European Journal of Radiology 114 (2019) 111–119

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

Comparison of diagnostic performance between two prostate imaging reporting and data system versions: A systematic review

Wei Li[a], Chun Xin[a,*], Lanlan Zhang[b], Anding Dong[a], Haibing Xu[a], Yiman Wu[a]

[a] Department of Medical Imaging, Jiangsu Vocational College of Medicine, Yancheng City, China
[b] Department of Pediatrics, Yancheng Maternal and Child Health Hospital, Yancheng, China

ARTICLE INFO

Keywords:
Prostate cancer
Diagnostic performance
Prostate imaging reporting and data system
Multi-parametric MRI

ABSTRACT

Objective: To compare the diagnostic performance between Prostate Imaging Reporting and Data System version 1(PI-RADS v1) and PI-RADS v2 for detection of prostate cancer (PCa).

Methods: A systematic literature search was performed from inception to September 31, 2018, in following databases MEDLINE, EMBASE, Cochrane Library, Google Scholar, in addition to Chinese National Knowledge Infrastructure (CNKI) and Wanfang Data database. Sensitivity and specificity of individual studies along with summary estimates were calculated and presented in forest plots. Multiple subgroup analyses and meta-regression were performed to investigate the heterogeneity. Quality of included studies was assessed using the Quality Assessment of Diagnostic Accuracy Studies-2 tool.

Results: 14 studies involving head to head comparison between PI-RADS v1 and v2 were included, with a total of 1682 patients. The pooled sensitivity for PI-RADS v1 and PI-RADS v2 were 85.37% (95% CI 79.29%–89.89%) and 87.79% (95% CI 79.84%–92.88%), with pooled specificity of 77.26% (95% CI 57.29%–89.58%) and 76.73% (95% CI 58.69%–88.44%), respectively. There was no substantial difference in diagnostic accuracy between PI-RADS v1 and PI-RADS v2 ($P = 0.57$ for sensitivity and $P = 0.96$ for specificity). Multiple subgroup analyses and meta-regression suggested these two scoring systems had comparable diagnostic performance on magnetic field strength, zonal anatomy, and outcome assessment. For the transitional zone, it seemed that PI-RADS v2 had higher sensitivity than PI-RADS v1 (90.1% vs. 80.59%), but the difference was not substantial ($P = 0.17$).

Conclusion: PI-RADS v2 has slightly higher sensitivity but at the expense of minor decreased specificity. Thus, on the whole PI-RADS v1 and PI-RADS v2 have comparable diagnostic accuracy.

1. Introduction

Prostate cancer (PCa) has been the number one form of cancer among males in western countries [1]. In spite of most types of PCa grow slowly and may need no intervention, the others are so aggressive that spread quickly. Therefore, early detection and location of PCa will play an important role in proper treatment, then leading to a better chance of successful treatment. However, some traditional diagnostic tools for PCa such as serum prostate-specific antigen (PSA), digital rectal examination (DRE) and trans-rectal ultrasound (TRUS) guided biopsy could not fulfill the requirement because of limitations, inconvenience and insufficient accuracy [2,3]. The development of MRI technology has been recognized as one of the most effective examinations approaches in the detection, location and phase dividing of prostatic symptom due to its better spatial resolution and soft tissue contrast. However, depending on conventional MRI sequences such as T1-weighted imaging (T1WI) and T2-weighted imaging (T2WI) could not satisfy the requirement of sensitivity and specificity for PCa as expected. Therefore, multi-parametric magnetic resonance imaging (mpMRI) examination including dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI), diffusion-weighted imaging (DWI) and magnetic resonance spectroscopy imaging (MRS, or MRSI) were introduced to diagnosis and classification of PCa, and the diagnostic performance was confirmed by published studies [4–6]. Nevertheless, those advanced techniques are not widely applied to PCa, because of lacking standardized diagnostic criteria for reporting of results, which could lead to considerable variability in interpretation among radiologists and sometimes even controversial [7].

To standardize the assessment and reporting PCa with mpMRI, in 2012, the European Society of Urogenital Radiology (ESUR) released

* Corresponding author at: Department of Medical Imaging, Jiangsu Vocational College of Medicine, Yancheng City, China.
E-mail addresses: hfjs2000@126.com (W. Li), hfjs2000@outlook.com (C. Xin), contribute_sci@126.com (L. Zhang), 153902705@qq.com (A. Dong), 11368@jsmc.edu.cn (H. Xu), wamzqy0661@sina.com (Y. Wu).

https://doi.org/10.1016/j.ejrad.2019.03.016
Received 2 January 2019; Received in revised form 13 March 2019; Accepted 19 March 2019

the first edition of Prostate Imaging Reporting and Data System (PI-RADS v1), a scoring system based on expert consensus and provided guideline for specific mpMRI protocols (T2WI, DCE-MRI, DWI and MRS). It was the initial attempt to introduce a standardized scoring system to detection of PCa, and several subsequent studies had validated its diagnostic performance and reproducibility. A meta-analysis reported that the summary estimates of sensitivity and specificity achieved 78% and 79%, respectively. However, lacking detailed methods to calculate overall scores impeded the wide clinical application of the PI-RADS v1, and literature reported that different MRI sequences may have an advantage over others in different regions of the prostate [8]. Thus in 2015, an updated version of PI-RADS was released to address the weakness existed in PI-RADS v1.

After PI-RADS v2 released, many studies had validated the diagnostic performance and reproducibility of it. A meta-analysis [9] revealed the pooled sensitivity and specificity for PI-RADS v2 were 89% and 73%, furthermore, this study also made a comparison between two PI-RADS versions, found that PI-RADS v2 had a substantially higher sensitivity (95% vs. 88%, P = 0.04), while for specificity there was no significant difference. However, more studies published recently reported that there was no considerable discrepancy in diagnostic performance between PI-RADS v1 and PI-RADS v2. Therefore, this systematic review aimed to make a detailed comparison between the two PI-RADS versions for the detection of PCa.

2. Methods

This systematic review and meta-analysis was performed and reported compliance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines [10]. The primary outcome for this meta-analysis was a comprehensive comparison between PI-RADS v1 and PI-RADS v1v2 on diagnostic performance.

2.1. Literature search

A systematic literature search was performed from inception to Sep 31, 2018, in following databases: MEDLINE, EMBASE, Cochrane Library, Google Scholar, in addition to Chinese National Knowledge Infrastructure (CNKI) and Wanfang Data database, with no language restrictions. Two reviewers independently identified eligible studies involving head to head comparison on diagnostic performance between PI-RADS v1 and PI-RADS v2. The searched was supplemented by manually screening the reference list of identified articles and reviews. Search terms, synonyms and abbreviations for the present study including: ([prostate cancer] OR [prostatic cancer] OR [prostate tumor] OR [prostatic tumor] OR [PCa]) AND ([multiparametric] OR [multiparametric]) AND ([MRI] OR [MR]) AND ([prostate imaging reporting and data system] OR [pi-rads] OR [pirads]).

2.2. Inclusion criteria and exclusion criteria

Studies were included only if they met all of the following criteria: 1) participants were cases or suspicions of PCa, 2) using mpMRI, with protocol of T2WI, DCE-MRI, DWI to detection of PCa, 3) both PI-RADS v1 and PI-RADS v1v2 scoring systems were used in a single study, 4) reliable reference such as histopathological examination of radical prostatectomy (RP) or biopsy were used to validate the examination results, and 5) all studies should contain sufficient details for reconstructing the 2 × 2 contingency table to evaluate the diagnostic accuracy.

Studies were excluded if they fulfill any of the following criteria: 1) studies with ordinary MRI protocols but not performed mpMRI, 2) involved only one scoring system and deficient in head to head comparison, 3) for other prostate-related diseases such as benign prostate hyperplasia (BPH) rather than PCa, 4) could not obtain the detailed data for reconstruction of the 2 × 2 tables, and 5) reviews, meta-analyses, guidelines, conference abstracts and editorials letters. Two reviewers independently undertook the literature searches, and disagreements were resolved through consensus discussion with the third reviewer.

2.3. Data extraction and quality assessment

A data collection form was developed to standardize the data extraction process. We extracted the interesting data from included studies as follows: 1) demographic characteristics, including patient age, prostate volume, prostate-specific antigen (PSA) levels, Gleason scores and prevalence of PCa, 2) study characteristics, which including authors, publication year, country and institution, study design and sample size, number of readers and experience, reference standard, interval between MRI and RP, analysis locations, cutoff values, true positive, false positive, false negative and true negative numbers both for PI-RADS v1 and PI-RADS v2, definition of clinically significant PCa (CS PCa), outcomes assessed (any or CS PCa), and kappa values, and 3) technical characteristics of mpMRI, including MRI manufacturer and model, magnetic field strength, sequences and types of coil used for diagnosis. In the case of accuracy was reported for different radiologists, we selected the most experienced or the first one when the experience was not reported.

Methodological quality was evaluated utilizing the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAs-2) tool [11], which was the most used tool for evaluation of diagnostic accuracy studies. In this scoring system there are 4 domains need to assess: patient selection domain describes the strategy used for selecting patients and the patients included; index test domain illustrates the test studies and how to perform and interpret; reference standard domain describes the reference standard employed and how to perform and interpret; finally, the flow and timing domain illustrates the flow of patient included and excluded. Each of these 4 domains evaluated “yes”, “no” or “unclear”. The procedure of data exaction and literature quality evaluation were conducted by two reviewers independently, and all disagreements were resolved by means of discussion with the third reviewer to achieve a consensus.

2.4. Data synthesis and statistical analysis

Sensitivity, specificity and corresponding 95% confidence intervals (95% CIs) for each study were calculated and presented in the forest plots to investigate the heterogeneity, and pooled estimates of sensitivity and specificity were calculated using the bivariate model [12,13] and hierarchical summary receiver operating characteristic (HSROC) model [14]. The Summary ROC curves for PI-RADS v1 and PI-RADS v2 were plotted to visually present the comparison between two versions.

2.5. Subgroup analyses and meta-regression

Subgroup analyses and meta-regression were performed to explore causes of heterogeneity, and following covariates were added to the bivariate model to assess whether associated with heterogeneity: 1) magnet field strength of MRI 1.5 T, 2) magnet field strength of MRI 3.0 T, 3) outcome assessment of any PCa, 4) outcome assessment of CS PCa, 5) analyzing for peripheral zone (PZ) PCa, and 6) analyzing for transition zone (TZ) PCa. Publication bias was evaluated by utilizing the Deeks funnel plot, in which the asymmetry as used to test the statistical significance. The “metandi” and “midas” modules in Stata 15.2 (StataCorp, Texas, USA) were used to perform calculations and statistical analysis, along with RevMan 5.3(Review Manager, Cochrane Nordic). The P values < 0.05 was regarded as statistically significant.

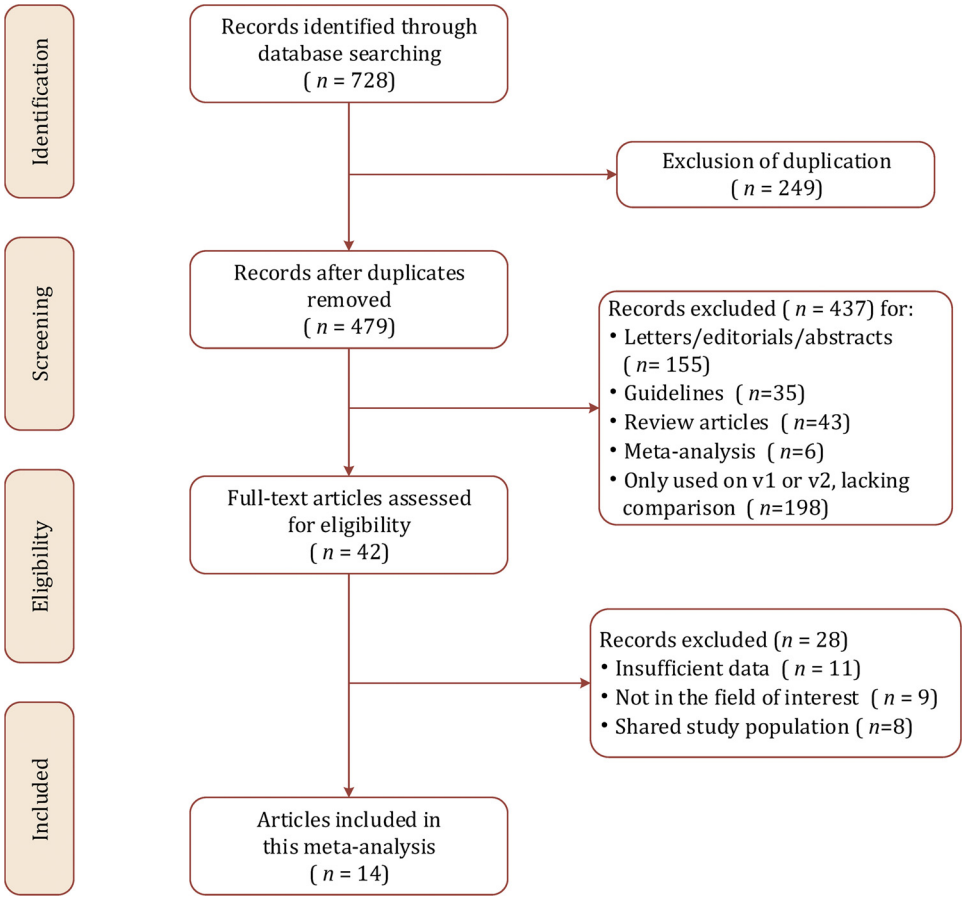

Fig. 1. PRISMA flow diagram showing the study selection process.

3. Results

3.1. Literature search

Initially, literature searching yielded 728 results. Of them, 249 were removed for duplicates and after manually screening, 198 were excluded for just used one scoring system of PI-RADS v1 or PI-RADS v2 and lacking head to head comparison. Full text reviewing was performed for the remaining 42 articles, then 28 were ruled out for following reasons: 1) did not include outcomes of interested, 2) overlapping patient population with other studies, 3) unable to reconstruct the 2 × 2 contingency table owing to deficient data. Finally, 14 original articles with a total of 1682 patients involving directed comparison of diagnostic performance between PI-RADS v1 and v2 were included in this systematic review [15–28]. Fig. 1 shows a concise overview of the literature selection process.

3.2. Characteristics of included studies

The detailed demographic characteristics are presented in Table 1. The sample size of the study population ranged from 30 to 401, with a mean age of 61.7–73.45 years of old. The Gleason scores for studies ranged from 5 to 10, with a mean PSA level of 5.9–26.3 ng/ml. All studies reported that the MRI examinations were performed prior to biopsy or RP. Only 4 studies [15,18,19,23] were described as prospective, whereas the remaining 10 studies [16,17,20–22,24–28] were retrospective.

Table 2 shows the summarized study and technical characteristics. 1.5 T MRI scanners were performed in 4 studies [15–18], and all of the remaining 10 studies used 3.0 T scanners [19–28]. Only one study [17] used the endorectal coil, but there were respectively six [18,19,22,24,26,27] and four studies [16,17,20,28] used body coil and pelvic phase array (PPA) coil, and the remaining 4 studies did not use any type of coil. Nearly all studies used mpMRI sequences of T2WI, DCE, and DWI, and 6 studies [17–19,24,27,28] introduced ADC which derived from DWI to improve the diagnostic accuracy. The radical prostatectomy was used as reference standard in 3 studies [18,20,27], TRUS in 4 studies [15,16,21,28], and MRI-GB in 7 studies [17,19,22–26]. In almost all included studies, there were 2 radiologists to interpretation the MRI image, with experience of 1–16 years. In one study [21], however, the PI-RADS scoring was performed by one reader (with 10 years of experience). In 8 studies, PCa were evaluated according to zonal anatomy divided as peripheral zone and transitional zone [16,17,22,24,20–28], however, 3 studies [15,21,23] did not differentiate PZ and TZ, and there were respectively one study only assessed PZ and TZ [19,20]. There were 5 studies [16,18–21] assessed CS cancer and 6 studies [22–26,28] assessed any type of PCa, for the remaining 3 studies [15,17,27], both of them were evaluated. Almost all studies reported that readers were blinded to clinical stage and histopathological findings, however, in one study [26], the radiologists were aware of the patients' PSA levels, and 2 studies [21,25] did not explicitly describe the blinding.

3.3. Quality assessment

In summary, quality assessment for studies was not substantially high, primarily attributed to patient selection domain, details are shown in Fig. 2. As for patient selection domain, 10 studies [16,17,20–22,24–28] were regarded as high risk of bias because of a retrospective in nature. Two studies were assigned high concern for applicability, as for these 2 studies [19,20] only analyzed peripheral zone or transition zone. In the index test domain, 4 studies

Table 1
Demographic characteristics.

Study	Year	Country	Period	Study Design	Patients Number	Age Mean ± SD/Range	PSA Mean ± SD/Range	Gleason Scores median/range	Reference	Clinical Significant
Auer	2016	Austria	NR	Retrospective	50	63 ± 8	7.3 ± 4.3	7 (3 + 4)/6-10	RP	GS ≥ 4 + 3
Feng	2016	China	2013.6/2015.7	Retrospective	401	64.4 ± 9/34-88	10.7/0.2–1763	7 (4 + 3)/6–8	MRI-TRUS	NR
Bao	2017	China	2015.9/2016.2	Prospective	30	67 ± 8.42	17.38 ± 17.92	7/6-9	MRI-TRUS	GS ≥ 7
Krishna	2017	Canada	2012.4/2015.5	Retrospective	52	61.7 ± 5.7	5.9 ± 2.9	7 (3 + 4)	RP	GS ≥ 7
Wang	2018	China	2015.9/2013.7	Prospective	77	73.20 ± 6.78	26.34 ± 19.90	≥6	MRI-TRUS	GS ≥ 7
Kasel-Seibert	2016	Germany	2013.7/2015.3	Retrospective	82	65/48-81[a]	13/1-111[a]	7 (3 + 4)/6-9	MRI-GB	GS ≥ 7/ EPE
Tan	2017	USA	2013.5/2013/12	Retrospective	106	66.5/43-79[b]	7.9/5.6-10.6[b]	7 (3 + 4)/6-10	MRI-GB	GS ≥ 3 + 4
Hoffmann	2017	Austria	2014.4/2015.8	Prospective	58	66.7/47-81	24.7/9.25 ± 4.73	≥6	TRUS	EPE
Tewes	2016	Germany	2012.12/2014.12	Retrospective	54	69.6 ± 9.6	8.7 ± 4.9	6/6-9	MRI-GB	GS ≥ 3 + 4
Polanec	2016	Austria	2011.6/2015.9	Retrospective	65	65.3/62.3-87.4	10.78/4.2-74.5	7 (4 + 3)/6-9	MRI-GB	NR
Yun	2018	China	2012.4/2016.10	Retrospective	77	73.45 ± 6.59	17.23 ± 13.04	6-10	TRUS	GS ≥ 7
Visschere	2016	Belgium	2011.5/2014.12	Retrospective	245	66/44-85	9.0/1.4-935.5	7 (3 + 4)/ ≥ 6	TRUS	3 + 4/ ≥ 0.5cc/ EPE
Ju	2017	China	2009.12/2015.12	Retrospective	175	65.4 ± 8.5	4.5-100	≥6	TRUS	GS ≥ 7
Matsuoka	2017	Japan	2007.8/2015.3	Prospective	210	67/50–81	7.0/2.9–30.0	5-10	RP	EPE

Notes: EPE, extraprostatic extension; GS, Gleason score; MRI-GB, magnetic resonance imaging-guided biopsy; MRI–TRUS, fusion of magnetic resonance imaging and transrectal ultrasound images; NR, not report; PSA, prostate-specific antigen; RP, radical prostatectomy; SD, standard deviation; TRUS, systematic transrectal ultrasound-guided biopsy.

[a] median.

[b] interquartile range.

Table 2
Study characteristics.

Study	Field Strength	Coil	Cutoff v1/v2	Localization	Analysis	Readers	Experience (years)	κ values v1/v2	Blinding	Protocols	Outcome assessed
Auer	3.0 T	Body	4/4	PZ/TZ	Lesion	2	> 5/2	NR	Yes	T2/DWI (ADC)/DCE	Both
Feng	3.0 T	Body	4/4	PZ/TZ	Lesion	2	4/5	NR	Yes	T1/T2/DWI/DCE	Any PCa
Bao	3.0 T	None	10/4	Whole	Patient	2	> 6	NR	Yes	T2/DWI/DCE	Any PCa
Tewes	3.0 T	Body	4/3[a]	PZ/TZ	Lesion	2	5/2	0.62/0.64 0.54/0.53[b]	Yes	T2/DWI (ADC)/DCE	Any PCa
Hoffmann	1.5 T	None	4	Whole	Patient	2	NR	0.56/0.51	Yes	T2/DWI/DCE	Both
Tan	3.0 T	None	3	PZ/TZ	Lesion	2	10/16	NR	NR	T2/DWI/DCE	Any PCa
Polanec	3.0 T	Body	10/4	PZ/TZ	Lesion	2	> 150 cases/year	0.81/0.71	Yes	T2/DWI/DCE	Any PCa
Kasel-Seibert	1.5 T	PPA + ERC	4/4	PZ/TZ	Lesion	2	NR	0.55/0.68	Yes	T2/DWI (ADC)/DCE	Both
Krishna	3.0 T	PPA	10/4	PZ	Lesion	2	11/16	0.52/0.41	Yes[c]	T2/DWI/DCE/MRS	CS PCa
Yun	3.0 T	PPA	3/3	PZ/TZ	Lesion	2	2/5	0.78/0.88	Yes[c]	T2/DWI (ADC)/DCE	Any PCa
De Visschere	3.0 T	None	3	Whole	Patient	1	10	NR	NR	T2/DWI/DCE/MRS	CS PCa
Ju	1.5 T	PPA	11/4	PZ/TZ	Lesion	2	1/6	0.39/0.53	Yes	T2/DWI/DCE	CS PCa
Wang	3.0 T	Body	11/4	TZ	Lesion	2	2/8	NR	Yes	T2/DWI (ADC)/DCE	CS PCa
Matsuoka	1.5 T	Body	3/3	Whole	Patient	2	5/10	0.51/0.78	Yes	T2/DWI (ADC)/DCE	CS PCa

Notes: ADC, apparent diffusion coefficient; CS, clinical significant; DCE, dynamic contrast-enhanced magnetic resonance imaging; DWI, diffusion-weighted imaging; ERC, endorectal; MRS, magnetic resonance spectroscopy imaging; NR, not report; PCa, prostate cancer; PPA, pelvic phase array; PZ, peripheral zone; T2, T2-weighted imaging; TZ, transitional zone.

[a] cutoff = 3 for peripheral zone and cutoff = 4 for transitional zone.

[b] For clinically prostate cancer.

[c] Blinded but aware that patients had biopsy-proven PCa.

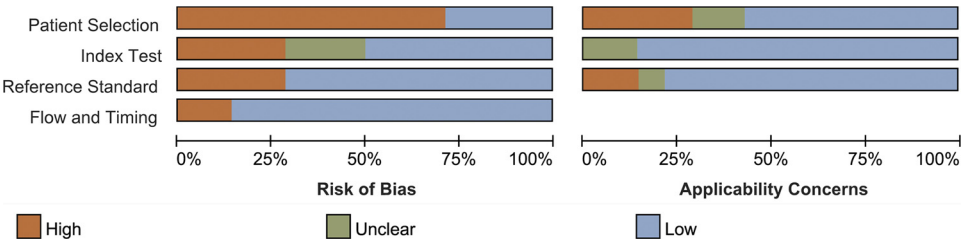

Fig. 2. Grouped bar charts show the risk of bias and concerns for applicability of included studies.

[16,22,24,27] were assessed as high risk of bias because of cutoff values for PCa were not established before interpretation. There was no study assigned high concern regarding applicability in this domain. With regard to reference standard domain, 4 studies [15,16,21,28] only used TRUS as reference standard was deemed had high risk of bias, whereas the other 10 studies including 3 [18,20,27] used radical prostatectomy, 3 [19,22,23] used MRI-TRUS, and 4 [17,24–26] used MRGB were regarded had low risk of bias. Finally, for flow and timing domain, two studies [15,21] had a high risk of bias owing to patients received different reference standard.

3.4. Synthesis of included studies

For PI-RADS v1, the sensitivity ranged from 69% to 99%, with specificity from 6% to 98%, and the pooled estimates of sensitivity and specificity were 85.37% (95% CI 79.29%–89.89%) and 77.26% (95% CI 57.29%–89.58%), respectively. On the other hand, for PI-RADS v2, the sensitivity ranged from 58% to 100%, with specificity of 14%–100%, pooled estimates of sensitivity and specificity were 87.79% (95% CI 79.84%–92.88%) and 76.73% (95% CI 58.69%–88.44%), respectively. Detailed diagnostic performances for two PI-RADS scoring systems are presented in Figs. 3 and 4.

Overall, it seemed that PI-RADS v2 performed slightly better in sensitivity, however, the difference was not substantial (P = 0.57), moreover, the improvement was at the expense of minor diminished specificity. The Deeks' funnel plot with the P = 0.33 for slop coefficient demonstrated that the likelihood of publication bias was low.

3.5. Comparison of diagnostic accuracy between two PI-RADS versions

Multiple subgroup analyses and meta-regression were performed to evaluate various clinical settings (Fig. 5 and Supplementary Table 1). For any PCa, the pooled sensitivity and specificity for PI-RADS v1 were 86.96% (95% CI 80.4%–91.55%) and 71.86% (95% CI 41.5%–90.19%), whereas PI-RADS v2 showed a higher sensitivity of 91.26% (95% CI 79.64%–96.54%), with minute declined specificity of 71.71% (95% CI 47.05%–87.85%). However, the discrepancy both for sensitivity and specificity were not statistically significant (P = 0.7 and P = 0.96). With respect to CS PCa, PI-RADS v1 and PI-RADS v2 had nearly identical sensitivity of 85.77% (75.2%–92.3%) and 85.25% (95% CI 77.7%–90.56%), respectively, while showed a little difference in specificity, with 81.85% (95% CI 60.19%–93.07%) and 79.67% (95% CI 58.46%–91.6%), but not substantial (P = 0.85). For PZ, our analysis suggested that PI-RADS v2 performed better than PI-RADS v1, irrespective of sensitivity (88.77% vs. 91.79%) and specificity (85.57% vs. 88.78%). However, neither sensitivity nor specificity reached substantially significant different, for both P = 0.58. For TZ, the pooled sensitivity for PI-RADS v1 was 80.59% (95% CI 73.9%–85.89%), with specificity of 77.62% (95% CI 68.77%–84.52%), while for PI-RADS v2, the pooled sensitivity and specificity were 90.1% (95% CI 79.88%–95.43%) and 84.03% (95% CI 75.33%–90.07%), respectively. Although it seemed that the differences on TZ were more dramatically, neither sensitivity nor specificity had statistically significantly different (P = 0.17 and P = 0.42, respectively).

Regarding magnet field strength, our analyses showed that regardless of 1.5 T or 3.0 T scanners, there was no significant difference between PI-RADS v1 and PI-RADS v2. For 4 studies [15–18] used the 1.5 T scanner, pooled sensitivity for PI-RADS v1 and PI-RADS v2 were 76.46% (95% CI 68.06%–83.2%) and 83.66% (95% CI 72.63%–90.81%), respectively, with a P value of 0.29. By comparison, the pooled specificity were 58.31% (95% CI 32.07%–80.56%) and 60.5% (95% CI 34.49%–81.67%), respectively, with a P value of 0.91. For 10 studies [19–28] used 3.0 T scanner, the pooled sensitivity were 88.32% (95% CI 81.9%–92.67%) for PI-RADS v1 and 89.71% (95% CI 78.46%–95.42%) for PI-RADS v2, with trivial differences (P = 0.78). Similarly, a P value of 0.92 indicated that the difference in specificity (83.6%, 95% CI 61.41%–94.23% vs. 82.45%, 95% CI 61.92%–93.14%) was also insignificant. Fig. 6 shows the SROC curves for two PI-RADS guidelines.

3.6. Heterogeneity exploration

The Higgins I^2 statistics indicated substantial heterogeneity for sensitivity (I^2 = 80.25% for PI-RADS v1 and I^2 = 83.1% for PI-RADS v2, respectively), and considerable heterogeneity with respect to specificity (I^2 = 96.17% for PI-RADS v1 and I^2 = 94.4% for PI-RADS v2, respectively), we then made meta-regression in terms of several potential variables. For PI-RADS v1, our analyses showed that the magnetic field strength (1.5 T vs. 3.0 T) contributed to the substantial heterogeneity of sensitivity (76% vs. 88%, P = 0.03), and whether use of coil (any coil vs. none) was a substantial factor for specificity (86% vs. 45%, P = 0.03). For all of the other variables, including the cutoff value, outcome assessment, type of analysis, study design, and reference standard, were not significant factors. With regard to PI-RADS v2, our analyses found that none of these variables was the significant factor influencing the heterogeneity. More details are presented in Supplementary Table 2.

4. Discussion

In this systematic review, we performed a detailed comparison between PI-RADS v1 and PI-RADS v2 on the basis of 14 studies. According to our analyses, there was no significant difference between these two scoring systems, with 85.37% vs. 87.79% for sensitivity and 77.26% vs. 76.73% for specificity. Multiple subgroup analyses and meta-regression based on magnetic field strength (1.5 T vs. 3.0 T), outcome assessment (any PCa vs. CS PCa) and zonal anatomy (PZ vs. TZ), cutoff value, suggested consistent results. The PI-RADS v2 showed slightly higher sensitivity when compared to PI-RADS v1, but at the expense of minor diminished specificity, therefore the overall diagnostic performance for PI-RADS v1 and PI-RADS v2 was comparable. One of the most important revision in PI-RADS v2 is that imaging protocols are weighted in accordance with zonal anatomy (DWI for the PZ and T2WI for the TZ), in order to improve the diagnostic performance. Other primary changes from PI-RADS v1 to PIRADSv2 as follows: 1) using an overall 5-point scale to combine findings throughout all MRI protocols, 2) for overall score, DCE sequence is regarded secondary to DWI and T2WI, because many studies questioned whether DCE plays a role in the

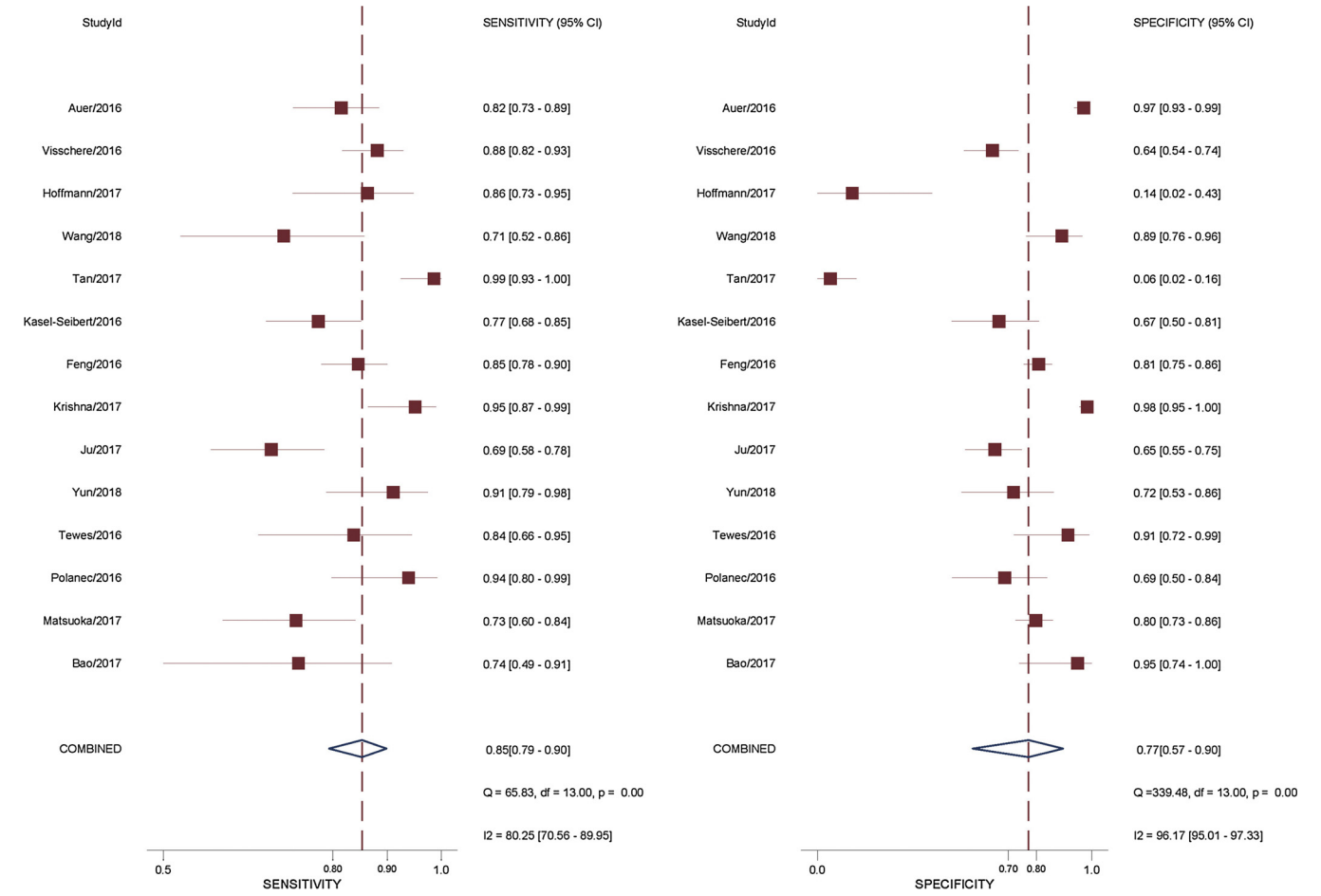

Fig. 3. Coupled forest plots of pooled sensitivity and specificity for PI-RADS v1. PI-RADS, Prostate Imaging Reporting and Data System; CI, confidence interval.

detection of PCa [29].

We performed subgroup analyses and meta-regression to investigate the influence of variant clinical settings. Based on several studies analyzed according to zonal anatomy, it seemed that the revision did not bring about significant improvement in diagnostic performance irrespective of PZ or TZ. Feng et al found that PI-RADS v2 had significantly higher accuracy than v1 regarding TZ (96.36% vs. 76.36%, P = 0.003) [22], whereas 3 other studies found that even though PI-RADS v2 showed a better performance the difference not reached substantial in comparison with PI-RADS v1. Likewise, for lesions locate in PZ, most studies included suggested that there was no significant difference in diagnostic performance between these two scoring systems, though several studies reported that PI-RADS v2 had better diagnostic performance in PZ [24–28]. It is interesting that 2 studies found that for PZ, PI-RADS v1 performed even better than PI-RADS v2 [20,26]. The other clinical settings such as magnetic field strength (1.5 T and 3.0 T), outcome assessment (any PCa and CS PCa), and cutoff value showed there was no significant difference between these two scoring systems.

Hamoen et al made a diagnostic meta-analysis on PI-RADS v1 based on 14 studies, then obtained the pooled sensitivity of 78% and specificity of 79% [8]. A more recently diagnostic meta-analysis which included 13 studies for PI-RADS v2, showed a pooled sensitivity of 85%, with a specificity of 71% [30]. In another study including 21 original articles conducted by Woo et al, the pooled sensitivity and specificity were 89% and 73%. Moreover, they compared the diagnostic accuracy between PI-RADS v1 and PI-RADS v2 based on 6 articles involving head to head comparison, then obtained pooled sensitivity of 88% vs. 95% (P = 0.04), suggested that PI-RADS v2 had a substantial advantage over PI-RADS v1, but the pooled specificity of 75% vs. 73% (P = 0.90) suggested a minor decline in PI-RADS v2 [9]. In our study, however, almost all studies included reported that there was no significant difference in diagnostic performance between these two scoring systems. Moreover, it is interesting that Auer et al [27] reported that for overall, PI-RADS v1 behaved substantial better than PI-RADS v2 (96% vs. 90%, P < 0.001 for the area under the curve, AUC), especially for PZ (97% vs. 92%, P < 0.001 for AUC).

As an important indication for scoring systems, the inter-reader agreement, which assessed using Cohen's kappa (κ) statistics were reported by more than half of studies. For PI-RADS v1, the κ values ranged from 0.39 to 0.81, whereas it varied from 0.41 − 0.88 for PI-RADS v2, a slightly higher. 4 studies reported that the inter-observer agreement in PI-RADS v2 was better [16–18,28], whereas there were 3 studies [15,20,26] showed that PI-RADS v1 had higher κ values. In the study of Tewes et al [24], for any PCa the PI-RADS v2 had a better inter-reader agreement (0.62 versus 0.64), while for CS PCa the situation was opposite (0.54 versus 0.5). A study [32] focused on the comparison of inter-reader agreement between PI-RADS v1 and v2 suggested that the κ values for these two guidelines nearly identical (for both κ = 0.71). However, Kasel-Seibert et al [17] demonstrated that diagnostic performance was better for the inexperienced reader as well when working with PI-RADS v2 supports its improved reproducibility even for less experienced physicians.

Apart from the two PI-RADS guidelines, Bao et al [23] made a comparison between these two scoring systems with the 5-point Likert scale, found that the Likert scale had identical sensitivity with PI-RADS v1 but better than PI-RADS v2, however, it showed the lowest specificity among these 3 scoring systems. Another study conducted by Rosenkrantz et al [31] systematically compared the Likert scale with PI-

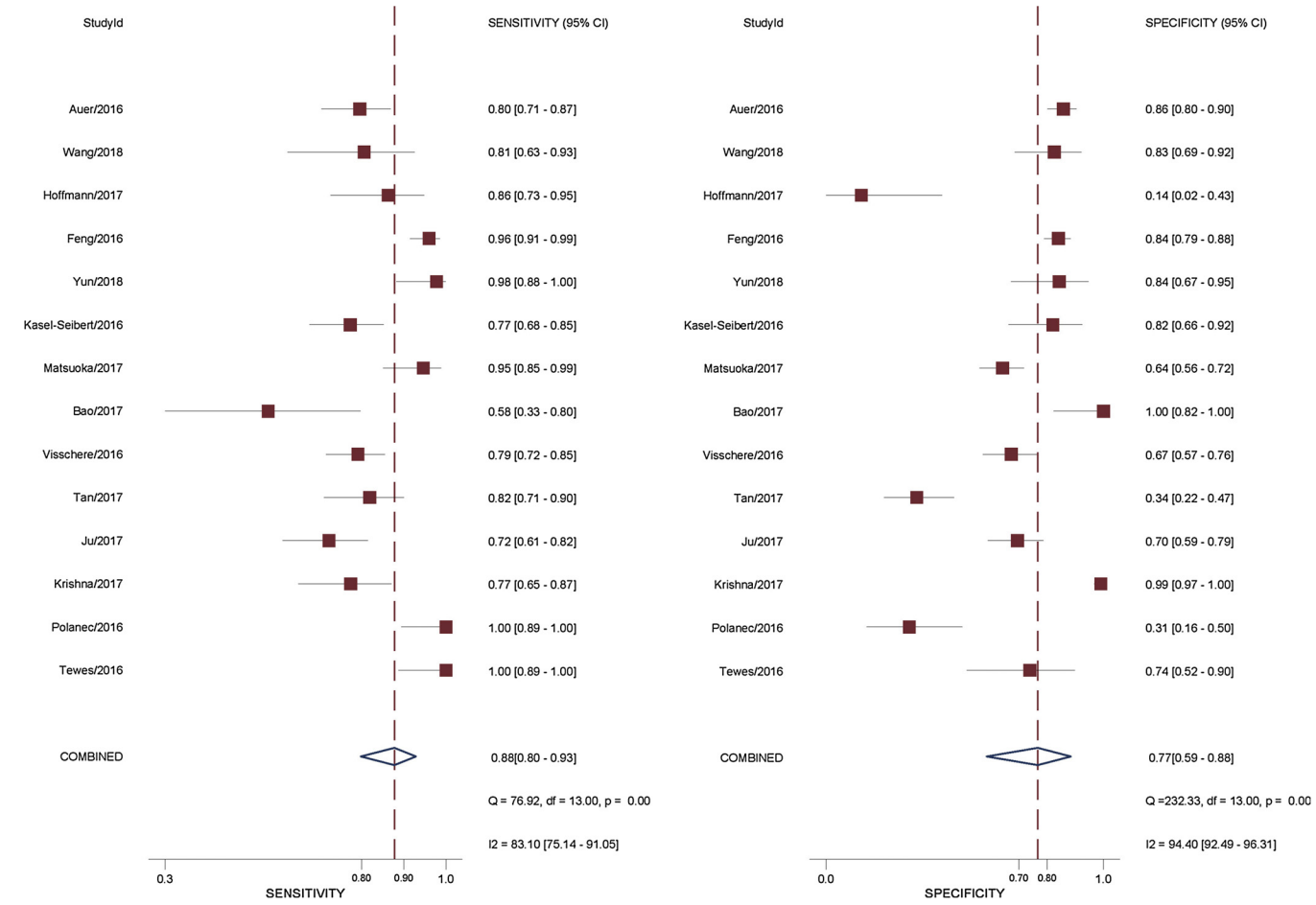

Fig. 4. Coupled forest plots of pooled sensitivity and specificity for PI-RADS v2. PI-RADS, Prostate Imaging Reporting and Data System; CI, confidence interval.

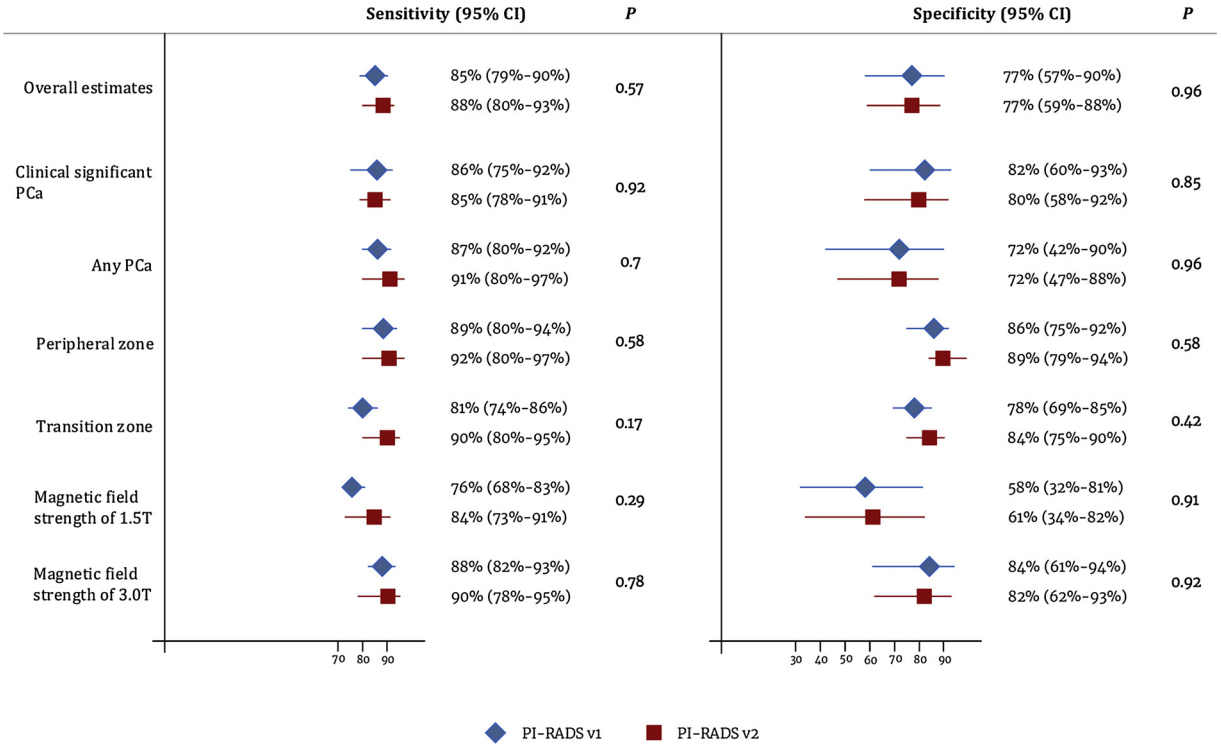

Fig. 5. Pooled estimates with 95% CI for two PI-RADS versions. PI-RADS, Prostate Imaging Reporting and Data System; CI, confidence interval.

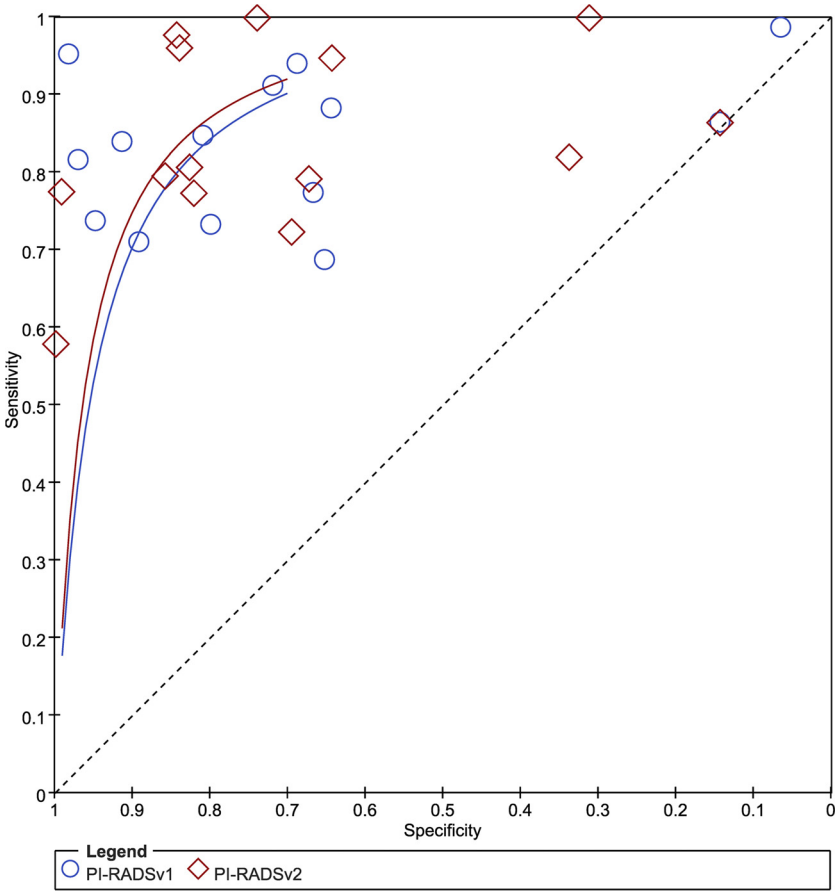

Fig. 6. Comparison of SROC curves between two PI-RADS versions. PI-RADS, Prostate Imaging Reporting and Data System.

RADS v1, the diagnostic accuracy of the Likert scale was substantially higher, in 2 of 3 experienced radiologists. However, a shortage of studies renders it difficult to draw a comparison between the Likert scale and PI-RADS guidelines.

Up to date, 3 meta-analyses had been published to evaluate the diagnostic performance of PI-RADS for detection of PCa, including one for PI-RADS v1 [8] and two for PI-RADS v2 [9,30]. To our knowledge, however, there was no one on the subject of comparison between these two PI-RADS versions based on head to head comparison studies. Generally, findings from this systematic review suggested that either PI-RADS v1 or PI-RADS v2 has high accuracy and favorable inter-reader agreement, which was consistent with 3 earlier meta-analyses, and we also found that there was no significant difference in diagnostic performance between them, irrespective of sensitivity and specificity. This is different from a recently published article [9] based on 6 head to head comparison studies, in which the summary estimated suggested that PI-RADS v2 had a substantial advantage over PI-RADS v1 (88% vs. 95%, P = 0.04,), the pooled specificity of 75% vs. 73% (P = 0.90) suggested a minor decline in PI-RADS v2.

Some limitations existed in our systematic review. Nearly all studies included were retrospective in nature, leading to a high risk of bias for patient selection domain. In view of deficient prospective studies, it is not practicable to perform analysis using only these studies, moreover, the derived results would not represent the existing literature. Another limitation is that there is considerable heterogeneity across pooled analysis, which had an effect on the applicability of pooled estimates of sensitivity and specificity. Meta-regression was conducted to investigate the heterogeneity, and the results demonstrated that both magnetic field strength and use of coil were the substantial contributors. However, we could not completely explain the heterogeneity, for insufficient information of all study characteristics. Finally, several studies did not differentiate PZ and TZ when performing analysis or merely analyzed one of them, which may affect the overall evaluations of two PI-RADS scoring systems.

5. Conclusion

Either PI-RADS v1 or PI-RADS v2 shows favorable diagnostic performance for detection of the PCa, and there is no substantial difference between these two scoring systems. It appears that PI-RADS v2 has slightly higher sensitivity, but at the expense of minor decreased specificity.

Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of interests

None of the authors have any actual or potential conflicts of interest to declare in relation to this article.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.ejrad.2019.03.016.

References

[1] R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, 2016, CA Cancer J. Clin. 56 (106) (2015).

[2] T.J. Caverly, R.A. Hayward, E. Reamer, et al., Presentation of benefits and harms in US Cancer screening and prevention guidelines: systematic review, J. Natl. Cancer Inst. 108 (2016), https://doi.org/10.1093/jnci/djv436 djv436.
[3] N.E. Abraham, N. Mendhiratta, S.S. Taneja, Patterns of repeat prostate biopsy in contemporary clinical practice, J. Urol. 193 (2015) 1178–1184, https://doi.org/10.1016/j.juro.2014.10.084.
[4] T. Tamada, T. Sone, H. Higashi, et al., Prostate Cancer detection in patients with total serum prostate-specific antigen levels of 4-10 ng/mL: diagnostic efficacy of diffusion-weighted imaging, dynamic contrast-enhanced MRI, and T2-Weighted imaging, AJR Am. J. Roentgenol. 197 (2011) 664–670, https://doi.org/10.2214/AJR.10.5923.
[5] J.C. Vilanova, C. Barceló-Vidal, J. Comet, et al., Usefulness of prebiopsy multifunctional and morphologic MRI combined with free-to-Total prostate-specific antigen ratio in the detection of prostate Cancer, Am. J. Roentgenol. 196 (2011) W715–W722, https://doi.org/10.2214/AJR.10.5700.
[6] N.B. Delongchamps, M. Rouanne, T. Flam, et al., Multiparametric magnetic resonance imaging for the detection and localization of prostate cancer: combination of T2-weighted, dynamic contrast-enhanced and diffusion-weighted imaging, BJU Int. 107 (2011) 1411–1418, https://doi.org/10.1111/j.1464-410X.2010.09808.x.
[7] L.D. Mrcs, H.U.A. Mrcs, C.A. Frcr, et al., Scoring systems used for the interpretation and reporting of multiparametric MRI for prostate cancer detection, localization, and characterization: could standardization lead to improved utilization of imaging within the diagnostic pathway? J. Magn. Reson. Imaging 37 (2013) 48–58.
[8] E.H.J. Hamoen, Rooij M. de, J.A. Witjes, et al., 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. 67 (2015) 1112–1121, https://doi.org/10.1016/j.eururo.2014.10.033.
[9] S. Woo, C.H. Suh, S.Y. Kim, et al., 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. 72 (2017) 177–188, https://doi.org/10.1016/j.eururo.2017.01.042.
[10] A. Liberati, D.G. Altman, J. Tetzlaff, et al., The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration, Epidemiol. Biostat. Public Health 6 (2009) e1–e34.
[11] P.F. Whiting, A.W.S. Rutjes, M.E. Westwood, et al., QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies, Ann. Intern. Med. 155 (2011) 529–536, https://doi.org/10.7326/0003-4819-155-8-201110180-00009.
[12] J.B. Reitsma, A.S. Glas, A.W.S. Rutjes, et al., Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews, J. Clin. Epidemiol. 58 (2005) 982–990, https://doi.org/10.1016/j.jclinepi.2005.02.022.
[13] H. Chu, S.R. Cole, Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach, J. Clin. Epidemiol. 59 (2006) 1331–1332, https://doi.org/10.1016/j.jclinepi.2006.06.011.
[14] C.M. Rutter, C.A. Gatsonis, A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations, Stat. Med. 20 (2001) 2865–2884.
[15] R. Hoffmann, C. Logan, M. O'Callaghan, et al., Does the prostate imaging-reporting and data system (PI-RADS) version 2 improve accuracy in reporting anterior lesions on multiparametric magnetic resonance imaging (mpMRI)? Int. Urol. Nephrol. 50 (2018) 13–19, https://doi.org/10.1007/s11255-017-1753-1.
[16] Ju, Comparison of PI-RADS v1 and PI-RADS v2 in diagnosing Clinically Significant Prostate Cancer with 1.5 T Mp-MRI, [master's thesis] Suzhou, Soochow University, 2019.
[17] M. Kasel-Seibert, T. Lehmann, R. Aschenbach, et al., Assessment of PI-RADS v2 for the detection of prostate Cancer, Eur. J. Radiol. 85 (2016) 726, https://doi.org/10.1016/j.ejrad.2016.01.011.
[18] Y. Matsuoka, J. Ishioka, H. Tanaka, et al., Impact of the prostate imaging reporting and data system, version 2, on MRI diagnosis for extracapsular extension of prostate Cancer, Am. J. Roentgenol. 209 (2017) W76–W84, https://doi.org/10.2214/AJR.16.17163.
[19] X. Wang, J. Bao, X. Ping, et al., The diagnostic value of PI-RADS V1 and V2 using multiparametric MRI in transition zone prostate clinical cancer, Oncol. Lett. 16 (2018) 3201–3206, https://doi.org/10.3892/ol.2018.9038.
[20] S. Krishna, M. McInnes, C. Lim, et al., Comparison of prostate imaging reporting and data system versions 1 and 2 for the detection of peripheral zone gleason score 3 + 4 = 7 cancers, AJR Am. J. Roentgenol. 209 (2017) W365–W373, https://doi.org/10.2214/AJR.17.17964.
[21] P. De Visschere, E. Pattyn, P. Ost, et al., Comparison of the Prostate Imaging Reporting and Data System (PI-RADS) version 1 and 2 in a cohort of 245 patients with histopathological reference and long-term follow-up, J. Belg. Soc. Radiol. 100 (2016), https://doi.org/10.5334/jbr-btr.1147.
[22] Z.Y. Feng, L. Wang, X.D. Min, et al., Prostate Cancer detection with multiparametric magnetic resonance imaging: prostate imaging reporting and data system version 1 versus version 2, Chin. Med. J. 129 (2016) 2451–2459, https://doi.org/10.4103/0366-6999.191771.
[23] J. Bao, X. Wang, M. Zhu, et al., Evaluation of the PI-RADS scoring system for detection of prostate cancer with targeted MRI-TRUS fusion-guided biopsy, J. Pract. Radiol. 33 (2017) 1217–1221.
[24] S. Tewes, N. Mokov, D. Hartung, et al., Standardized reporting of prostate MRI: comparison of the prostate imaging reporting and data system (PI-RADS) version 1 and version 2, PLoS One 11 (2016) e0162879, , https://doi.org/10.1371/journal.pone.0162879.
[25] N. Tan, W.-C. Lin, P. Khoshnoodi, et al., In-bore 3-T MR-guided transrectal targeted prostate biopsy: prostate imaging reporting and data system version 2-based diagnostic performance for detection of prostate cancer, Radiology 283 (2017) 130–139, https://doi.org/10.1148/radiol.2016152827.
[26] S. Polanec, T.H. Helbich, H. Bickel, et al., Head-to-head comparison of PI-RADS v2 and PI-RADS v1, Eur. J. Radiol. 85 (2016) 1125–1131, https://doi.org/10.1016/j.ejrad.2016.03.025.
[27] T. Auer, M. Edlinger, J. Bektic, et al., Performance of PI-RADS version 1 versus version 2 regarding the relation with histopathological results, World J. Urol. 35 (2017) 687–693, https://doi.org/10.1007/s00345-016-1920-5.
[28] Y. Yun, X. Chen, W. Xia, et al., Comparison of prostate imaging reporting and data system version 2 and version 1, J. Clin. Interv. Radiol. Isvir 37 (2018) 1136–1141.
[29] B.G. Hansford, Y. Peng, Y. Jiang, et al., Dynamic contrast-enhanced MR imaging curve-type analysis: is it helpful in the differentiation of prostate cancer from healthy peripheral zone? Radiology 275 (2015) 448–457, https://doi.org/10.1148/radiol.14140847.
[30] L. Zhang, M. Tang, S. Chen, et al., A meta-analysis of use of Prostate Imaging Reporting and Data System Version 2 (PI-RADS V2) with multiparametric MR imaging for the detection of prostate cancer, Eur. Radiol. 27 (2017) 5204–5214, https://doi.org/10.1007/s00330-017-4843-7.
[31] W. Rosenfeld, J. Conry, L. Lagae, et al., Efficacy and safety of perampanel in adolescent patients with drug-resistant partial seizures in three double-blind, placebo-controlled, phase III randomized clinical studies and a combined extension study, Eur. J. Paediatr. Neurol. 19 (2015) 435–445, https://doi.org/10.1016/j.ejpn.2015.02.008.
[32] A.S. Becker, A. Cornelius, C.S. Reiner, et al., Direct comparison of PI-RADS version 2 and version 1 regarding interreader agreement and diagnostic accuracy for the detection of clinically significant prostate cancer, Eur. J. Radiol. 94 (2017) 58–63, https://doi.org/10.1016/j.ejrad.2017.07.016.