Published in final edited form as:

AJR Am J Roentgenol. 2017 September ; 209(3): W145–W151. doi:10.2214/AJR.16.17551.

Interreader Agreement of Prostate Imaging Reporting and Data System Version 2 Using an In-Bore MRI-Guided Prostate Biopsy Cohort: A Single Institution's Initial Experience

Daniel I. Glazer¹, William W. Mayo-Smith¹, Nisha I. Sainani¹, Cheryl A. Sadow¹, Mark G. Vangel², Clare M. Tempny¹, and Ruth M. Dunne¹

¹Department of Radiology, Brigham and Women's Hospital, 75 Francis St, Boston, MA 02115

²Department of Radiology, Massachusetts General Hospital, Boston, MA

Abstract

OBJECTIVE—The purpose of this study is to determine the interobserver agreement of the Prostate Imaging Reporting and Data System version 2 (PI-RADSv2) for diagnosing prostate cancer using in-bore MRI-guided prostate biopsy as the reference standard.

MATERIALS AND METHODS—Fifty-nine patients underwent in-bore MRI-guided prostate biopsy between January 21, 2010, and August 21, 2013, and underwent diagnostic multiparametric MRI 6 months or less before biopsy. A single index lesion per patient was selected after retrospective review of MR images. Three fellowship-trained abdominal radiologists (with 1–11 years' experience) blinded to clinical information interpreted all studies according to PI-RADSv2. Interobserver agreement was assessed using Cohen kappa statistics.

RESULTS—Thirty-eight lesions were in the peripheral zone and 21 were in the transition zone. Cancer was diagnosed in 26 patients (44%). Overall PI-RADS scores were higher for all biopsy-positive lesions (mean $\pm$ SD, 3.9 ± 1.1) than for biopsy-negative lesions (3.1 ± 1.0 ; $p < 0.0001$) and for clinically significant lesions (4.2 ± 1.0) than for clinically insignificant lesions (3.1 ± 1.0 ; $p < 0.0001$). Overall suspicion score interobserver agreement was moderate ($\kappa = 0.45$). There was moderate interobserver agreement among overall PI-RADS scores in the peripheral zone ($\kappa = 0.46$) and fair agreement in the transition zone ($\kappa = 0.36$).

CONCLUSION—PI-RADSv2 scores were higher in the biopsy-positive group. PI-RADSv2 showed moderate interobserver agreement among abdominal radiologists with no prior experience using the scoring system.

Keywords

multiparametric MRI; PI-RADS; prostate cancer

In the United States, prostate cancer is the most common cancer diagnosed in men, with 161,360 new cases and 26,730 deaths projected in 2017 [1]. Despite the frequency with

Address correspondence to D. I. Glazer (dglazer@bwh.harvard.edu).

Supplemental Data: Available online at www.ajronline.org.

which prostate cancer is diagnosed, the sensitivity of serum markers is surprisingly low, with a prostate-specific antigen (PSA) level of 4.0 ng/mL being 21% sensitive for detecting any prostate cancer and 51% sensitive for detecting high-grade cancers (Gleason score ≥ 8) [2]. Measuring the percentage of free PSA, which is unbound from protease inhibitors, can provide greater sensitivity in distinguishing benign from malignant causes of total PSA elevation, though some overlap remains [3]. Because of these suboptimal test characteristics, more men are undergoing multiparametric MRI with T2-weighted imaging, DWI, and dynamic contrast-enhanced (DCE) MRI as a method for diagnosing higher-grade prostate cancer [4–6]. Indeed, prostate MRI has even recently been evaluated as a screening test for the general population [7]. In an attempt to standardize reporting, the European Society of Urogenital Radiology published a set of guidelines, the Prostate Imaging Reporting and Data System (PI-RADS) [8]. This system (PI-RADSv1) provided a PI-RADS sum score (on a scale of 4–20) based on review of four pulse sequences (T2-weighted imaging, DWI, DCE-MRI, and MR spectroscopic imaging).

The interreader reproducibility and accuracy of the first version of the PI-RADS guidelines has been evaluated in several studies and has been shown to be moderate, with improved performance in the peripheral zone (PZ) [9–12]. More recent data have suggested that different pulse sequences should be used on the basis of the location of the lesion, with DWI considered the most important sequence in the PZ, and T2-weighted imaging the most important in the transition zone (TZ) [13]. In addition, the value of DCE-MRI, particularly in the TZ, has been called into question [13, 14]. To address these concerns, the American College of Radiology and the European Society of Urogenital Radiology have published an updated version of the PI-RADS guidelines [15]. The key change in PI-RADSv2 is the concept of a dominant pulse sequence, with the DWI score driving the overall suspicion score in the PZ and the T2-weighted imaging score driving the overall suspicion score in the TZ [16]. Other important changes include DCE-MRI to be scored as positive or negative (rather than using a more complicated curve-type analysis); a positive DCE-MRI score will increase the overall suspicion score by 1 point, but only if it makes a clinically relevant difference (from PI-RADS score 3 to PI-RADS score 4); and an overall suspicion score from 1 to 5 should be reported [16]. To date, the interobserver agreement of this version of PI-RADS has only been evaluated in a small number of studies and shown to have moderate-to-substantial reproducibility [17–19].

The purpose of this study is to determine the interobserver agreement of PI-RADSv2 for diagnosing prostate cancer using in-bore MRI-guided prostate biopsy as the reference standard.

Materials and Methods

Patients

This retrospective cohort study was compliant with the HIPAA, and informed consent was waived by the Brigham and Women's Hospital institutional review board. Between January 2010 and August 2013, a total of 90 consecutive in-bore MRI-guided prostate transperineal biopsies were performed in 87 men (mean age, 66.2 years; range, 48–81 years). This cohort of patients has been studied previously in a prospective clinical observation study of in-bore

prostate biopsy, but no data regarding PI-RADSV2 scores were reported [20]. Indication for biopsy was at least one suspected lesion at 3-T multiparametric MRI and at least one of the following: elevated prostate-specific antigen (PSA) level, prior negative transrectal ultrasound (TRUS)-guided biopsy, or inability to undergo TRUS-guided biopsy because of rectal surgery. Initially, 87 patients underwent both multiparametric MRI and MRI-guided biopsy; 16 patients were excluded because they were not treatment naive. Because PI-RADSV2 requires evaluation of T2-weighted imaging, DWI, and DCE-MRI, 12 patients were excluded from the study because the diagnostic MRI examination lacked one of these sequences or was compromised by artifact. The final study population, therefore, consisted of 59 patients (Fig. 1). The 59 included patients (mean age, 65.1 years; range, 48–77 years) had a mean prostate volume of 63.1 cm³ (range, 21.7–195.8 cm³) and a mean PSA level of 12.2 ng/mL (range, 1.7–42.0 ng/mL). The mean ($\pm$ SD) PSA level in patients with cancer was 15.1 $\pm$ 11.0 ng/mL compared with 10.1 $\pm$ 8.0 ng/mL in patients without cancer (p = 0.05). The mean age of patients with cancer was higher than that of patients without cancer (68.0 $\pm$ 6.7 vs 62.8 $\pm$ 6.8 years; p = 0.005). The mean time between diagnostic prostate MRI and MRI-guided biopsy was 64 days (range, 1–180 days). The indication for biopsy for the included patients was prior negative TRUS-guided biopsy with a focal lesion on MRI (n = 38), elevated PSA level (n = 16), and inability to undergo a TRUS-guided biopsy because of prior rectal surgery (n = 6).

Lesion Characteristics

Thirty-eight lesions were in the PZ and 21 lesions were in the TZ. The prostate biopsy yielded a cancer diagnosis in 26 patients (44%). Non-cancerous lesions included 30 biopsies revealing normal prostatic tissue and three showing chronic inflammation (one prostatic intraepithelial neoplasia). The distribution of Gleason scores was as follows: 3 + 3 (n = 6), 3 + 4 (n = 7), 4 + 3 (n = 6), 4 + 4 (n = 4), and 4 + 5 (n = 3), indicating that 20 of 26 tumors were considered clinically significant.

MRI Protocol

Prebiopsy multiparametric MRI was performed using a 3-T MRI device (Signa HDxt 3 T, GE Healthcare) with both endorectal (Prostate eCoil, Medrad) and pelvic phased-array coils. Bowel peristalsis was suppressed by administering 1 mg of glucagon intramuscularly unless there was a contraindication. The imaging sequences used were multiplanar T2-weighted imaging, DWI (b = 500 and 1400 s/mm²), and DCE-MRI after IV injection of gadopentetate dimeglumine (Magnevist, Bayer HealthCare Pharmaceuticals) at 0.1 mmol/kg of body weight at a rate of 3 mL/s via a power injector [21]. Typically, the DCE-MRI pulse sequence consists of 50 consecutive acquisitions over approximately 5 minutes to fully evaluate the enhancement characteristics of the normal prostate and any lesion. In the subset of patients without a rectum, no endorectal coil was placed (n = 6).

In-Bore MRI-Guided Biopsy

All biopsies were performed using a transperineal approach by a single interventional radiologist in either a 3-T wide-bore MRI system (Verio 3 T, Siemens Healthcare) or a ceiling mounted 3-T wide-bore MRI system (IMRIS, IMRIS; and Verio, Siemens Healthcare) [20]. Patients were placed into the bore of the magnet in a supine position with

legs placed in the lithotomy position using an in-house device. The perineum was prepared and draped in sterile fashion, and a needle guidance template was positioned. IV conscious sedation was initiated with midazolam and fentanyl. Next, the prebiopsy multiparametric MR images were fused with the intraprocedural images using nonrigid registration to allow optimal target localization [22]. Using intraprocedural images as well as the fusion overlay, an 18-gauge side-cutting MRI-compatible core biopsy needle (MRI Bio Gun, E-ZEM; Single Action Biopsy Device, US Biopsy; Semi-Automatic Biopsy Gun, Invivo; or Fully Automatic Biopsy Gun, Invivo) was introduced through the perineal template and into the target, and multiple biopsy samples (typically 2–3; range, 1–5) were obtained.

Histopathologic Evaluation

Tissue samples were collected, fixed in formalin, and labeled according to the target number. This detailed labeling allowed correlation between the biopsy specimen and the specific target with prostate. After fixation, the specimens were processed in the standard manner and stained using H and E. All cases were reviewed by a genitourinary pathologist. All cores containing prostate cancer were assigned a Gleason score. Tumors with Gleason scores of 3 + 4 and higher were considered clinically significant [15]. Tumors with Gleason scores of 3 + 3, inflammatory lesions, and benign biopsy results were considered clinically insignificant.

Lesion Scoring

Multiparametric MR images from the diagnostic scan were independently reviewed by three abdominal radiologists (with 1, 4, and 11 years' experience in interpreting prostate MRI) who were blinded to the prostate biopsy results. A single lesion per patient was scored according to the published PI-RADSv2 system [15]. For each lesion, the readers were also asked to document its location on a standard sector map. This study was performed before implementation of the PI-RADS scoring system into the institution's clinical workflow; therefore, a short educational module was provided to each reader before image interpretation. The single index lesion per patient to be scored by the readers was selected after retrospective review of preprocedure multiparametric MR images by two nonblinded radiologists. To be certain that this index lesion was targeted for biopsy, interprocedural images from the transperineal biopsy were also reviewed and correlated with the preprocedure images. Finally, to ensure that each reader was scoring the selected index lesion, each reader received an imaging atlas with the index lesion marked with an ROI on T2-weighted and DWI sequences.

Statistical Analysis

Interobserver agreement for overall PI-RADS score, T2 score, and DWI score was assessed using Cohen weighted kappa statistics. The interobserver agreement for DCE-MRI was assessed using unweighted Cohen kappa statistics. A lesion was considered DCE-MRI positive if two of three or all three of the readers scored it as positive. Correlation coefficients were calculated for each reader pair (reader 1 vs reader 2, reader 1 vs reader 3, and reader 2 vs reader 3) and then were averaged to obtain the overall interobserver agreement. Agreement was defined as substantial ($\kappa \geq 0.61$), moderate ($\kappa = 0.41$ – 0.60), fair ($\kappa = 0.21$ – 0.40), and poor ($\kappa \leq 0.20$). Histologic assessment from in-bore MRI-guided

prostate biopsy was the reference standard for lesion characterization. Chi-square tests and *t* tests were used to compare categorical and continuous variables, respectively. Statistical analyses were performed using Stata (version 13, StataCorp).

Results

PI-RADS Scores

The mean PI-RADS scores were 3.9 ± 1.1 for biopsy-positive lesions and 3.1 ± 1.0 for biopsy-negative lesions ($p < 0.0001$). The mean DWI scores were 3.6 ± 1.1 for biopsy-positive lesions and 3.0 ± 0.9 for biopsy-negative lesions ($p < 0.0001$). The mean T2-weighted imaging scores were 3.8 ± 1.1 for biopsy-positive lesions and 2.9 ± 1.0 for biopsy-negative lesions ($p < 0.0001$). Twenty-two of 59 lesions were scored as DCE-MRI positive. Biopsy-positive lesions were more likely to have a positive DCE-MRI score than were biopsy-negative lesions (15/26 [58%] vs 7/33 [21%]; $p = 0.004$). When only clinically significant lesions were considered (Gleason score of 3 + 4 and higher), the mean overall suspicion score was 4.2 ± 1.0 compared with 3.1 ± 1.0 for clinically insignificant lesions ($p < 0.0001$). Mean PI-RADS scores are summarized in Tables 1 and 2. Example cases are shown in Figures 2–4.

Interobserver Agreement

The interobserver agreement for overall PI-RADS suspicion scores was moderate ($\kappa = 0.45$) for all lesions, moderate ($\kappa = 0.46$) for PZ lesions, and fair ($\kappa = 0.36$) for TZ lesions. The agreement for T2-weighted imaging scores was moderate overall ($\kappa = 0.46$), moderate for PZ lesions ($\kappa = 0.44$), and fair for TZ lesions ($\kappa = 0.36$). The agreement for DWI scores was moderate overall ($\kappa = 0.50$), moderate for PZ lesions ($\kappa = 0.51$), and moderate for TZ lesions ($\kappa = 0.43$). The agreement for DCE-MRI scores was moderate overall ($\kappa = 0.55$), substantial for PZ lesions ($\kappa = 0.67$), and fair for TZ lesions ($\kappa = 0.33$). Interobserver agreement for zonal location (PZ vs TZ) was substantial ($\kappa = 0.80$) and also substantial for specific sector ($\kappa = 0.66$). Kappa statistics are summarized in Table 3, with pairwise comparisons shown in Table S1, which can be viewed in the *AJR* electronic supplement to this article (available at www.ajronline.org).

Discussion

The findings in our study reveal that PI-RADSv2 has overall moderate interobserver agreement ($\kappa = 0.45$). Even given the moderate level of agreement, overall PI-RADS scores and the component T2-weighted imaging scores, DWI scores, and DCE-MRI scores were significantly higher in biopsy-positive lesions compared with biopsy-negative lesions. Agreement on zone and sector location was substantial ($\kappa = 0.80$ and 0.66 , respectively). There was greater interobserver agreement for PZ lesions ($\kappa = 0.46$) than for TZ lesions ($\kappa = 0.36$).

The moderate agreement observed in this study suggests that, despite significant improvements over the original version, some interobserver variability still exists. However, this represents a continued step forward and validates the changes between PI-RADSv1 and PI-RADSv2 [23]. Specifically, the poor performance of DCE-MRI in the TZ support the

conclusions of Baur et al. [13] regarding the value of using a dominant pulse sequence and lessening the influence of DCE MR images. Because of this, at our institution, we report all multiparametric MRI examinations of the prostate with a PI-RADSv2 score. Not only does this standardization provide improved value to our referring clinicians, but the detailed sector location of any lesion aids in the performance of an MRI-guided biopsy [24]. The conclusions reached by our study are concordant with previously published work PI-RADSv1. Although the specifics of the scoring systems are different, Rosenkrantz et al. [10] observed moderate overall inter-reader agreement ($\kappa = 0.45$) with better performance in the PZ ($\kappa = 0.51$) than in the TZ ($\kappa = 0.29$) in their study of PI-RADSv1. A second study comparing PI-RADSv1 scoring to MRI-guided biopsy results by Schimmöller et al. [25] reported good-to-moderate interreader agreement depending on the sequence (T2-weighted imaging, $\kappa = 0.55$; DWI, $\kappa = 0.64$; DCE-MRI, $\kappa = 0.65$), slightly higher than in the current study.

Although there are few studies investigating the performance of PI-RADSv2, an article by Muller and colleagues [17] found moderate interobserver agreement ($\kappa = 0.42$), very similar to that observed in our study. Muller et al. found that PI-RADSv2 has similar performance characteristics in the PZ and the TZ, representing an improvement over PI-RADSv1. In a two-reader retrospective cohort study, Kasel-Seibert et al. [19] found substantial interreader agreement ($\kappa = 0.68$) which was equivalent in the PZ ($\kappa = 0.69$) and the TZ ($\kappa = 0.68$). Finally, a six-reader study by Rosenkrantz et al. [18] found moderate agreement in the PZ ($\kappa = 0.593$) and moderate agreement in the TZ ($\kappa = 0.509$) for lesions with a PI-RADS score of 4 or greater.

As shown in the current study, PI-RADSv2 has moderate reproducibility in treatment-naïve patients. Because the updated guidelines focus on directing clinical management and diagnosing clinically relevant prostate cancer, a PI-RADSv2 score can convey important information to our urologist and oncologist colleagues. Specifically, a PI-RADS score of 1 or 2 can give the treating provider confidence that clinically significant prostate cancer is unlikely to be present. However, PI-RADS scores of 4 or higher should prompt an intervention (typically MRI-ultrasound fusion biopsy or MRI-guided biopsy at our institution) because clinically significant cancer is likely to be present.

Our study has several limitations. First, the study was not designed to assess lesion detection or conspicuity. Readers were asked to score a preselected target to ensure that all readers were, in fact, assessing the same lesion. The study population was enriched with patients who had prior negative TRUS-guided biopsy. This could have led to selection bias, although this cohort frequently undergoes multiparametric MRI. Additionally, given the fact that a number of patients had prior negative TRUS-guided biopsies, this may have increased the percentage of TZ tumors. There is also a selection bias in that the study favors MRI-visible focal prostatic lesions because these were the indications for the MRI-targeted biopsy. In addition, PI-RADS had not been routinely used in clinical practice at our institution, and this could have led to increased variability. However, we attempted to correct for this by conducting a brief training session before the scoring sessions. The pathologic diagnosis of prostate cancer or benignity was established with targeted MRI-guided biopsy rather than prostatectomy. It is possible that there could have been sampling error or the Gleason score

of the tumor may have been different if the entire lesion had been examined. The retrospective nature of the study could have introduced bias. Finally, our single-institution experience using readers fellowship-trained in abdominal imaging may not be generalizable to practicing radiologists throughout the country.

In summary, PI-RADSv2 scores were higher in the biopsy-positive group. PI-RADSv2 showed moderate interobserver agreement among abdominal radiologists with no prior experience using the scoring system.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

D. I. Glazer received grant R25 CA89017, M. G. Vangel received grant P41 EB 015898, and C. M. Tempny received grant P41 EB0115898 from the National Institutes of Health.

References

1. American Cancer Society. Cancer facts & figures 2017. Atlanta, GA: American Cancer Society; 2017.
2. Meigs JB, Barry MJ, Oesterling JE, Jacobsen SJ. Interpreting results of prostate-specific antigen testing for early detection of prostate cancer. *J Gen Intern Med*. 1996; 11:505–512. [PubMed: 8905498]
3. Catalona WJ, Partin AW, Slawin KM, et al. Use of the percentage of free prostate-specific antigen to enhance differentiation of prostate cancer from benign prostatic disease: a prospective multi-center clinical trial. *JAMA*. 1998; 279:1542–1547. [PubMed: 9605898]
4. Hamoen EH, 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; 67:1112–1121. [PubMed: 25466942]
5. Penzkofer T, Tempny-Afdhal CM. Prostate cancer detection and diagnosis: the role of MR and its comparison with other diagnostic modalities—a radiologist's perspective. *NMR Biomed*. 2014; 27:3–15. [PubMed: 24000133]
6. George AK, Pinto PA, Rais-Bahrami S. Multiparametric MRI in the PSA screening era. *Biomed Res Int*. 2014; 2014:465816. [PubMed: 25250323]
7. Nam RK, Wallis CJ, Stojcic-Bendavid J, et al. A pilot study to evaluate the role of magnetic resonance imaging for prostate cancer screening in the general population. *JUrol*. 2016; 196:361–366. [PubMed: 26880413]
8. Barentsz JO, Richenberg J, Clements R, et al. ESUR prostate MR guidelines 2012. *Eur Radiol*. 2012; 22:746–757. [PubMed: 22322308]
9. Rosenkrantz AB, Lim RP, Haghighi M, Somberg MB, Babb JS, Taneja SS. Comparison of inter-reader reproducibility of the prostate imaging reporting and data system and Likert scales for evaluation of multiparametric prostate MRI. *AJR*. 2013; 201:W612–W618. [web]. [PubMed: 24059400]
10. Rosenkrantz AB, Kim S, Lim RP, et al. Prostate cancer localization using multiparametric MR imaging: comparison of Prostate Imaging Reporting and Data System (PI-RADS) and Likert scales. *Radiology*. 2013; 269:482–492. [PubMed: 23788719]
11. Vaché T, Bratan F, Mège-Lechevallier F, Roche S, Rabilloud M, Rouvière O. Characterization of prostate lesions as benign or malignant at multi-parametric MR imaging: comparison of three scoring systems in patients treated with radical prostatectomy. *Radiology*. 2014; 272:446–455. [PubMed: 24937690]

12. Kuru TH, Roethke MC, Rieker P, et al. Histology core-specific evaluation of the European Society of Urogenital Radiology (ESUR) standardised scoring system of multiparametric magnetic resonance imaging (mpMRI) of the prostate. *BJU Int.* 2013; 112:1080–1087. [PubMed: 23937255]
13. Baur AD, Maxeiner A, Franiel T, et al. Evaluation of the prostate imaging reporting and data system for the detection of prostate cancer by the results of targeted biopsy of the prostate. *Invest Radiol.* 2014; 49:411–420. [PubMed: 24598440]
14. Junker D, Schäfer G, Edlinger M, et al. Evaluation of the PI-RADS scoring system for classifying mpMRI findings in men with suspicion of prostate cancer. *Biomed Res Int.* 2013; 2013:252939. [PubMed: 24396825]
15. 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. [PubMed: 26427566]
16. Barentsz JO, Weinreb JC, Verma S, et al. Synopsis of the PI-RADS v2 guidelines for multiparametric prostate magnetic resonance imaging and recommendations for use. *Eur Urol.* 2016; 69:41–49. [PubMed: 26361169]
17. 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; 277:741–750. [PubMed: 26098458]
18. 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. [PubMed: 27035179]
19. Kasel-Seibert M, Lehmann T, Aschenbach R, et al. Assessment of PI-RADS v2 for the detection of prostate cancer. *Eur J Radiol.* 2016; 85:726–731. [PubMed: 26971415]
20. Penzkofer T, Tuncali K, Fedorov A, et al. Transperineal in-bore 3-T MR imaging-guided prostate biopsy: a prospective clinical observational study. *Radiology.* 2015; 274:170–180. [PubMed: 25222067]
21. Hegde JV, Mulkern RV, Panych LP, et al. Multi-parametric MRI of prostate cancer: an update on state-of-the-art techniques and their performance in detecting and localizing prostate cancer. *J Magn Reson Imaging.* 2013; 37:1035–1054. [PubMed: 23606141]
22. Fedorov A, Tuncali K, Fennessy FM, et al. Image registration for targeted MRI-guided transperineal prostate biopsy. *J Magn Reson Imaging.* 2012; 36:987–992. [PubMed: 22645031]
23. Rosenkrantz AB, Oto A, Turkbey B, Westphalen AC. Prostate Imaging Reporting and Data System (PI-RADS), version 2: a critical look. *AJR.* 2016; 206:1179–1183. [PubMed: 26913638]
24. Wibmer A, Vargas HA, Sosa R, Zheng J, Moskowitz C, Hricak H. Value of a standardized lexicon for reporting levels of diagnostic certainty in prostate MRI. *AJR.* 2014; 203:W651–W657. [web]. [PubMed: 25415731]
25. Schimmöller 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; 23:3185–3190. [PubMed: 23756958]

Fig. 1.
Flowchart for selection of patient population for study.

Fig. 2. 66-year-old man with prostate-specific antigen level of 5.92 ng/mL

A–D, T2-weighted image (**A**) shows 1.1-cm T2-hypointense mass (*arrow*) in left anterior peripheral zone that is markedly hyperintense on DW image (*arrow*, **B**) and markedly hypointense on apparent diffusion coefficient map (*arrow*, **C**). Mass (*arrow*) shows early focal enhancement on dynamic contrast-enhanced MR image (**D**). Mass was scored as Prostate Imaging Reporting and Data System score 4 by all three readers. Targeted MRI-guided biopsy revealed Gleason score 4 + 3 cancer. 66-year-old man with prostate-specific antigen level of 5.92 ng/mL.

A–D, T2-weighted image (**A**) shows 1.1-cm T2-hypointense mass (*arrow*) in left anterior peripheral zone that is markedly hyperintense on DW image (*arrow*, **B**) and markedly hypointense on apparent diffusion coefficient map (*arrow*, **C**). Mass (*arrow*) shows early focal enhancement on dynamic contrast-enhanced MR image (**D**). Mass was scored as Prostate Imaging Reporting and Data System score 4 by all three readers. Targeted MRI-guided biopsy revealed Gleason score 4 + 3 cancer.

Fig. 3. 73-year-old man with prostate-specific antigen level of 9.4 ng/mL

A–D, T2-weighted image (**A**) shows 1.9-cm T2-hypointense mass (*arrow*) in anterior transition zone that is isointense on DW image (*arrow*, **B**) and markedly hypointense on apparent diffusion coefficient map (*arrow*, **C**). Mass (*arrow*) shows early focal enhancement on dynamic contrast-enhanced MR image (**D**). Mass was scored as Prostate Imaging Reporting and Data System score 5 by all three readers. Targeted MRI-guided biopsy revealed Gleason score 4 + 3 cancer.

Fig. 4. 68-year-old man with prostate-specific antigen level of 8.4 ng/mL

A–D, T2-weighted image (**A**) shows 0.7-cm linear or wedge-shaped region of T2 signal hypointensity (*arrow*) in right peripheral zone that is mildly hyperintense on DW image (*arrow*, **B**) and mildly hypointense on apparent diffusion coefficient map (*arrow*, **C**). Mass (*arrow*) does not show early focal enhancement on dynamic contrast-enhanced MR image (**D**). Mass was scored as Prostate Imaging Reporting and Data System (PI-RADS) score 3 by two readers and PI-RADS score 4 by third reader. Targeted MRI-guided biopsy revealed benign prostatic tissue.

TABLE 1

Pooled Prostate Imaging Reporting and Data System (PI-RADS) Scores for Malignant and Benign Lesions

PI-RADS Component	Malignant Lesions (<i>n</i> = 26)	Benign Lesions (<i>n</i> = 33)	<i>p</i>
Overall suspicion score	3.9 ± 1.1	3.1 ± 1.0	< 0.0001
T2-weighted imaging	3.8 ± 1.1	2.9 ± 1.0	< 0.0001
DWI	3.6 ± 1.1	3.0 ± 0.9	< 0.0001
DCE-MRI, no. of positive lesions	15	7	0.004

Note—Except where noted otherwise, data are mean ± SD score. Dynamic contrast-enhanced (DCE) MRI scores were considered positive if two of three or three of three readers scored them as positive.

TABLE 2

Pooled Prostate Imaging Reporting and Data System (PI-RADS) Scores for Clinically Significant and Clinically Insignificant Lesions

	Clinically Significant Lesions	Clinically Insignificant	
PI-RADS Component	(<i>n</i> = 20) ^a	Lesions (<i>n</i> = 39)	<i>p</i>
Overall suspicion score	4.15 ± 0.99	3.07 ± 0.99	< 0.0001
T2-weighted imaging	4.0 ± 1.12	2.92 ± 0.97	< 0.0001
DWI	3.83 ± 1.08	2.97 ± 0.90	< 0.0001
DCE-MRI, no. (%) of positive lesions	13 (65)	9 (23)	0.002

Note—Except where noted otherwise, data are mean ± SD score. Dynamic contrast-enhanced (DCE) MRI was considered positive if two of three or three of three readers scored it as positive.

^a Gleason score of 3 + 4 and higher.

TABLE 3

Multireader Kappa Statistics for Overall Prostate Imaging Reporting and Data System (PI-RADS) Suspicion Score and Component Scores

PI-RADS Component	κ
All lesions ($n = 59$)	
Overall suspicion score	0.45
T2-weighted imaging	0.46
DWI	0.50
DCE-MRI	0.55
Zone	0.80
Sector	0.66
Peripheral zone lesions ($n = 38$)	
Overall suspicion score	0.46
T2-weighted imaging	0.44
DWI	0.51
DCE-MRI	0.67
Transition zone lesions ($n = 21$)	
Overall suspicion score	0.36
T2-weighted imaging	0.36
DWI	0.43
DCE-MRI	0.33

Note—All kappa values are weighted except for dynamic contrast-enhanced (DCE) MRI, zone, and sector evaluation, which are unweighted.