute).

3. Results

3.1. Study population

Table 1 lists characteristics for the study population ($n = 409$). The median age was 65 yr (interquartile range [IQR] 60–69) and median PSA was 7.9 ng/ml (IQR 5.5–12.1). A radiologist identified at least one lesion in 83% of men. The mean number of lesions per subject on MRI was 1.2 (standard deviation 0.9) and the total number of lesions was 503.

The 409 MRI scans were divided among nine radiologists. Table 2 shows patient and prostate cancer risk factors by radiologist. The number of MRI scans read by each radiologist ranged from 18 to 70. Variation in the number of studies read by each radiologist resulted from differences in the frequency spent on the MRI service. There were no significant differences in patient age, biopsy indication, or mean number of lesions identified across radiologists. PSA differed across radiologists ($p = 0.02$).

Please cite this article in press as: Sonn GA, et al. Prostate Magnetic Resonance Imaging Interpretation Varies Substantially Across Radiologists. Eur Urol Focus (2017), https://doi.org/10.1016/j.euf.2017.11.010

Table 1 – Patient demographics and magnetic resonance imaging results for the full cohort (n = 409).

Characteristic	Result
Age (yr)	
Mean (standard deviation)	64.1 (8.1)
Median (interquartile range)	65 (60–69)
Prior evaluation, *n* (%)	
First biopsy	143 (35)
Prior biopsy	266 (65)
Prior negative biopsy	125 (47)
Prior positive biopsy	141 (53)
Prostate-specific antigen (ng/ml)	
Mean (standard deviation)	9.9 (6.9)
Median (interquartile range)	7.9 (5.5–12.1)
Prostate volume (ml)	
Mean (standard deviation)	64.7 (39.3)
Median (interquartile range)	53.0 (38.7–75.8)
Overall PIRADS score, *n* (%)	
Number of lesions	70 (17.1)
1	1 (0.2)
2	19 (4.6)
3	76 (18.6)
4	146 (35.7)
5	97 (23.7)
Number of lesions identified	
Mean (standard deviation)	1.2 (0.9)
Median (interquartile range)	1 (1–2)
Number of targeted cores per lesion of interest (*n*)	
Mean (standard deviation)	3.2 (1.1)
Median (interquartile range)	3 (2–4)
Mean number of systematic cores per patient, *n* (median)	12.4 (12)

PIRADS = Prostate Imaging Reporting and Data System.

3.2. MRI interpretation

While radiologists did not differ in the mean number of lesions assigned per patient, there were differences in the PIRADS score distribution. Figure 1 shows the variation in PIRADS score assignment by radiologist. The radiologists also differed in cancer yields. Figure 2 shows the proportion of lesions containing clinically significant cancers and all cancers based on PIRADS score.

On average, clinically significant cancer was found in 13% of PIRADS 2, 12% of PIRADS 3, 38% of PIRADS 4, and 63% of PIRADS 5 lesions. When stratified by radiologist, the significant cancer yield ranged from 3% to 27% for PIRADS 3, from 23% to 65% for PIRADS 4, and from 40% to 80% for PIRADS 5 lesions. Figure 3 quantifies the variability in significant cancer yield for PIRADS 3–5 lesions across radiologists. While a higher PIRADS score was associated with a higher likelihood of identifying significant cancer for all nine radiologists, the strength of this correlation (slope of the red lines) varied by radiologist. The mean AUC for all radiologists was 0.73 (range 0.69–0.81) for detection of clinically significant cancer and 0.72 (range 0.67–0.81) for any cancer (Fig. 4).

In our study, if biopsy were restricted to men with a PIRADS score $\geq$3 as proposed in the PROMIS trial, 90 men (22%) with PIRADS $\leq$2 would have avoided biopsy (range 12–30% across radiologists). Of these 90 men, 22 (24%) had clinically significant prostate cancer on biopsy. This proportion of false negatives ranged from 13% to 60% across radiologists.

3.3. Adjusted results

Supplementary Table 2 shows the significance of factors associated with detection of clinically significant prostate cancer in the multivariable logistic regression models. PIRADS score was the dominant contributor to model fit, while radiologist and a history of active surveillance remained independently associated with clinically significant cancer on biopsy in the fully adjusted model. Use of PIRADS v1 versus v2 was not significant (p = 0.053). Pairwise comparisons between radiologists further illustrated variation in the detection of clinically significant prostate cancer (Supplementary Fig. 1).

We found no significant change in performance as the study progressed across study period quartiles (p = 0.11) or with increasing number of cases read (p = 0.66). The interaction between individual radiologist volume and radiologist was not significant (p = 0.79). No significant difference was seen in performance between the five high-volume radiologists and the four low-volume radiologists on multivariable analysis (p = 0.2). In contrast to clinically significant cancer, PIRADS score ($p < 0.0001$) and a history of active surveillance ($p < 0.0001$) were the only factors independently associated with detection of any cancer in fully adjusted models.

Table 2 – Patient and prostate cancer risk factors by radiologist.

Characteristic	Radiologist									p value
	1	2	3	4	5	6	7	8	9	
Total MRI studies (*n*)	65	65	65	65	20	22	19	18	70	
Mean age (yr)	63.5	63.7	64.7	64.6	66.1	64.7	65.4	64.7	62.9	0.98
Median PSA (ng/ml)	7.6	7.5	6.8	6.6	11.9	9.4	10.4	9.6	8.7	0.02
First biopsy (%)	27.7	32.3	39.7	46.2	26.3	23.8	31.6	35.3	39.7	0.62
Prior negative biopsy (%)	33.9	32.3	27.0	20.0	52.6	33.3	36.9	29.4	29.4	
Prior positive biopsy (%)	38.5	35.4	33.3	33.9	21.1	42.9	31.6	35.3	30.9	
Mean lesions of interest (*n*)	1.1	1.2	1.1	1.5	1.5	1.4	1.5	1.2	1.1	0.23

MRI = magnetic resonance imaging; PSA = prostate-specific antigen.

Please cite this article in press as: Sonn GA, et al. Prostate Magnetic Resonance Imaging Interpretation Varies Substantially Across Radiologists. Eur Urol Focus (2017), https://doi.org/10.1016/j.euf.2017.11.010

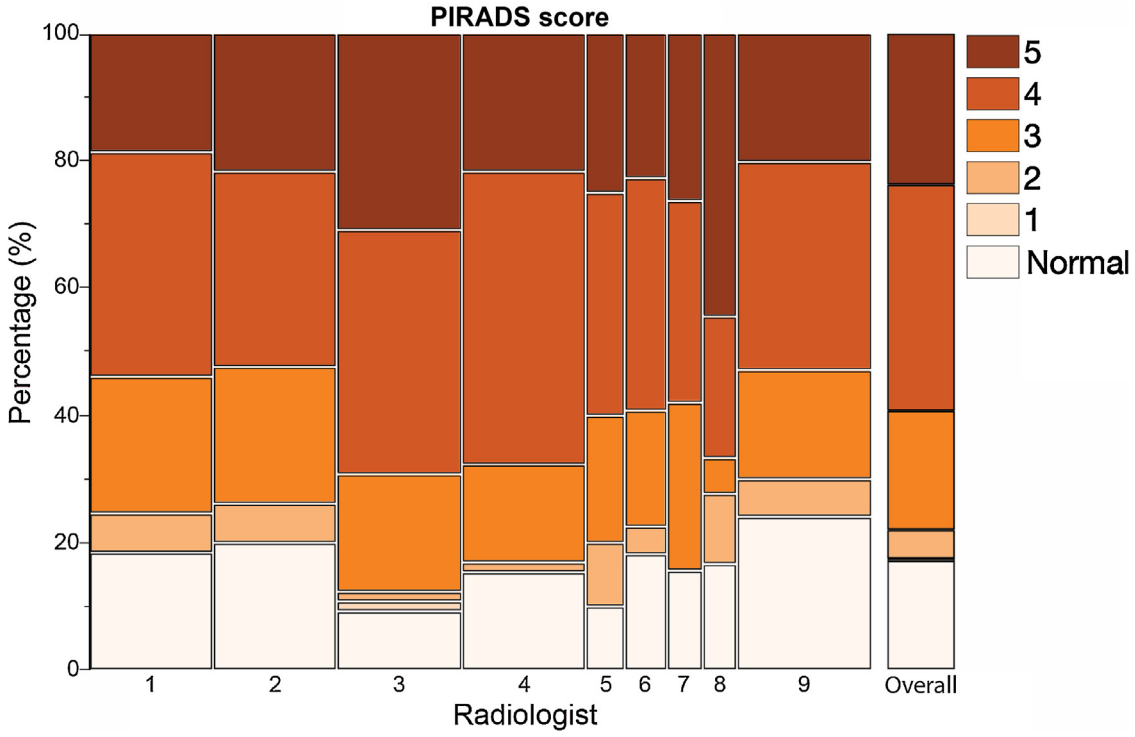


Fig. 1 – Distribution of overall PIRADS score by radiologist among 409 patients. Stratified across radiologists, 18–44% of patients were classified as PIRADS 5, 22–46% as PIRADS 4, 5–26% as PIRADS 3, and 2–10% as PIRADS 2. The column width reflects the number of magnetic resonance images read by each radiologist. PIRADS = Prostate Imaging Reporting and Data System.

4. Discussion

Our study has four important findings. First, the PIRADS score distribution varied across radiologists (Fig. 1). Previous studies have shown differences in PIRADS score distributions [3,10]. As only institutional averages have been reported, it is not clear if this variation is due to differences in patient population, technical features of MRI acquisition, or interpretation of results. Since our study was based at a single institution with a set scanning protocol, we were able to directly assess differences in interpretation among a group of radiologists. We found significant variation in PIRADS score distributions between radiologists. More importantly, we found significant variation in the detection of clinically significant cancer by PIRADS score between individual radiologists.

Second, the correlation between PIRADS score and the presence of cancer (both all cancers and clinically significant cancers) varied across radiologists. Cancer was found in 75% of PIRADS 5, 55% of PIRADS 4, 24% of PIRADS 3, and 21%

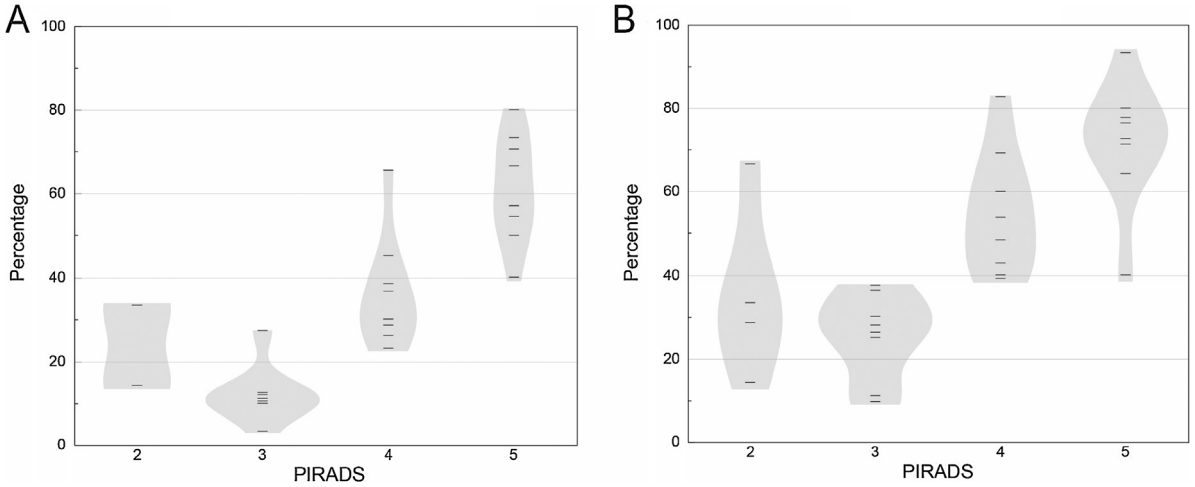


Fig. 2 – Per-lesion cancer yield for each radiologist by PIRADS score for (A) clinically significant cancer and (B) any prostate cancer. Each dash represents the performance of an individual radiologist. For example, the clinically significant cancer yield for PIRADS 5 lesions ranges from 40% to 80% across radiologists. The width of the gray shading associated with each dash reflects the number of lesions classified as that PIRADS score by that radiologist. PIRADS = Prostate Imaging Reporting and Data System.

Please cite this article in press as: Sonn GA, et al. Prostate Magnetic Resonance Imaging Interpretation Varies Substantially Across Radiologists. Eur Urol Focus (2017), https://doi.org/10.1016/j.euf.2017.11.010

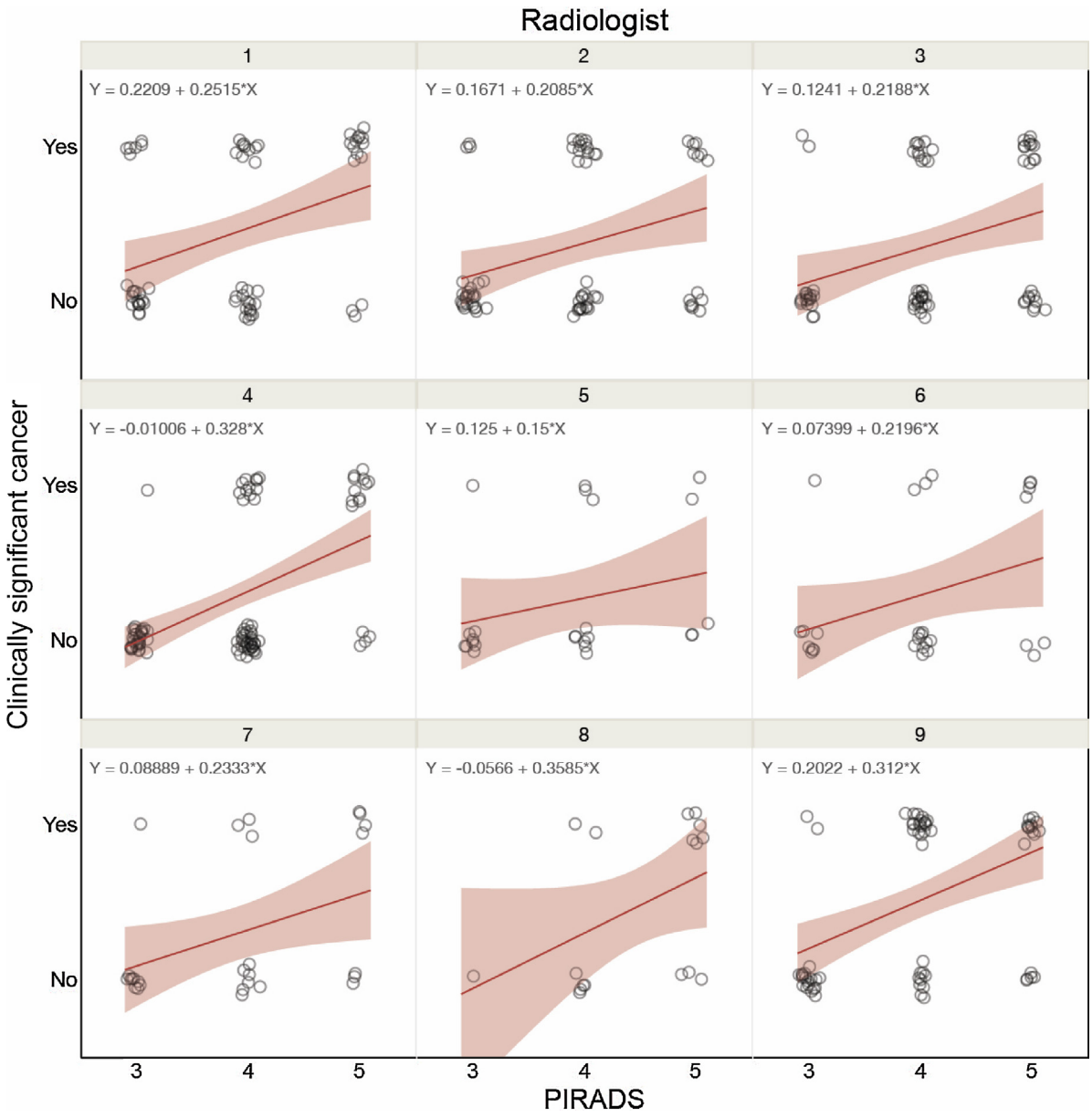


Fig. 3 – Detection of clinically significant prostate cancer for lesions scored as PIRADS ≥3 for each radiologist. A higher PIRADS score is associated with a higher likelihood of clinically significant cancer for all nine radiologists, but the strength of this association (slope of the red line) varies by radiologist. PIRADS = Prostate Imaging Reporting and Data System.

of PIRADS 2 lesions. The strong correlation between level of suspicion on MRI and cancer yield replicates that reported in the literature [3,4,9–12]. However, reporting only the institutional averages masks the underlying variation across radiologists. For example, we found that the presence of clinically significant cancer in PIRADS 5 lesions ranged from 40% to 80%. Variation persisted in adjusted analyses (Fig. 4 and Supplementary Fig. 1).

Third, the proportion of men with a PIRADS score <3 (mean 22%, range 12–30%) and those with a PIRADS score <3 found to have clinically significant prostate cancer (mean 24%, range 13–60%) varied across radiologists. Recommendations from the PROMIS trial suggest that men with PIRADS scores <3 can safely be observed without prostate biopsy [1]. At our site, the proportion of men who would avoid biopsy ranged from 12% to 30%, depending on which radiologist interpreted the scans. More importantly, the proportion of men with a PIRADS score <3 found to have clinically significant cancer on biopsy varied widely across radiologists (13–60%), calling into question the safety of universal adoption of biopsy recommendations from PROMIS.

Fourth, in multivariable analysis controlling for factors associated with prostate cancer, PIRADS score, history of active surveillance, and the radiologist reading the study were associated with clinically significant prostate cancer. The PIRADS score was the most powerful factor associated with biopsy outcome, reconfirming the role of MRI in detecting prostate cancer. However, the fact that the radiologist remained independently associated with detection

Please cite this article in press as: Sonn GA, et al. Prostate Magnetic Resonance Imaging Interpretation Varies Substantially Across Radiologists. Eur Urol Focus (2017), https://doi.org/10.1016/j.euf.2017.11.010

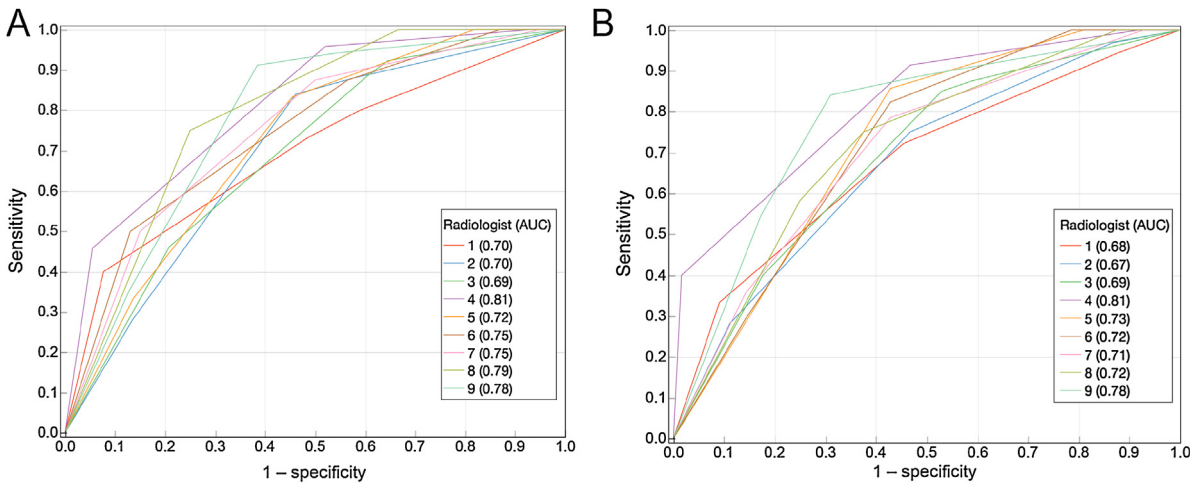


Fig. 4 – Performance of the assigned PIRADS score for detection of (A) clinically significant prostate cancer and (B) any cancer by radiologist. PIRADS = Prostate Imaging Reporting and Data System.

of clinically significant cancer suggests that physicians should be cautious in extrapolating published results to their local radiology practices, unless the MRI and biopsy results have been directly validated. In addition, sensitivity analyses did not detect differences in cancer yield performance over time, either collectively or for individual radiologists. This suggests that quality improvement initiatives are necessary to improve the real-world effectiveness of the PIRADS scoring system, as increased experience over time may be insufficient [13].

Our study is novel because it is the first to systematically evaluate the accuracy of MRI interpretation across radiologists of varying experience in real-time routine clinical care. Most previous publications assessing interobserver agreement used a smaller number of radiologists who retrospectively reviewed MRIs and scored predetermined index lesions. These studies included expert radiologists [7,16] or a mixture of experts and novices [14,17,18] and mostly showed good overall accuracy, with moderate interobserver reproducibility in MRI scoring. A meta-analysis of 21 studies (3857 patients) by Woo et al [15] revealed excellent sensitivity overall (0.89), but substantial heterogeneity across individual studies. This may be because of differences in study design, patient population, image acquisition, scanner type, or radiologist interpretation. By contrast, our study isolated the effect of the radiologist by using the same imaging protocol and drawing patients from the same population. It evaluated variation among attending radiologists with varying prostate MRI experience performing routine clinical care in which the MRIs are read according to the standard workflow. Radiologists do not know if cancer is present and attempt to identify all cancer foci. While some analyses did not control for random variation in patient characteristics across radiologists, variation persisted in adjusted analyses. Therefore, it is unlikely that patient-level differences would explain the large variation in MRI interpretation across radiologists.

In addition to the strength of our real-world study design, targeted and systematic biopsy histopathology was available to compare with MRI-detected lesions and with areas that appeared normal. By using biopsy outcomes instead of prostatectomy, we could include men without cancer. This design precisely recapitulates how MRI is used in clinical practice to select patients for biopsy and guide biopsy targeting. Our large sample size, which allowed multivariable statistical comparison of a large number of radiologists, is another strength. Finally, all biopsies were performed by a single urologist with extensive targeted biopsy experience, thereby minimizing variations in technique that could contribute to differences in outcomes.

The study has some limitations. First, because each MRI study was read by a single radiologist, we could not evaluate interobserver agreement. Thus, it is possible that fundamental differences between patients could contribute to the variation in interpretations. However, as stated above, it is unlikely that this entirely accounts for the large degree of variation. Furthermore, we purposefully chose our design to allow evaluation of consistency across radiologists within routine clinical care. Having all radiologists retrospectively read all MRI scans could not meet this objective because biopsies were targeted on the basis of the clinical read. Second, the study includes interpretations using both PIRADS v1 and v2; studies using PIRADS v1 were not re-read because the objective was to evaluate consistency in routine care. Furthermore, use of PIRADS v1 versus v2 was not significant in multivariable analysis. Third, unlike the PROMIS trial, in which all men underwent a perineal template mapping biopsy, we used the combination of targeted and systematic sampling. Thus, it is possible that some cancers, including clinically significant cancers, were missed because they were not seen on MRI or targeted at biopsy. Fourth, because all MRIs and radiologists came from a single center and all radiologists had fellowship training and/or extensive clinical experience, it is likely that our results underestimate the extent of interobserver variation present across all practice settings.

Notwithstanding these limitations, our study provides important insights into variation in prostate MRI results due to differences between radiologists. While numerous publications have demonstrated the potential of MRI to improve

Please cite this article in press as: Sonn GA, et al. Prostate Magnetic Resonance Imaging Interpretation Varies Substantially Across Radiologists. Eur Urol Focus (2017), https://doi.org/10.1016/j.euf.2017.11.010

cancer diagnosis, use of MRI in different practice settings could be confounded by differences in radiologist interpretation. Internal validation of MRI interpretation with biopsy outcomes should be considered for each site. In addition, our findings suggest a role for MRI training modules and radiologist feedback to improve consistency in clinical practice. We have now implemented a multidisciplinary prostate imaging conference in which prostate MRI scans are retrospectively reviewed alongside pathology results, and structured performance reports are sent to each radiologist.

5. Conclusions

MRI is a powerful tool for prostate cancer diagnosis when performed and interpreted by expert radiologists. However, its performance varies across radiologists. In centers establishing new prostate MRI and biopsy programs, we advise internal validation before widespread adoption. Furthermore, unwanted variation in MRI interpretation suggests a target for quality improvement strategies to reduce such variation.

Author contributions: Geoffrey A. Sonn 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: Sonn, Fan, Ghanouni, Brooks.
Acquisition of data: Sonn, Fan, Ghanouni, Loening, Daniel, To'o, Thong.
Analysis and interpretation of data: Sonn, Fan, Leppert, To'o.
Drafting of the manuscript: Sonn, Ghanouni, Leppert, Wang.
Critical revision of the manuscript for important intellectual content: Sonn, Fan, Ghanouni, Leppert, Wang, Brooks, Loening, Thong.
Statistical analysis: Fan, Leppert.
Obtaining funding: None.
Administrative, technical, or material support: Sonn, Fan, Ghanouni, Leppert.
Supervision: Sonn.
Other: None.

Financial disclosures: Geoffrey A. Sonn 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.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.euf.2017.11.010.

References

[1] Ahmed HU, El-Shater Bosaily A, Brown LC, et al. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. Lancet 2017;389:815–22. http://dx.doi.org/10.1016/S0140-6736(16)32401-1.

[2] Meng X, Rosenkrantz AB, Mendhiratta N, et al. Relationship between prebiopsy multiparametric magnetic resonance imaging (MRI), biopsy indication, and MRI-ultrasound fusion-targeted prostate biopsy outcomes. Eur Urol 2016;69:512–7. http://dx.doi.org/10.1016/j.eururo.2015.06.005.

[3] Rastinehad AR, Waingankar N, Turkbey B, et al. Comparison of multiparametric MRI scoring systems and the impact on cancer detection in patients undergoing MR US fusion guided prostate biopsies. PLoS One 2015;10:e0143404. http://dx.doi.org/10.1371/journal.pone.0143404.

[4] Mehralivand S, Bednarova S, Shih JH, et al. Prospective evaluation of Prostate Imaging Reporting and Data System, version 2 using the International Society of Urological Pathology prostate cancer grade group system. J Urol 2017;198:583–90. http://dx.doi.org/10.1016/j.juro.2017.03.131.

[5] Barentsz JO, Richenberg J, Clements R, et al. ESUR prostate MR guidelines 2012. Eur Radiol 2012;22:746–57. http://dx.doi.org/10.1007/s00330-011-2377-y.

[6] 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.

[7] Rosenkrantz AB, Ginocchio LA, Cornfeld D, et al. Interobserver reproducibility of the PI-RADS version 2 lexicon: a multicenter study of six experienced prostate radiologists. Radiology 2016;280:793–804. http://dx.doi.org/10.1148/radiol.2016152542.

[8] Moore CM, Kasivisvanathan V, Eggener S, et al. Standards of Reporting for MRI-targeted Biopsy Studies (START) of the prostate: recommendations from an international working group. Eur Urol 2013;64:544–52. http://dx.doi.org/10.1016/j.eururo.2013.03.030.

[9] Sonn GA, Natarajan S, Margolis DJA, et al. Targeted biopsy in the detection of prostate cancer using an office based magnetic resonance ultrasound fusion device. J Urol 2013;189:86–92. http://dx.doi.org/10.1016/j.juro.2012.08.095.

[10] Wysock JS, Rosenkrantz AB, Huang WC, et al. A prospective, blinded comparison of magnetic resonance (MR) imaging-ultrasound fusion and visual estimation in the performance of MR-targeted prostate biopsy: the PROFUS trial. Eur Urol 2014;66:343–51. http://dx.doi.org/10.1016/j.eururo.2013.10.048.

[11] Pinto PA, Chung PH, Rastinehad AR, et al. Magnetic resonance imaging/ultrasound fusion guided prostate biopsy improves cancer detection following transrectal ultrasound biopsy and correlates with multiparametric magnetic resonance imaging. J Urol 2011;186:1281–5. http://dx.doi.org/10.1016/j.juro.2011.05.078.

[12] Venderink W, van Luijtelaar A, Bomers JGR, et al. Results of targeted biopsy in men with magnetic resonance imaging lesions classified equivocal, likely or highly likely to be clinically significant prostate cancer. Eur Urol 2017. http://dx.doi.org/10.1016/j.eururo.2017.02.021.

[13] Rosenkrantz AB, Ayoola A, Hoffman D, et al. The learning curve in prostate MRI interpretation: self-directed learning versus continual reader feedback. Am J Roentgenol 2017;208:W92–100. http://dx.doi.org/10.2214/AJR.16.16876.

[14] 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. http://dx.doi.org/10.1148/radiol.2015142818.

[15] Woo S, Suh CH, Kim SY, Cho JY, Kim SH. Diagnostic performance of Prostate Imaging Reporting and Data System version 2 for detection of prostate cancer: a systematic review and diagnostic meta-analysis. Eur Urol 2017;177–88. http://dx.doi.org/10.1016/j.eururo.2017.01.042.

[16] Zhao C, Gao G, Fang D, et al. The efficiency of multiparametric magnetic resonance imaging (mpMRI) using PI-RADS version 2 in the diagnosis of clinically significant prostate cancer. Clin Imaging 2016;40:885–8. http://dx.doi.org/10.1016/j.clinimag.2016.04.010.

[17] Vargas HA, Hötker 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. http://dx.doi.org/10.1007/s00330-015-4015-6.

[18] 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. http://dx.doi.org/10.1002/jmri.25372.

Please cite this article in press as: Sonn GA, et al. Prostate Magnetic Resonance Imaging Interpretation Varies Substantially Across Radiologists. Eur Urol Focus (2017), https://doi.org/10.1016/j.euf.2017.11.010