Orjinal Araştırma / Original Article

Interreader Agreement in Multiparametric Prostate Magnetic Resonance Imaging: A Head-to-Head Comparison between PI-RADSv2 and v2.1

Multiparametrik Prostat Manyetik Rezonans Görüntülemede Okuyucular Arası Uyum: PI-RADSv2 ve v2.1'in Bire Bir Karşılaştırılması

ⒾMehmet Coşkun¹, ⒾMerve Horoz Dönmez², ⒾKazım Ayberk Sinci², ⒾKürşad Dönmez³, ⒾEngin Uluç²

¹Department of Radiology, Health Science University Dr. Behçet Uz Child Disease and Surgery Training and Research Hospital, İzmir, Turkey
²Department of Radiology, İzmir Kâtip Çelebi University Atatürk Training and Research Hospital, İzmir, Turkey
³Department of Urology, İzmir Kâtip Çelebi University Atatürk Training and Research Hospital, İzmir, Turkey

Abstract

Aim: The purpose of this study is to compare PI-RADSv2 with v2.1 in terms of interreader agreement.

Material and Method: Two hundred-two patients who had both multiparametric prostate magnetic resonance imaging (mpMRI) and 12 quadrant systematic biopsies were included in this retrospective study. Acquisition parameters were totally complied with proposal of PI-RADSv2 and 2.1 guidelines. mpMRIs were evaluated by two radiologists independently. Index lesion's score was used to determine diagnostic performance of the systems. Gleason ≥ 3+4 tumors were considered clinically significant prostate cancer (CSCa). Kappa statistic was used to determine interreader agreement on overall PI-RADS scores. Area under the curve (AUC) was calculated in detection of CSCa using receiver operating characteristic curve (ROC).

Result: The numbers of cases with PI-RADSv2 and v2.1 scores from 1 to 5 were 46, 21, 13, 41, 81 and 46, 26, 9, 43, 78, by reader 1 and 51, 10, 18, 43, 80 and 51, 15, 19, 37, 80 by reader 2, respectively. There was "substantial" agreement between two readers for both scoring systems. Kappa values were 0.600 in PI-RADSv2 and 0.624 in PI-RADSv2.1. Ninety-eight (48.5%) patients had CSCa. AUC values of PI-RADSv2 and v2.1 scoring systems were 0.861 and 0.851 for reader 1, 0.873 and 0.883 for reader 2, respectively.

Conclusion: Interreader agreement was "substantial" in mpMRI and slightly improved with PI-RADSv2.1. Diagnostic performance of the two systems were almost equal.

Keywords: Gleason, image-guided biopsy, multiparametric MRI, prostate cancer

Öz

Amaç: Bu yazının amacı, okuyucular arası uyum açısından PI-RADSv2 ile v2.1'in karşılaştırılmasıdır.

Gereç ve Yöntem: Multiparametrik prostat manyetik rezonans görüntüleme (mpMRG) ve 12 kadran sistematik biyopsi yapılan 202 olgu, bu retrospektif çalışmayı oluşturmaktadır. Görüntüleme parametrelerinde PI-RADSv2 ve v2.1'in sunduğu tüm önerilere uyuldu. mpMRG'ler iki radyolog tarafından ayrı ayrı değerlendirildi. Sistemlerin tanısal performansını belirlemede indeks lezyon esas alındı. Gleason ≥ 3+4 tümörler klinik anlamlı prostat kanseri (KAK) kabul edildi. Genel PI-RADS skoru için okuyucular arası uyumun belirlenmesinde kappa istatistiği kullanıldı. KAK tespitinde eğrinin altındaki alan (EAA), alıcı işletim karakteristiği eğrisi kullanılarak hesaplandı.

Bulgular: PI-RADSv2 ve v2.1 için skoru 1'den 5'e olan olgu sayısı okuyucu 1 için sırasıyla 46, 21, 13, 41, 81 ve 46, 26, 9, 43, 78 iken, okuyucu 2 için sırasıyla 46, 21, 13, 41, 81 ve 46, 26, 9, 43, 78 idi. Her iki skorlama sistemi için iki okuyucu arasındaki uyum "iyi" düzeydeydi. Kappa değerleri PI-RADSv2 için 0,600, PI-RADSv2.1 için 0,624 idi. Doksan-sekiz (%48,5) olguda KAK görüldü. PI-RADSv2 ve v2.1 skorlama sistemi için EAA değerleri okuyucu 1 için sırasıyla 0,861 ve 0,851, okuyucu 2 için sırasıyla 0,873 ve 0,883 idi.

Sonuç: Okuyucular arası uyum mpMRG'de "iyi" olup PI-RADSv2.1 ile hafif artmıştır. Her iki sistemin tanısal performansı neredeyse eşittir.

Anahtar Kelimeler: Gleason, görüntüleme eşliğinde biyopsi, multiparametrik MRG, prostat kanseri

INTRODUCTION

Prostate cancer is the second most common cancer of men in Turkey.[1] The incidence is higher in developed countries, reflecting a greater use of screening and diagnostic tools. [2] Multiparametric prostate magnetic resonance imaging (mpMRI) is popular in diagnosing and local staging.[3] The publication of the prostate imaging-reporting and data system (PI-RADS) guideline introduced a standardization in acquisition and reporting.[4] The guideline was revised in 2019 and the current version 2.1 (v2.1) has been published.[5]

The main problem in the evaluation of mpMRI was inconsistency among readers.[6-9] The reason is subjective signal definitions as "mild, moderate, marked" in PI-RADSv2.[6] These qualitative definitions facilitate use of mpMRI while it causes differences among readers in interpretation.[7,9] The current PI-RADSv2.1 presents minor revisions.[5] For example, the addition of "a more pronounced signal change than any other focus in the same zone" provides a more detailed definition of category 4 in peripheral zone (PZ).[3] These updates in signal description propose to increase the compatibility between readers.[5]

The purpose of this study is to compare PI-RADSv2 with v2.1 in terms of interreader agreement. Additionally, the PI-RADS scores were also compared with each other in the diagnosis of clinically significant prostate cancer (CSCa).

MATERIAL AND METHOD

Patient selection

This retrospective study was approved by the local ethic committee (approval number: 12.05.2020-651). The patients who had mpMRI and 12 quadrant systematic biopsies were included between 2019 and 2020. Informed consent was obtained from all participants. mpMRI indications were elevated prostate-specific antigen (PSA ≥ 4) and/or abnormal digital rectal examination (DRE). Biopsy indications were elevated or rising PSA or suspicion PCa in DRE or abnormal mpMRI (PI-RADS score ≥ 3). PI-RADS scoring is not possible after treatment so the patients treated before mpMRI were excluded. Two patients were also excluded due to severe susceptibility artifact secondary to hip prosthesis.

Technical parameters of the mpMRIs

All mpMRIs were acquired on a 1.5T scanner (Aera, Siemens Healthineers, Erlangen, Germany). The protocol included the following sequences: Turbo spin-echo T2-weighted imaging (T2WI) with axial, sagittal, and coronal orientations (Axial T2WI parameters were as follows: repetition time, 5660 msec; echo time, 99 msec; the field of view, 200×180 mm; acquisition matrix, 320×288; slice thickness, 3 mm with no gap), a diffusion-weighted imaging (DWI) with an axial orientation (repetition time, 4000 msec; echo time, 76 msec; b-values, 0, 200, 600 and 1400 sec/mm2; the field of view, 200×180 mm; acquisition matrix, 100×90; slice thickness, 3 mm with no gap) with apparent diffusion coefficient (ADC) mapping, and

dynamic contrast-enhanced (DCE) sequences with an axial orientation (repetition time, 2.48 msec; echo time, 1.52 msec; the field of view, 260×215 mm; acquisition matrix, 160×108; slice thickness 3 mm with 0.3 mm gap; temporal resolution, 7 sec). All parameters were complied with proposal of the PI-RADSv2.1 guideline.[5]

Radiological evaluation

mpMRIs were evaluated by two radiologists individually (reader 1 with 5 years of experience in prostate imaging; reader 2 with 4 years of experience in this field). The radiologists were blinded to any clinical or pathological information. The radiologists scored the lesions two times using PI-RADSv2 and PI-RADSv2.1. Interreader agreement was calculated for PI-RADSv2 and v2.1 respectively. The score of the index lesion was considered in patient-based statistical analysis.

Major scoring revisions in PI-RADSv2.1 were as follows: In PZ, category 2, "linear/wedge shaped" on ADC and high b-value images; category 3, "discrete and different from background" instead of "mild"; category 4, "more pronounced signal change than any other focus in the same zone" instead of "marked". In transition zone (TZ), round, completely encapsulated nodules were clarified as category 1. Atypical nodules were classified as category 2. Overall PI-RADS score upgraded from 2 to 3 for the lesion with T2WI score of 2, and DWI score of 4 or 5.

Pathological assessment

The pathological evaluation was based on the pathology reports. Tumors were graded by the genitourinary pathologists as proposed by the International Society of Urological Pathology (ISUP) in 2014. Accordingly, Gleason 3+3 tumors were categorized as ISUP 1, Gleason 3+4 tumors as ISUP 2, Gleason 4+3 tumors as ISUP 3, Gleason 4+4 tumors as ISUP 4, and Gleason ≥ 4+5 tumors as ISUP 5. ISUP ≥ 2 tumors were considered CSCa as defined in the PI-RADSv2.1.[5] The highest scored biopsy core was considered in patient-based statistical analysis.

Statistical analysis

Statistical analyzes were done using SPSS version 20.0 (IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp.). Kappa statistic was used to determine interreader agreement on overall PI-RADS scores. Accordingly, it was classified as follows: 0.01–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial, and 0.81–0.99, almost perfect.

The PI-RADSv2 and v2.1 scores of the readers were compared in diagnosing CSCa using patient-based analysis. Area under the curve (AUC) was calculated using receiver operating characteristic curve (ROC).

RESULTS

The median age of 202 patients included in this study was 67 (standard deviation, ±7; range, 46-84). The median PSA level was 7.7 ng/dL (±16.7, 1.21-119).

Reader 1 was assigned no lesion in 46 patients. Reader 1 detected 249 lesions in PZ, 29 lesions in TZ lesions in 156 patients. Reader 2 was assigned no lesion in 51 patients. Reader 2 detected 216 lesions in PZ, 25 lesions in TZ in 151 patients. Forty-eight and thirty-six patients had multifocal lesions by reader 1 and reader 2, respectively (**Table 1**). Median dimension of the lesions 15 mm for both readers.

The numbers of cases with PI-RADSv2 and v2.1 scores from 1 to 5 were 46, 21, 13, 41, 81 and 46, 26, 9, 43, 78 by reader 1, respectively (**Table 2**). The numbers of cases with PI-RADSv2 and v2.1 scores from 1 to 5 were 51, 10, 18, 43, 80 and 51, 15, 19, 37, 80 by reader 2, respectively (**Figure 1 and 2**). There was substantial agreement between two readers for both scoring systems overall. Kappa values were 0.600 in PI-RADSv2 and 0.624 in PI-RADSv2.1.

Table 1. Lesion counts on mpMRIs by two different radiologists

| Number of lesions	Reader 1	Reader 2
No lesion	46	51
One lesion	156	151
Two lesions	28	22
Three lesions	14	10
Four lesions	6	4

Table 2. The numbers of cases with PI-RADSv2 and v2.1 by two radiologists

Type of scores	PI-RADSv2, R1 (n)	PI-RADSv2.1, R1 (n)	PI-RADSv2, R2 (n)	PI-RADSv2.1, R2 (n)
Score 1	46	46	51	51
Score 2	21	26	10	15
Score 3	13	9	18	19
Score 4	41	43	43	37
Score 5	81	78	80	80

R, reader; PI-RADS, prostate imaging-reporting and data system; v2, version 2; v2.1, version 2.1; n: number of patients.
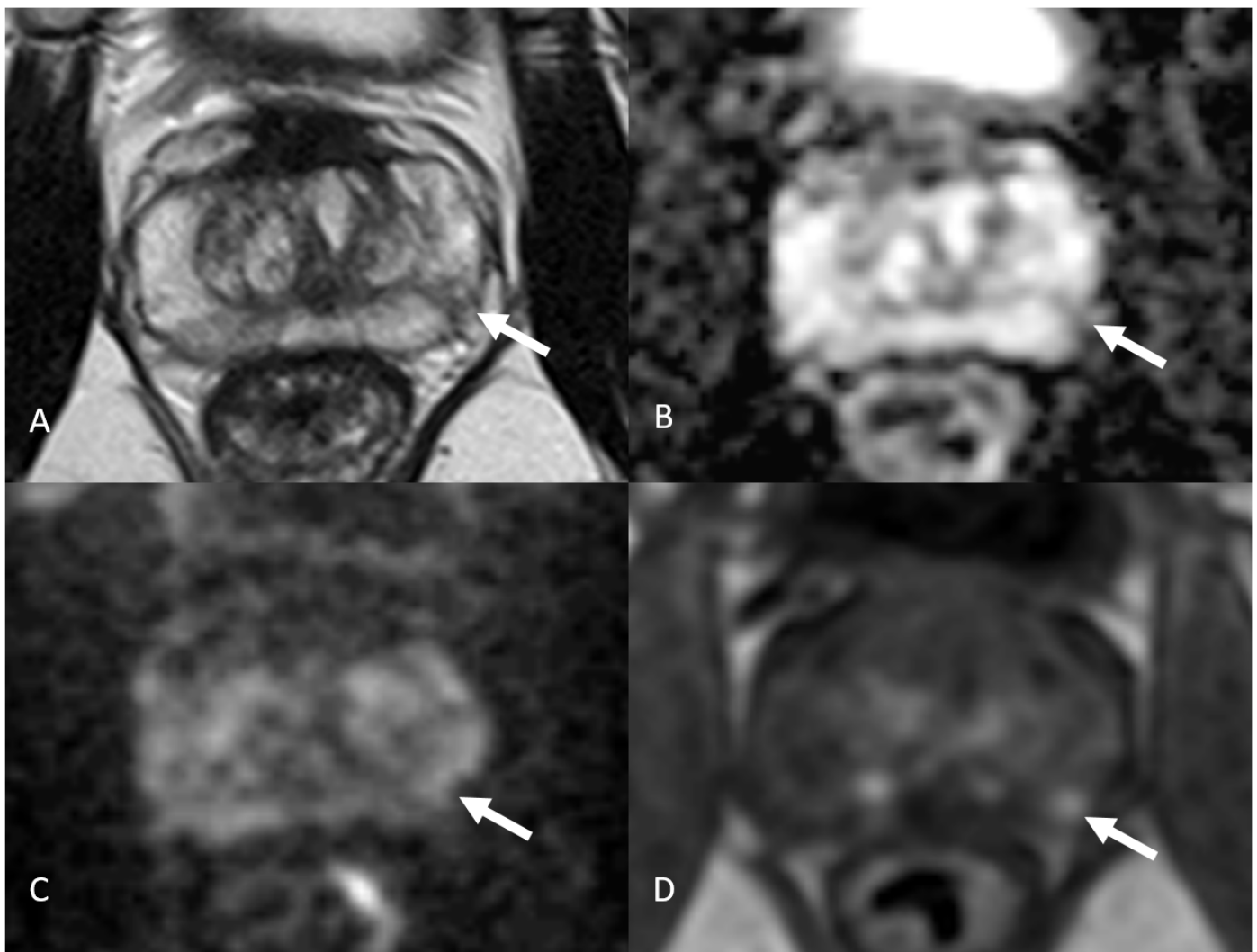
Figure 1. Multiparametric MRI of the 47 year-old man with a PSA of 3.16 ng/dl (a. T2WI, b. ADC map, c. high b-value DWI and d. DCE-MRI). Arrows are showing a left mid PZ lesion. Reader 1 assigned this lesion as category 4 (3+1) for both PI-RADSv2 and v2.1. Reader 2 assigned category 4 (3+1) using PI-RADSv2 but category 2 using PI-RADSv2.1 because of linear-wedge shaped appearance. Systematic biopsy resulted in Gleason 3+3 tumor with 40% involvement on two core from left mid PZ.
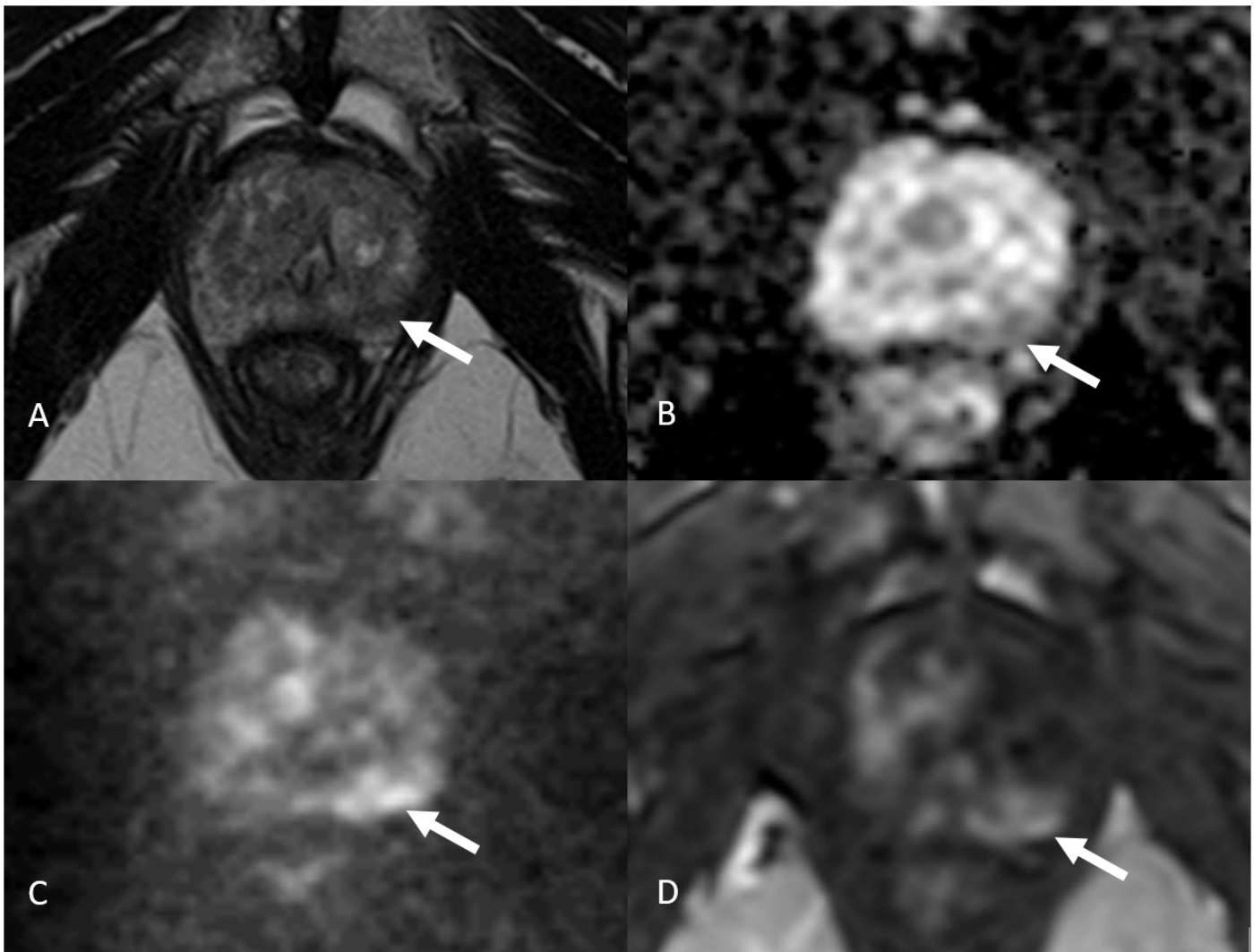
Figure 2. Multiparametric MRI of the 69 year-old man with a PSA of 4.68 ng/dl (a. T2WI, b. ADC map, c. high b-value DWI and d. DCE-MRI). Arrows are showing a left apical PZ lesion with 15mm in diameter. Reader 1 and reader 2 assigned this lesion as category 5 for both PI-RADSv2 and v2.1. Systematic biopsy resulted in Gleason 3+4 tumor with 20% involvement on four cores from left apical and mid PZ.

Pathological outcomes were showed that there was no tumor in 64 (31.7%) patients. The numbers of cases with ISUP scores from 1 to 5 were 40, 44, 28, 13 and 13, respectively. Ninety-eight (48.5%) patients had CSCa. In patient-based analysis, AUC values of PI-RADSv2 and v2.1 scoring systems were 0.861 and 0.851 for reader 1, 0.873 and 0.883 for reader 2, respectively (**Figure 3**).

DISCUSSION

This is one of the earliest studies comparing the PI-RADS guidelines in terms of interreader agreement and diagnosis of CSCa. The results were reflecting minimal progression on interreader reproducibility with PI-RADSv2.1. Detection of CSCa did not improve with current version substantially.

The PI-RADSv2.1 still assigns a score from 1 to 5 to predict likelihood of CSCa. It introduced minor updates for
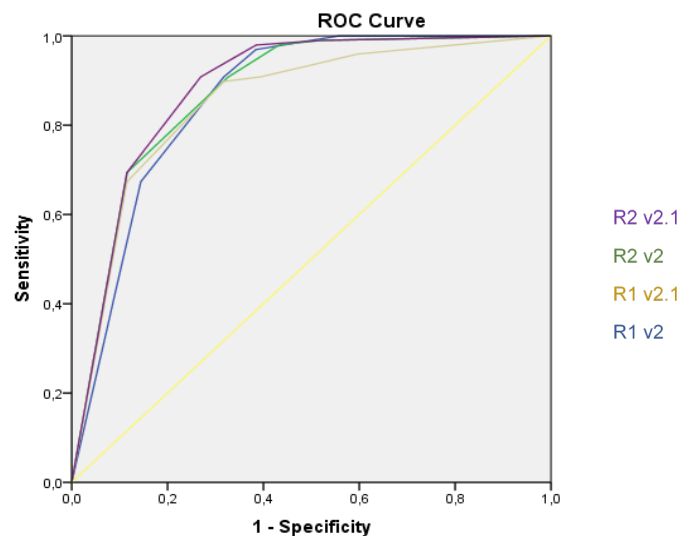
Figure 3. ROC curves of two different scoring systems among two readers in detection of clinically significant prostate cancer

interpretation of mpMRI.[5] DWI category 2 was defined as indistinct hypointense on ADC in PZ. It was upgraded to category 3 when it was defined as mild-moderate hypointense on ADC and isointense-mildly hyperintense on high b-value DWI in the former version.[4] The revised version included morphologic appearance additional to signal descriptive.[3] Category 2 is stated as linear-wedge shaped hypointense on ADC and/or linear-wedge shaped hyperintense on high b-value DWI currently. "Discrete and different from the background" is added to former signal description in clarifying category 3 in PZ.[5] In our study, the numbers of lesions assigned as score 2 were increased using PI-RADSv2.1 by two readers. We observed that some lesions assigned as score 3 in the former version were downgraded to score 2 using PI-RADSv2.1. The term "linear-wedged shaped appearance" was catalyzed downgrading from category 3 into category 2. We did not search for the pathological outcome of those downgraded lesions since the purpose of this work was different. This would be an attractive point for a large-scale study investigating only downgraded lesions.

Round and completely encapsulated nodules are typical benign nodules and evaluated in category 1 now, while they may be scored as category 2 in PI-RADSv2. This revision clarifies the definition of benign typical nodule.[3] In the study of Linhares Moreira et al, the most common score change was observed in TZ with downgrading of typical BPH nodules from category 2 into category 1 with the PI-RADSv2.1.[10] In our study, the number of category 1 lesions did not change for both readers. We had already reported it as "category 1 with typical BPH nodules" in the former version. If we reported all BPH nodules as score 2, we would hardly be able to report as score 1 considering that BPH nodules were quite common.

Interreader agreement was fair to moderate in PI-RADSv2.[8,9,11,12] PI-RADS guidelines were constituted by the experienced radiologists that make uncertain the reproducibility across inexperienced readers. Seven radiologists had a moderate agreement (kappa = 0.591) in detection of index lesion using PI-RADSv2.1.[13] Tamada et al.[14] researched on TZ cancers and reported that interreader agreement was moderate (kappa = 0.580) for PI-RADSv2 and substantial (kappa = 0.645) for PI-RADSv2.1. These findings were replicated in another study comparing the PI-RADSv2 and v2.1 in interpretation of TZ lesion and concluded that PI-RADSv2.1 had higher agreement than PI-RADSv2 (kappa = 0.700 and 0.622, respectively).[15] These findings were also supported for PZ lesions in the study of Bhayana et al.[16] They found that interreader agreement was higher in PI-RADSv2.1 (kappa = 0.64) than PI-RADSv2 (kappa = 0.51). Unlikely, Hötker et al.[17] claimed that interreader agreement was higher in the former version (Kappa = 0.57 for PI-RADSv2 and Kappa = 0.51 for PI-RADSv2.1). But the difference of experience between the readers was higher in this study. Additionally, the potential reason for this decrease may being less familiar with the new version according to the authors. In our study, we directly compared both versions and found that interreader agreement improved slightly with current version (kappa = 0.600 in PI-RADSv2 vs kappa = 0.624 in PI-RADSv2.1). Additional descriptors may enhance the interreader agreement.

There were a few studies searching diagnostic performance of PI-RADSv2.1 which was published recently in September 2019. Wei et al reported that PI-RADSv2.1 had higher diagnostic accuracy with an AUC of 0.929 than PI-RADSv2 with an AUC of 0.899 in diagnosing of CSCa in TZ (15). In another study, both versions had almost identical performance (AUCs of PI-RADSv2 and v2.1 0.874 and 0.879, respectively) in detection of CSCa (17). In our study, PI-RADSv2 and v2.1 scoring systems had parallel performance and AUC values were 0.861 and 0.851 for reader 1, 0.873 and 0.883 for reader 2, respectively.

There were several limitations of this study. Firstly, this was a retrospective study with a small sample size. The results needed to be supported with large-scale prospective study. Secondly, the patients were not sampled with targeted biopsy. Targeted biopsy may enhance the accuracy of mpMRI and provides more accurate radio-pathological concordance. The aim of this work was to compare the two score systems not diagnostic accuracy of mpMRI. Comparison of the systems with each other may trivialize this limitation. Thirdly, index lesion in mpMRI was considered in detection of CSCa. This was problematic in cases with multifocal tumors and prevented lesion-based analysis. So, we introduced patient-based results of CSCa detection. Current findings should also be confirmed by prospective randomized studies.

CONCLUSION

Interreader agreement was substantial and slightly improved with PI-RADSv2.1. Diagnostic performance of PI-RADSv2 and v2.1 were similar in detection of CSCa.

ETHICAL DECLARATIONS

Ethics Committee Approval: This study was approved by institutional Review Board of İzmir Katip Çelebi University (Non-Interventional Clinical Studies, 12.05.2020, 651).

Informed Consent: Due to the retrospective design of the study, informed consent of the patients was not necessary.

Referee Evaluation Process: Externally peer-reviewed.

Conflict of Interest Statement: The authors have no conflict of interest to declare.

Financial Disclosure: The authors declared that this study received no financial support.

Author Contributions: The authors declare that they have all participated in the design, execution and analysis of the paper, and that they have approved the final version.

REFERENCES

1. iarc.fr [World Health Organization]. Lyon: International agency for research on cancer, GLOBOCAN 2018. [updated October 2020; cited 1 November 2020] available from: https://gco.iarc.fr/today/data/factsheets/populations/792-turkey-fact-sheets.pdf.

2. Rawla P. Epidemiology of Prostate Cancer. World J Oncol. 2019;10(2):63-89.

3. Walker SM, Türkbey B. PI-RADSv2.1: Current status. Turk J Urol. 2020.

4. Weinreb JC, Barentsz JO, Choyke PL, Cornud F, Haider MA, Macura KJ, et al. PI-RADS Prostate Imaging - Reporting and Data System: 2015, Version 2. Eur Urol. 2016;69(1):16-40.

5. Turkbey B, Rosenkrantz AB, Haider MA, Padhani AR, Villeirs G, Macura KJ, et al. Prostate Imaging Reporting and Data System Version 2.1: 2019 Update of Prostate Imaging Reporting and Data System Version 2. Eur Urol. 2019;76(3):340-51.

6. Smith CP, Harmon SA, Barrett T, Bittencourt LK, Law YM, Shebel H, et al. Intra- and interreader reproducibility of PI-RADSv2: A multireader study. J Magn Reson Imaging. 2019;49(6):1694-703.

7. Rosenkrantz AB, Babb JS, Taneja SS, Ream JM. Proposed Adjustments to PI-RADS Version 2 Decision Rules: Impact on Prostate Cancer Detection. Radiology. 2017;283(1):119-29.

8. Greer MD, Shih JH, Lay N, Barrett T, Bittencourt L, Borofsky S, et al. Interreader Variability of Prostate Imaging Reporting and Data System Version 2 in Detecting and Assessing Prostate Cancer Lesions at Prostate MRI. AJR Am J Roentgenol. 2019:1-8.

9. Girometti R, Giannarini G, Greco F, Isola M, Cereser L, Como G, et al. Interreader agreement of PI-RADS v. 2 in assessing prostate cancer with multiparametric MRI: A study using whole-mount histology as the standard of reference. J Magn Reson Imaging. 2019;49(2):546-55.

10. Linhares Moreira AS, De Visschere P, Van Praet C, Villeirs G. How does PI-RADS v2.1 impact patient classification? A head-to-head comparison between PI-RADS v2.0 and v2.1. Acta Radiol. 2020:284185120941831.

11. Muller BG, Shih JH, Sankineni S, Marko J, Rais-Bahrami S, George AK, et al. Prostate Cancer: Interobserver Agreement and Accuracy with the Revised Prostate Imaging Reporting and Data System at Multiparametric MR Imaging. Radiology. 2015;277(3):741-50.

12. Greer MD, Brown AM, Shih JH, Summers RM, Marko J, Law YM, et al. Accuracy and agreement of PIRADSv2 for prostate cancer mpMRI: A multireader study. J Magn Reson Imaging. 2017;45(2):579-85.

13. Brembilla G, Dell'Oglio P, Stabile A, Damascelli A, Brunetti L, Ravelli S, et al. Interreader variability in prostate MRI reporting using Prostate Imaging Reporting and Data System version 2.1. Eur Radiol. 2020;30(6):3383-92.

14. Tamada T, Kido A, Takeuchi M, Yamamoto A, Miyaji Y, Kanomata N, et al. Comparison of PI-RADS version 2 and PI-RADS version 2.1 for the detection of transition zone prostate cancer. Eur J Radiol. 2019;121:108704.

15. Wei CG, Zhang YY, Pan P, Chen T, Yu HC, Dai GC, et al. Diagnostic Accuracy and Inter-observer Agreement of PI-RADS Version 2 and Version 2.1 for the Detection of Transition Zone Prostate Cancers. AJR Am J Roentgenol. 2020.

16. Bhayana R, O'Shea A, Anderson MA, Bradley WR, Gottumukkala R, Mojtahed A, et al. PI-RADS versions 2 and 2.1: Interobserver Agreement and Diagnostic Performance in Peripheral and Transition Zone Lesions Among Six Radiologists. AJR Am J Roentgenol. 2020.

17. Hötker AM, Blüthgen C, Rupp NJ, Schneider AF, Eberli D, Donati OF. Comparison of the PI-RADS 2.1 scoring system to PI-RADS 2.0: Impact on diagnostic accuracy and inter-reader agreement. PLoS One. 2020;15(10):e0239975.