

Comparison of likert and PI-RADS v2 scoring in the diagnosis of prostate cancer

Ibrahim Inan¹, Ahmet Aktan², Sidika Seyma Ozkanli³, Asif Yildirim⁴, Ahmet Aslan², Senem Senturk Gucl², Sukru Mehmet Erturk¹

¹Adiyaman University, Faculty of Medicine Department of Radiology, Adiyaman, Turkey

²Medeniyet University, Goztepe Training and Research Hospital, Department of Radiology, Istanbul, Turkey

³Medeniyet University, Goztepe Training and Research Hospital, Department of Pathology, Istanbul, Turkey

⁴Medeniyet University, Goztepe Training and Research Hospital, Department of Urology, Istanbul, Turkey

Copyright © 2018 by authors and Annals of Medical Research Publishing Inc.

Abstract

Aim: We aimed to compare the inter- and intra-observer variability and diagnostic performance of PI-RADS v2 and Likert scoring in the evaluation of prostate cancer on multiparametric prostate MR imaging.

Material and Methods: MRI findings of 53 patients who had histopathologic diagnosis of prostate cancer and 51 patients who had one or more negative transrectal ultrasound guided prostate biopsy were evaluated retrospectively. The images were assessed by three independent observers blinded to histopathologic results of the patients. Intra-observer and inter-observer variability were evaluated by Cohen's Kappa coefficient for all possible pairs of observers. ROC curve analysis was used to determine AUC to evaluate diagnostic performance of Likert and PI-RADS v2 scoring with a 1.5T MR in the diagnosis of prostate cancer.

Results: Kappa values for 4 or higher PI-RADS v2 scores were calculated and $\kappa=0.59-0.71$ for intra-observer and $\kappa=0.48-0.67$ for inter-observer variability were obtained. $\kappa=0.54-0.69$ for intra-observer and $\kappa=0.42-0.75$ for inter-observer reliability were obtained for 4 or higher Likert scoring. AUC was 72% for PI-RADS v2 scores and was 75.9% for Likert scores for determining Gleason 6 or higher prostate cancers. There was not a statistically significant difference in the comparison of Likert and PI-RADS v2 ROC curves.

Conclusion: PI-RADS v2 has no significant superiority compared to Likert scoring in intra-observer and inter-observer reliability. Likert and PI-RADS v2 scoring has no difference in the terms of diagnostic capability of prostate cancers.

Keywords: Prostate Cancer; Magnetic Resonance Imaging; PI-RADS v2; Likert.

INTRODUCTION

Multiparametric magnetic resonance imaging (mpMRI) has been used in the imaging of prostate gland and provides high sensitivity and specificity in the diagnosis of prostate cancer. In many radiology centers, mpMRI has been commonly used in the diagnosis and preoperative staging of the disease (1-3).

The widespread use of the mpMRI emerged different problems in the assessment of images. Different imaging properties of prostate cancers in different sequences, variable tissue characteristics of prostate zones, existence of transitional zone nodules which mimics prostate cancer, inability of detection of low grade prostate cancer foci are some issues which cause the need for use of objective criteria in the evaluation of mpMRI (1,4).

Likert scoring was used as a non-objective test based on overall impression of the observers, but the need of

objective criteria was debated in the ESUR meeting on 2010 and some suggestions were published based on the conclusions (5,6). In the following years, PI-RADS and the next version PI-RADS v2 were used as an objective assessment criteria in mpMRI and routine scoring for all mpMRI were recommended (1). Despite of the claim that PI-RADS v2 has provided objective assessment, both PI-RADS v2 and Likert scoring systems have similar inter-observer reliability according to some authors (7,8). Our aim in this study was to assess reliability of Likert and PI-RADS v2 in the evaluation of mpMRI.

MATERIAL and METHODS

Study design

The patients who underwent mpMRI between January 2015 and January 2016 in Istanbul Medeniyet University Goztepe Training and Research Hospital due to the suspicion of prostate cancer and who had histopathologic

Received: 09.08.2018 Accepted: 30.08.2018 Available online: 05.09.2018

Corresponding Author: Ibrahim Inan, Adiyaman University, Faculty of Medicine Department of Radiology, Adiyaman, Turkey

E-mail: ibrinan@gmail.com

diagnosis after the MRI were evaluated retrospectively. Written informed consent from all patients was obtained. Ethic committee approval was obtained from Istanbul Medeniyet University Goztepe Training and Research Hospital Clinical Researches Ethical Committee with the number of 2016/0007 and the date of 12.01.2016. Inclusion criteria were to have a mpMRI in our institute between January 2015 and January 2016 with the suspicion of prostate cancer and to have histopathologic diagnosis with at least one systematic core biopsy or total prostatectomy. One patient who had histopathologic diagnosis of prostate cancer after transurethral resection of prostate was also included.

The patients who had prostate surgery before the MRI, who received radiotherapy for prostate cancer before the MRI, and who had inadequate MR images for the evaluation such as missing sequence or intense artifacts were excluded. One patient who had extensive peripheral organ invasion causing difficulty of identification of primary cancer were also excluded. Patient selection was summarized on Figure 1. Finally, 104 patients were included in the study.

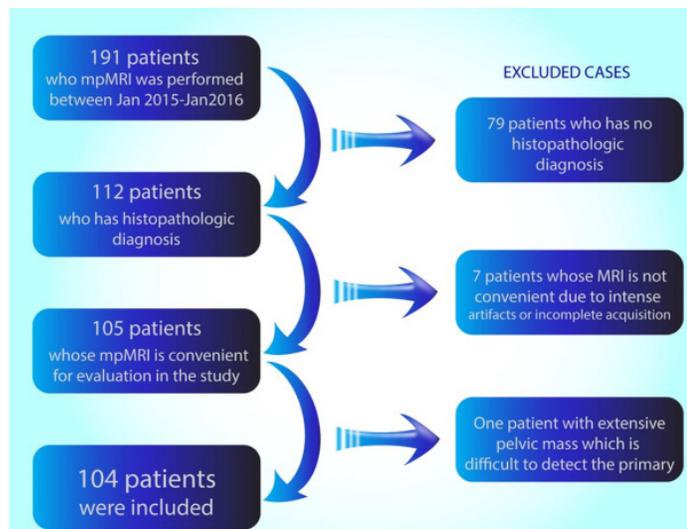


Figure 1. Patient selection

MRI protocol

Imaging was performed with a 1.5 Tesla MRI system (General Electric Optima 450w 1.5T, GE Medical Healthcare, US). Phased array body coil with 16 channels was used. Essential sequences were consisting of T2, DWI and dynamic T1 weighted images in MRI acquisition. Detailed parameters were summarized on Table 1. K-trans, Kep perfusion and ADC maps were obtained on the MRI workstation (Advantage Workstation 4.6, GE Medical Systems, Milwaukee, WI). Dynamic contrast enhanced (DCE) images were performed with 40 phases with the intervals of 8-12 seconds. Intravenous gadobutrol (Gadovist®, BAYER) with the dose of 0.1 mmol/kg and injection rate of 2 cc/sec and 20 ml bolus saline infusion after the contrast injection were applied for DCE.

Table 1. Sequences and parameters used in the image acquisition

MRI Sequences	T1 DCE* Perfusion (3D/FSPGR*)	T2 Fast Recovery Fast Spin Echo (FRFSE Propeller)	DWI (SE/EPI***)
Plane	Axial	Axial, Transverse and Coronal	Axial
Fat saturation	-	-	-
Time to repeat (msec)	4.1	4889	3445
Time to echo (msec)	1.9	121.9	72.3
Flip angle (°)	12	160	90
Slice thickness (mm)	4	3	3
Slice gap (mm)	0	0	0
FOV (mm x mm)	240 x 192	200 x 200	240 x 240
Matrix	160 x 160	320 x 320	96 x 96
NEX (number of excitations)	0.78	2.5	2
b value (sec/mm ²)	-	-	50, 800, 1000

Abbreviations: *Dynamic Contrast Enhanced Imaging **Fast Spoiled Gradient Echo Imaging ***Spin Echo/Echo-Planar Imaging

Data collection

Demographic data, PSA, free PSA, prostate volume of the patients were recorded. MR images were evaluated by three observers which had 12, 4 and 3 years of experience in abdominal magnetic resonance imaging. PSA, free PSA, PSA ratio were presented to observers during the mpMRI assessment, but they were blinded to the histopathologic diagnosis of the patients.

PI-RADS v2 scoring was made according to the up-to-date guideline by the observers (1). Likert scoring was made by the observers based on their overall impression about the images subjectively. Likert scores of 1 to 5 was used, as increasing score indicated increased cancer possibility according to their non-objective impression not based on certain criteria.

In the first stage of evaluation, three observers evaluated the mpMRI of the patients who were independent from each other and scored all patients according to PI-RADS v2. After a week, they assessed the images again blinded to prior PI-RADS v2 scores and scored subjectively according to Likert scoring. To avoid recall bias, reading order of patients was randomized for each evaluation process. After a month interval, they repeated the PI-RADS v2 and Likert scoring independently in the same way and blinded to the prior scores. Finally, for each patient, six PI-RADS v2 and six Likert scores were recorded. Scoring was made based on dominant lesion within the peripheral zone, central gland, and whole gland.

After the all evaluations were finished by the observers, final PI-RADS v2 and final Likert scores were defined by the consensus of all three observers, and it was used to categorize patients to compare inter-observer agreement

in different PI-RADS v2 groups. Also, final PI-RADS v2 and Likert scores were used to perform ROC analysis. Study design was summarized in Figure 2.

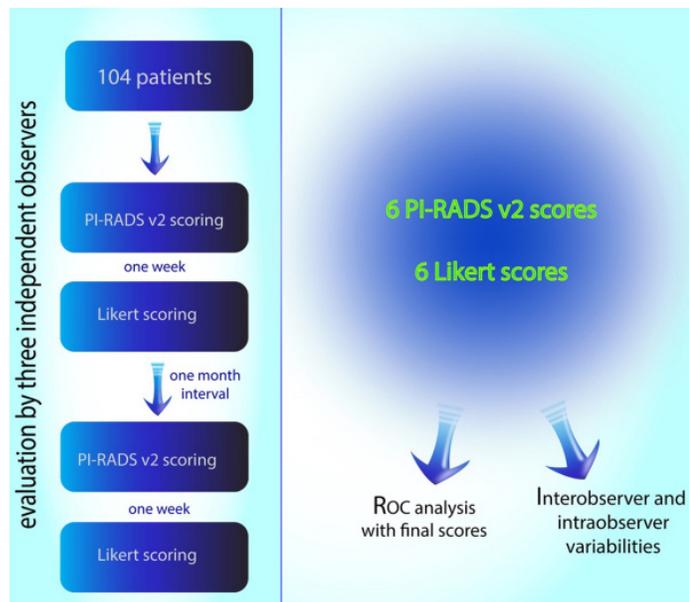


Figure 2. Study design

Scoring was made separately for peripheral zone and central gland and the bigger score was accepted as overall score (Figure 3). The evaluation according to different anatomic zones also assures to evaluate reliability for mpMRI for different anatomic zones in prostate gland.

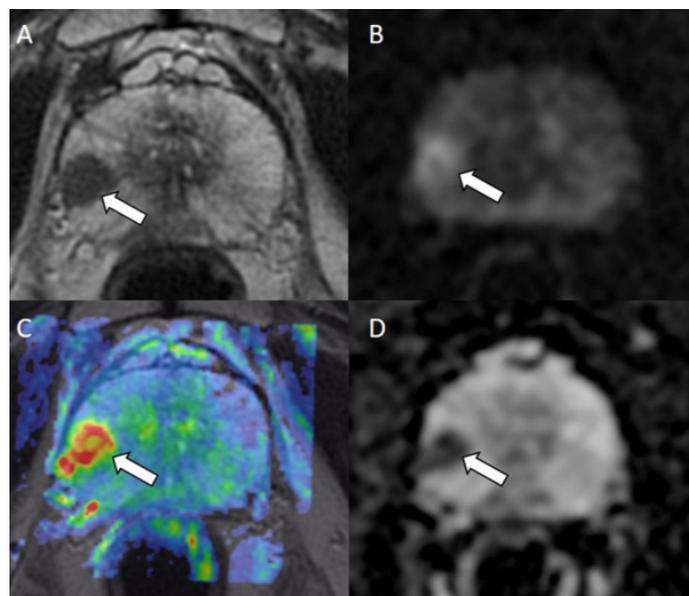


Figure 3. 67-year-old patient with PSA value of 19 ng/ml. The lesion within the right peripheral zone is shown by arrows on T2 (A), diffusion (B) weighted images, K_{trans} (C) and ADC maps (D). The mpMRI was scored as PI-RADS 4 and Likert 5 by all of the observers in all evaluations. After total prostatectomy, the lesion was diagnosed with Gleason 7 prostate cancer

Statistical Analysis

Descriptive data of the patients were expressed as mean, standard deviation (SD), minimum and maximum. Frequencies and ratios were calculated for nominal variables. Interobserver and intraobserver reliability was

calculated for 4 or more scoring of PI-RADS v2 and Likert with Cohen’s Kappa coefficient. ROC analysis was used to assess diagnostic performance of mean PI-RADS v2 and Likert scores. SPSS Statistics 21.0 (New York, United States) software was used in statistical analysis. Additionally, in the evaluation of inter-observer agreement of PI-RADS v2 in different tumor grades and groups, weighted kappa statistics were used.

RESULTS

The mean age of the 104 patients was 64.1 ± 6.27 . Out of the 104 patients, we had the data of PSA in 101 patients and free PSA data in 65 patients. PSA and related data are shown in table 2.

	n	Minimum	Maximum	Mean	Standard Deviation
Age	104	46	79	64.1	7.21
Prostate volume (mm ³)	104	18	230	58.66	33.01
Total PSA (ng/ml)	101	2.68	1000	30.82	139.30
fPSA (ng/ml)	65	0.48	4.9	1.55	0.77
PSA ratio	65	0.04	0.39	0.18	0.08
PSA density	101	0.05	20.00	0.49	2.09

The number of patients who had the diagnosis of prostate cancer was 53 (51%). Remaining 51 patients (49%) had benign histopathologic findings. The biopsy method was core needle biopsy in 79 patients (76%), total prostatectomy in 24 patients (23%), and transurethral resection of prostate gland in one patient who was diagnosed with prostate cancer.

In the intra-observer agreement of PI-RADS v2 and Likert scoring, calculated Kappa ranges and the mean Kappa values are shown in table 3. In the more than 3 scoring, both PI-RADS v2 and Likert showed weak intra-observer agreement on central gland and moderate intra-observer agreement on peripheral zone and overall score.

	Central gland	Peripheral zone	Overall score
PI-RADS v2			
range	0.438-0.507	0.594-0.760	0.592-0.712
mean	0.484	0.684	0.671
Likert			
range	0.463-0.597	0.572-0.787	0.538-0.689
mean	0.550	0.683	0.671

In the inter-observer agreement of PI-RADS v2 and Likert scoring, calculated Kappa ranges and the mean Kappa values are shown in table 4. In the evaluation of the central gland, intra-observer agreement of both PI-RADS v2 and Likert scoring was minimal. Peripheral zone showed higher interobserver agreement, which was moderate. However, the agreement for the overall scores was weak.

Table 4. Kappa values in the inter-observer agreement of PI-RADS v2 and Likert scoring

	Central gland	Peripheral zone	Overall score
PI-RADS v2			
range	0.244-0.508	0.561-0.741	0.481-0.672
mean	0.378	0.645	0.591
Likert			
range	0.124-0.534	0.470-0.827	0.420-0.749
mean	0.350	0.644	0.575

To get a detailed information about inter-observer agreement of different groups of the patients, we evaluated reliability in different categories as: benign, benign and low grade (Gleason 6), malignant (Gleason 6 and higher) and high grade (Gleason 7 or higher). Higher inter-observer agreement was found in the higher-grade tumors while benign and low-grade prostate cancers were reduced the reproducibility of PI-RADS v2 (Figure 4).

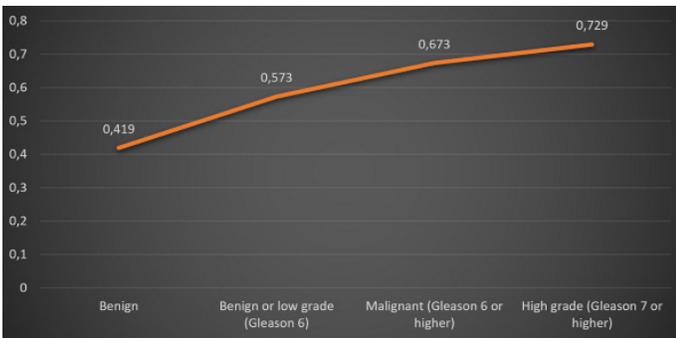


Figure 4. Weighted kappa values of inter-observer agreement of PI-RADS v2 scoring in different grade of prostate cancers

ROC analyze was performed by using final scores obtained from consensus of all observers for each patient. ROC curves of PI-RADS v2 and Likert (Figure 5) are shown. Data which were obtained after ROC analysis is summarized in Table 5.

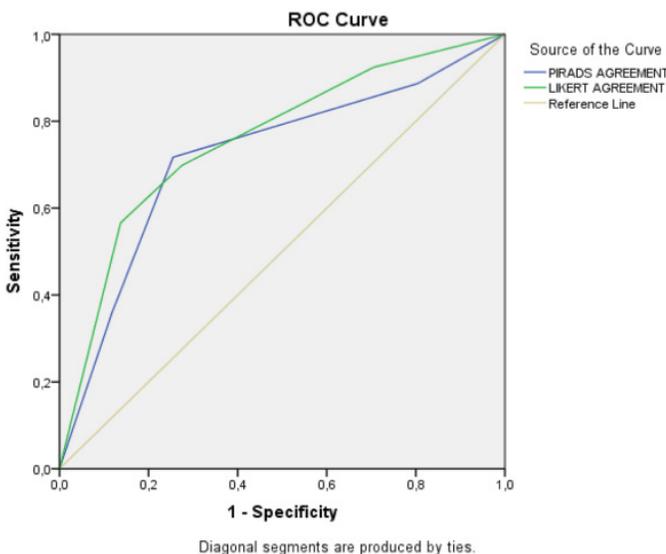


Figure 5. ROC curve which is obtained from the final PI-RADS v2 and Likert scores.

Table 5. Data obtained from the ROC analysis*

	AUC**	p***	Sensitivity	Specificity	Youden Index
PI-RADS v2	0.72	<0.0001	0,717	0,745	0.462
Likert	0.759	<0.0001	0,698	0,725	0.423

*Difference between ROC curves of PI-RADS v2 and Likert scoring systems in the diagnosis of prostate cancer is not significant (p=0.557).

**Best cut off interval for both Likert and PI-RADS v2 was found as 3-4 interval.

***AUC: Area under curve

DISCUSSION

Prostate cancer is very common in male population and mpMRI is used widely in the diagnosis and active surveillance as a diagnostic and non-invasive method. However, it is not easy to visualize the cancer foci in the mpMRI due to different imaging findings and complicated anatomic structure of the gland. PI-RADS v2 is used in the evaluation of mpMRI and claimed to provide more objective data comparing to Likert scale.

Muller et al. in 2015 evaluated 101 cases with mpMRI and fusion imaging guided biopsy. They found kappa values of 0.46 on PI-RADS and 0.55 on non-objective test which showed moderate inter-observer agreement on both scoring systems (9). In the study of Rosenkrantz et al. in 2013, inter-observer agreement was found higher in comparison to our study. Additionally, they also showed that experienced observers have a higher inter-observer agreement. They found similar kappa values for peripheral zone between Likert and PI-RADS, however in transitional zone Likert scoring showed higher agreement in contrast to ours (7). In the study of Rosenkrantz et al, prior version of PI-RADS was used.

Vache et al. in 2014 evaluated 115 cases and found that Likert scoring has a higher diagnostic performance with higher AUC in all observers. In our study, we found higher AUC in Likert scoring which was 75.9% while AUC in PI-RADS v2 scoring was 72%. However, the difference is not statistically significant (p=0.557). Nevertheless, kappa values of Likert scoring were higher than those of PI-RADS in the study of Vache et al in contrast to our study (10).

In the multicentric study of Rosenkrantz et al. in 2016 with six observers, 4 or higher scoring in PI-RADS v2 was assessed. Kappa value was calculated as 0.593 for peripheral zone and 0.509 for central gland (11). We found Kappa values of 0.561-0.741 in peripheral zone and 0.244-0.508 in central gland. The difference between peripheral zone and central gland is more evident in our study.

There are also some studies which compare diagnostic power of Likert and PI-RADS. In the study of Grey et al, AUC of PI-RADS was found 89% and higher than what we found. This might be due to the higher resolution of the 3 Tesla MRI system, or higher individual experience of the authors (12). Roethke et al compared the diagnostic performance of Likert and PI-RADS v2 with a 3 Tesla MRI and found that PI-RADS has a higher sensitivity and

specificity, in contrast to the study of Vache et al (10). As opposed to their findings, we found no difference between Likert and PI-RADS v2 in diagnostic performance (13). Renart-Penne et al in 2015 found no difference between Likert and PI-RADS in diagnostic performance as well (8).

In the current literature, as we know so far, intra-observer agreement hasn't been evaluated. This study pointed out that Likert and PI-RADS v2 have similar intra-observer agreement which is higher in peripheral zone in comparison to central gland.

Additionally, we further evaluated the possible reasons of low agreement in PI-RADS v2 and found that as the tumor grade increases, agreement values also increase. On PI-RADS 5 lesions, agreement is better than those of lower category lesions. This finding indicates that agreement is higher in the patient group which has the need for aggressive treatment.

We have some limitations in our study. Firstly, we used 1.5 Tesla MRI systems in the diagnosis without an endorectal coil. Even if the acquisition parameters were proper for prostate imaging, the magnetic power may affect the results. Secondly, not all patients included in the study were diagnosed with total prostatectomy as a gold standard. The other limitation is that, in the evaluation of inter-observer and intra-observer reliability, we dichotomized the PI-RADS v2 an Likert rating into two groups as <4 and ≥ 4 . Such dichotomization was used to evaluate inter-reader reliability in prior studies with cut off value of ≥ 3 or ≥ 4 (11, 14, 15). Best cut-off value for our study was <4 and ≥ 4 according to the ROC analysis results. Alternative statistical analysis such as Fleiss Kappa for multiple observer or weighted Kappa without dichotomization was also considered; however, Cohen's Kappa with dichotomization was preferred due to similar approach in prior studies. In addition, weighted Kappa was used in subgroup analysis.

CONCLUSIONS

In conclusion, PI-RADS v2 has not an evident superiority in comparison to Likert both in diagnostic accuracy and inter- and intra-observer agreement.

Competing interests: The authors declare that they have no competing interest.

Financial Disclosure: There are no financial supports

Ethical approval: Ethic comitte approval was obtained from Istanbul Medeniyet University Goztepe Training and Research Hospital Clinical Researches Ethical Committee with the number of 2016/0007 and the date of 12.01.2016.

REFERENCES

- 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.
- Lamb BW, Tan WS, Rehman A, et al. Is Prebiopsy MRI Good Enough to Avoid Prostate Biopsy? A Cohort Study Over a 1-Year Period. Clin Genitourin Cancer 2015;13:512-7.
- De Visschere PJ, Naesens L, Libbrecht L, et al. What kind of prostate cancers do we miss on multiparametric magnetic resonance imaging? Eur Radiol 2016;26:1098-107.
- Bomers JG, Barentsz JO. Standardization of multiparametric prostate MR imaging using PI-RADS. Biomed Res Int 2014;2014:431680.
- 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;269:482-92.
- Dickinson L, Ahmed HU, Allen C, et al. Magnetic resonance imaging for the detection, localisation, and characterisation of prostate cancer: recommendations from a European consensus meeting. Eur Urol 2011;59:477-94.
- Rosenkrantz AB, Lim RP, Haghghi M, et al. Comparison of interreader reproducibility of the prostate imaging reporting and data system and likert scales for evaluation of multiparametric prostate MRI. AJR Am J Roentgenol 2013;201:W612-8.
- Renard-Penna R, Mozer P, Cornud F, et al. Prostate Imaging Reporting and Data System and Likert Scoring System: Multiparametric MR Imaging Validation Study to Screen Patients for Initial Biopsy. Radiology 2015;275:458-68.
- 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.
- Vache T, Bratan F, Mege-Lechevallier F, et al. Characterization of prostate lesions as benign or malignant at multiparametric MR imaging: comparison of three scoring systems in patients treated with radical prostatectomy. Radiology 2014;272:446-55.
- 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.
- Grey AD, Chana MS, Popert R. Diagnostic accuracy of magnetic resonance imaging (MRI) prostate imaging reporting and data system (PI-RADS) scoring in a transperineal prostate biopsy setting. BJU Int 2015;115:728-35.
- 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.
- Puryoko AS, Bittencourt LK, Bullen JA, et al. Accuracy and Interobserver Agreement for Prostate Imaging Reporting and Data System, Version 2, for the Characterization of Lesions Identified on Multiparametric MRI of the Prostate. AJR Am J Roentgenol 2017;209:339-49.
- 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.