Diagnostic Performance of Prostate Imaging Reporting and Data System Version 2 for Detection of Prostate Cancer: A Systematic Review and Diagnostic Meta-analysis

Sungmin Woo¹, Chong Hyun Suh¹,²,³, Sang Youn Kim¹,*, Jeong Yeon Cho¹,², Seung Hyup Kim¹,²

¹Department of Radiology, Seoul National University College of Medicine, Seoul, Korea; ²Department of Radiology and Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Republic of Korea; ³Department of Radiology, Namwon Medical Center, Jeollabuk-do, Republic of Korea; ⁴Institute of Radiation Medicine and Kidney Research Institute, Seoul National University Medical Research Center, Seoul, Korea

Abstract

Context: In 2015, the updated Prostate Imaging Reporting and Data System version 2 (PI-RADS v2) for the detection of prostate cancer (PCa) was established. Since then, several studies assessing the value of PI-RADS v2 have been published.

Objective: To review the diagnostic performance of PI-RADS v2 for the detection of PCa.

Evidence acquisition: MEDLINE and EMBASE databases were searched up to December 7, 2016. We included diagnostic accuracy studies that used PI-RADS v2 for PCa detection, using prostatectomy or biopsy as the reference standard. The methodological quality was assessed by two independent reviewers using the Quality Assessment of Diagnostic Accuracy Studies-2 tool. Sensitivity and specificity of all studies were calculated. Results were pooled and plotted in a hierarchical summary receiver operating characteristic plot with further exploration using meta-regression and multiple subgroup analyses. Head-to-head comparison between PI-RADS v1 and PI-RADS v2 was performed for available studies.

Evidence synthesis: Twenty-one studies (3857 patients) were included. The pooled sensitivity was 0.89 (95% confidence interval [CI] 0.86–0.92) with specificity of 0.73 (95% CI 0.60–0.83) for PCa detection. Proportion of patients with PCa, magnetic field strength, and reference standard were significant factors affecting heterogeneity (p < 0.01). Multiple subgroup analyses showed consistent results. In six studies performing head-to-head comparison, PI-RADS v2 demonstrated higher pooled sensitivity of 0.95 (95% CI 0.85–0.98) compared with 0.88 (95% CI 0.80–0.93) for PI-RADS v1 (p = 0.04). However, the pooled specificity was not significantly different (0.73 [95% CI 0.47–0.89] vs 0.75 [95% CI 0.36–0.94], respectively; p = 0.90).

Conclusions: PI-RADS v2 shows good performance for the detection of PCa. PI-RADS v2 has higher pooled sensitivity than PI-RADS v1 without significantly different specificity.

Patient summary: We reviewed all previous studies using Prostate Imaging Reporting and Data System version 2 (PI-RADS v2) for prostate cancer detection. We found that the updated PI-RADS v2 shows significant improvement compared with the original PI-RADS v1.

© 2017 European Association of Urology. Published by Elsevier B.V. All rights reserved.

¹ These authors contributed equally.
* Corresponding author. Department of Radiology, Seoul National University Hospital, 101 Daehak-ro, Jongno-gu, Seoul 110-744, Korea. Tel.: +82 2 2072 4897; Fax: +82 2 743 6385.
E-mail address: iwishluv@empas.com (S.Y. Kim).
1. **Introduction**

Prostate cancer (PCa) is the second leading cause of cancer-related mortality in Western men [1]. With the recent technological advancements and growing availability, multiparametric magnetic resonance imaging (mpMRI) currently is increasingly being used for guiding several aspects of PCa management, including detection, staging, and treatment planning [2]. Despite abundant evidence in the literature reporting high accuracy of mpMRI for PCa diagnosis, widespread acceptance has been hampered by several factors including difficulty of interpretation, lack of standardized criteria for interpretation (ie, use of Likert scales based on the radiologist’s subjective level of suspicion for PCa), and resulting substantial inter-reader variability [3,4].

To bring standardization to the evaluation and reporting of mpMRI of the prostate, the European Society of Urogenital Radiology (ESUR) published a guideline termed Prostate Imaging Reporting and Data System (PI-RADS) in 2012 [5]. PI-RADS was generated based on expert consensus and provides a detailed scoring system for each MRI sequence (T2-weighted imaging [T2WI], diffusion-weighted imaging [DWI], dynamic contrast-enhanced MRI [DCE-MRI], and MR spectroscopy) for the presence of clinically significant PCa (csPCa). Several investigators have validated the accuracy and reproducibility of PI-RADS, and a recent meta-analysis reported pooled sensitivity and specificity of 0.78 and 0.79, respectively [6]. However, as there was no guideline for the generation of an overall score, different research groups utilized various measures for this purpose—some used a sum of the scores from each sequence (ranging from 3 to 15), whereas others used an overall score of 1–5 [7,8]. Furthermore, emerging data questioned the value of PI-RADS, and a recent meta-analysis reported sensitivity and specificity at specified cutoff values for evaluating the diagnostic performance of PI-RADSv2; and (5) studies had to be original articles.

To address these issues, the ESUR and American College of Radiology recently released the updated PI-RADS version 2 (PI-RADSv2) [11]. The main changes from PI-RADSv1 to PI-RADSv2 are the following: (1) introduction of dominant sequences according to zonal anatomy (DWI for the TZ and T2WI for the TZ), (2) limited contribution of DCE-MRI data as merely presence and absence of early focal enhancement, and (3) generation of an overall score (1–5) integrating findings across all MRI sequences.

Since then, several studies assessing the value of PI-RADSv2 have been published. However, the diagnostic performance of this new scoring system has not been evaluated systematically. Therefore, the purpose of our study was to assess the diagnostic performance of PI-RADSv2 for the detection of PCa. In addition, we aimed to compare the diagnostic performance of PI-RADSv1 and PI-RADSv2 in studies available for head-to-head comparison.

2. **Evidence acquisition**

This meta-analysis was performed and written according to the preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines [12].

2.1. **Literature search**

A computerized search of MEDLINE and EMBASE up to December 7, 2016, was performed in order to identify studies evaluating the diagnostic performance of PI-RADSv2 for the detection of PCa. The search query combined synonyms for PCa, MRI, and PI-RADS as follows: (prostate cancer OR prostatic cancer OR prostate neoplasm OR prostate neoplasm OR prostate tumor OR prostatic tumor OR prostate carcinoma OR prostatic carcinoma OR PCa) AND (magnetic resonance imaging OR MRI OR MR) AND (prostate imaging reporting and data system OR pi-rads OR pi rads OR pi rads OR pirads). Bibliographies of identified articles were also screened to expand the scope of search. Our search was limited to publications in English.

2.2. **Study selection**

2.2.1. **Inclusion criteria**

Studies were included if they satisfied all the following requirements according to the PICOS criteria [12]: (1) included patients with suspected or diagnosed PCa; (2) for index test, mpMRI of the prostate including all required sequences of T2WI, DWI, and DCE-MRI was performed and assessed with a PI-RADSv2 scoring system; (3) for comparison, a reference standard based on the histopathological examination of radical prostatectomy or biopsy was used; (4) results were reported in sufficient detail for the reconstruction of 2 × 2 tables and determination of sensitivity and specificity at specified cutoff values for evaluating the diagnostic performance of PI-RADSv2; and (5) studies had to be original articles.

2.2.2. **Exclusion criteria**

Studies were excluded if any of the following criteria were met: (1) studies involving <10 patients; (2) review articles, guidelines, consensus statements, letters, editorials, and conference abstracts; (3) studies using only PI-RADSv1 for the evaluation of mpMRI of the prostate; (4) studies focusing on topics other than using the PI-RADSv2 system for diagnosing PCa (ie, staging and prediction of biochemical recurrence); and (5) studies with overlapping patient population.

Two reviewers (S.W. and C.H.S., with 3 yr of experience in performing systematic reviews and meta-analyses) independently evaluated the eligibility of the selected studies from the literature. Disagreements, if present between the two reviewers, were resolved by consensus via discussion with a third reviewer (S.Y.K.).

2.3. **Data extraction and quality assessment**

We extracted the following data regarding study design and results from the selected studies using a standardized form:
1. Study characteristics—authors, year of publication, country and institution of origin, duration of patient recruitment, and study design (prospective vs retrospective and consecutive or not)

2. Demographic and clinical characteristics—sample size, number of patients with PCa, patient age, prostate-specific antigen (PSA) level and Gleason score, number of previous biopsies, and PCa diagnosis prior to mpMRI

3. Technical characteristics of mpMRI—scanner model and manufacturer, magnetic field strength (1.5 vs 3 T), coil type (endorectal vs pelvic phased array), and specific sequences used (T2WI, DWI, DCE, or MR spectroscopy)

4. Interpretation of mpMRI—number of reviewers and experience in prostate mpMRI, independent or consensus reading, and blinding to clinicopathological information

5. Reference standard—type of reference standard (radical prostatectomy, targeted biopsy, or systemic biopsy), interval between MRI and pathology, outcomes assessed (any PCa vs csPCa), definition of csPCa (studies assessing “clinically significant”, “aggressive”, or “high-grade” PCa were all considered to assess csPCa; however, only studies that used the definition as provided by the PI-RADSv2 guideline [Gleason score >7 [3 + 4], volume >0.5 ml, or extraprostatic extension] were considered not to have concern for applicability), separate analysis for the PZ and TZ, and type of analysis (per patient vs per lesion)

6. Diagnostic performance of PI-RADSv2 including criteria or cutoff values (in case of multiple readers, the results of the most experienced reader were extracted for this meta-analysis)

The methodological quality of the included studies was assessed using tailored questionnaires and criteria provided by Quality Assessment of Diagnostic Accuracy Studies-2 [13]. Data extraction and quality assessment were performed independently by two reviewers (S.W. and C.H.S.). All disagreements were resolved by consensus through discussion with the third reviewer (S.Y.K.).

2.4. Data synthesis and analysis

The diagnostic performance of PI-RADSv2 for the detection of PCa was the primary outcome for this meta-analysis. In addition, a comparison between the diagnostic performance of PI-RADSv2 and that of PI-RADSv1 using studies that reported head-to-head comparison data of the two PI-RADS versions was considered a secondary outcome.

Pooled estimates of sensitivity and specificity were calculated using hierarchical logistic regression modeling including bivariate modeling and hierarchical summary receiver operating characteristic (HSROC) modeling [14]. For graphical presentation of the results, an HSROC curve with 95% confidence region and prediction region was plotted. Publication bias was evaluated using the Deeks’ funnel plot, and statistical significance was tested with the Deeks’ asymmetry test [15].

We performed meta-regression analyses to investigate the cause of heterogeneity. The following covariates were considered for the bivariate model: (1) proportion of patients with PCa (>50% vs â‰¤50%), (2) magnet strength of MRI (3 vs 1.5 T), (3) use of endorectal coil, (4) cutoff value (â‰¥4 vs >3), (5) reference standard (radical prostatectomy vs biopsy), and (6) type of analysis (per patient vs per lesion).

In addition, multiple subgroup analyses were performed for cutoff value, outcome, and previous biopsy history to assess various clinical settings: (1) a cutoff value of >4 for all studies, (2) a cutoff value of >3 for all studies, (3) a cutoff value of >4 for determining any PCa, (4) a cutoff value of >3 for determining any PCa, (5) a cutoff value of >4 for determining csPCa, (6) a cutoff value of >3 for determining csPCa, (7) a cutoff value of >4 in studies using per-patient analysis, (8) a cutoff value of >4 in studies using per-lesion analysis, (9) studies analyzing PZ PCa, (10) studies analyzing TZ PCa, (11) patients without previous biopsies, and (12) patients with previous biopsies. The “metandi” and “midas” modules in Stata 10.0 (StataCorp LP, College Station, TX, USA) and “mada” package in R software version 3.2.1 (R Foundation for Statistical Computing, Vienna, Austria) were used for statistical analyses, with p < 0.05 signifying statistical significance.

3. Evidence synthesis

3.1. Literature search

A systematic literature search initially identified 287 articles. After removing 46 duplicates, screening of the 241 titles and abstracts yielded 105 potentially eligible articles. Full-text reviews were performed, and 84 studies were excluded for the following reasons: not in the field of interest (n = 80, including 68 studies that used only PI-RADSv1), insufficient data to reconstruct 2 × 2 tables (n = 2), and shared study population with other studies (n = 2). Ultimately, 21 original articles including a total of 3857 patients assessing the diagnostic performance of PI-RADSv2 were included in the meta-analysis [16–36]. No additional studies were identified via screening the bibliographies of these 21 studies. Among them, 15 studies including 3099 patients dealt with PI-RADSv2 alone, whereas six studies including 758 patients provided a head-to-head comparison between PI-RADSv1 and PI-RADSv2 [16,20,21,28,32,33]. The detailed study selection process is described in Fig. 1.

3.2. Characteristics of included studies

Patient characteristics are shown in Table 1. The size of the study population ranged from 49 to 456 patients, with the percentage of those with PCa ranging from 37% to 100%. The patients had a median age of 62–69.6 yr, median PSA of 3.97–15 ng/ml, and a Gleason score ranging from 5 to 10. Patients had already been diagnosed with PCa prior to MRI in all or some of the study populations in seven studies [16,17,22,26,27,32,35]. Biopsy was performed before MRI in seven studies [16,21–23,26,27,35], all patients were biopsy-naive in three studies [18,25,34], both patient types were included in three studies [28,32,33], and data regarding
previous biopsy were not reported in the remaining eight studies.

Characteristics of the studies are summarized in Table 2. MRI was performed using 3-T scanners in 16 studies [16–18,20,24–33,35,36], 1.5-T scanners in four studies [19,21–23], and either 3 or 1.5 T in one study [34]. Endorectal coils were used in seven studies [19,21–25,29]. In all studies, the mpMRI protocol consisted of T2WI, DWI, and DCE-MRI. The reference standard was radical prostatectomy in five studies [16,22,26,27,35], a combination of systematic and targeted biopsies in seven studies [17,20,23,25,29,34,36], and only targeted biopsy in seven studies [19,21,24,28,30,32,33]; the reference standard was not consistent throughout the study population in two studies [18,31]. PI-RADSv2 scoring was performed by one to five radiologists, either in consensus or independently. The level of experience of the radiologists was heterogeneous, ranging from 4 to 22 yr of experience in the prostate. In most studies, the readers were blinded; however, in three studies, the radiologists were aware that the patients had biopsy-proven PCa [22,26,35], and five studies were not explicit regarding blinding [18,19,30,32,36]. In the majority of the studies, the interval between MRI and the reference standard was less than 6 mo; however, the details were not reported in 10 studies [16–18,24,27–30,32,33]. PCa was separately assessed according to zonal anatomy in seven studies [20,24,25,28,30,31,33]. However, in one study [31], only PCa in the PZ could be evaluated, as no detailed data were provided in the article and the attempt to contact the authors for provision of further information was unsuccessful. Regarding the outcome assessed, seven studies evaluated any cancer [17,19,20,23,25,28,31,33], eight evaluated clinically significant cancer [18,22,26,27,30,34–36], and six evaluated both [16,21,24,29,32]. With regard to cutoff values, 13 studies used \( C > 4 \) [16,18–21,24–28,30,31,35], four studies used \( C > 3 \) [32–34,36], and four studies used both [17,22,23,29]. The location of PCa was separately reported by the PZ and TZ in six studies [20,24,25,28,30,33].

### 3.3. Quality assessment

Overall, the quality of the studies was not considered high, mainly due to the patient selection domain (Fig. 2). Regarding the patient selection domain, there was generally a high risk of bias as all but four of the studies were retrospective in nature [16–18,20,23,25–28,30,32–36]. Seven studies were considered to have high concern for applicability, as all or some of the patients had a pathological diagnosis of PCa prior to MRI [16,17,22,26,27,32,35]. Regarding the index test domain, there was a high risk of bias in nine studies. In three of nine studies, reviewers were aware that patients had biopsy-proven PCa [22,26,35]. In the other six studies, the cutoff value for determining PCa was not specified prior to interpretation [16,20,29,31,33,34]. Only one study had concern for applicability, as PI-RADSv2 scores were indirectly generated from existing clinical radiological
Table 1 – Patient characteristics

<table>
<thead>
<tr>
<th>First author (year)</th>
<th>Origin</th>
<th>Institution</th>
<th>Duration of patient recruitment</th>
<th>Patients (n)</th>
<th>Patients with PCa (n)</th>
<th>Age (yr) Median Range</th>
<th>PSA (ng/ml) Median Range</th>
<th>Gleason score Median Range</th>
<th>No. of previous biopsies</th>
<th>PCa diagnosis before MRI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auer (2016) [16]</td>
<td>Austria</td>
<td>Medical University of Innsbruck, Medizinische Hochschule Hannover</td>
<td>NR</td>
<td>50</td>
<td>NR</td>
<td>63* NR</td>
<td>7.3* NR</td>
<td>7 (3 + 4) 6–10</td>
<td>≥1 Yes (all)</td>
<td></td>
</tr>
<tr>
<td>Baldisserotto (2016) [17]</td>
<td>Brazil</td>
<td>Pontificia Universidade Católica do Rio Grande do Sul</td>
<td>2013.6–2015.6</td>
<td>54</td>
<td>33</td>
<td>65.9 a 53–81</td>
<td>8.4 a 3–31</td>
<td>7 (3 + 4) 6–9</td>
<td>NR Yes (some)</td>
<td></td>
</tr>
<tr>
<td>De Visschere (2017) [18]</td>
<td>Belgium</td>
<td>Ghent University Hospital</td>
<td>2011.5–2014.12</td>
<td>245</td>
<td>144</td>
<td>66 44–85</td>
<td>9 1.4–935.5</td>
<td>7 (3 + 4) ≥6</td>
<td>0 No</td>
<td></td>
</tr>
<tr>
<td>El-Samei (2016) [19]</td>
<td>Egypt</td>
<td>Al-Azhar University, El-Minia University</td>
<td>2014.5–2015.10</td>
<td>55</td>
<td>38</td>
<td>62 a 51–79</td>
<td>NR NR</td>
<td>7 6–10 No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feng (2016) [20]</td>
<td>China</td>
<td>Tongji Hospital</td>
<td>2013.6–2015.7</td>
<td>401</td>
<td>150</td>
<td>64.4 a 34–88</td>
<td>10.7 0.2–1763</td>
<td>7 (4 + 3) ≤6–≥8</td>
<td>NR NR</td>
<td></td>
</tr>
<tr>
<td>Kassel-Seibert (2016) [21]</td>
<td>Germany</td>
<td>University Hospital Jena</td>
<td>2013.7–2015.3</td>
<td>82</td>
<td>31</td>
<td>65 48–81</td>
<td>13 1–111</td>
<td>7 (3 + 4) 6–9</td>
<td>1–5 No</td>
<td></td>
</tr>
<tr>
<td>Lin (2016) [22]</td>
<td>Brazil, Taiwan</td>
<td>Riberão Preto School of Medicine, China Medical University, Ribeirão Preto School of Medicine</td>
<td>2011.5–2014.6</td>
<td>49</td>
<td>49</td>
<td>63 46–73</td>
<td>13.27 1.8–41.4</td>
<td>≥1 Yes (all)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Martorana (2016) [23]</td>
<td>Italy</td>
<td>University of Modena and Reggio Emilia, Università di Perugia</td>
<td>2014.1–2016.2</td>
<td>157</td>
<td>79</td>
<td>65 47–79 b</td>
<td>10.7 1–75</td>
<td>6 6–8 1–3 No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mertan (2016) [24]</td>
<td>USA</td>
<td>National Cancer Institute</td>
<td>2015.3–2015.9</td>
<td>62</td>
<td>38</td>
<td>65.5 50.3–76.6</td>
<td>7.1 0.5–863</td>
<td>7 (3 + 4) 6–9</td>
<td>NR NR</td>
<td></td>
</tr>
<tr>
<td>Muller (2015) [25]</td>
<td>USA, the Netherlands</td>
<td>National Cancer Institute, AMC University Hospital, Edward Hébert School of Medicine</td>
<td>2011.12–2014.5</td>
<td>94</td>
<td>94</td>
<td>62 37–79</td>
<td>8.51 0.7–51.1</td>
<td>7 (3 + 4) 6–10</td>
<td>0 No</td>
<td></td>
</tr>
<tr>
<td>Park (2016) [26]</td>
<td>South Korea</td>
<td>Samsung Medical Center</td>
<td>2012.1–2014.12</td>
<td>456</td>
<td>456</td>
<td>65 46–81</td>
<td>3.97 3.8–4.2</td>
<td>NR ≥6 ≥1 Yes (all)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Park (2016) [27]</td>
<td>South Korea</td>
<td>Yonsei University College of Medicine</td>
<td>2012.1–2013.3</td>
<td>425</td>
<td>425</td>
<td>65 NR</td>
<td>8.7 0.5 NR</td>
<td>5–6 5–10 ≥1 Yes (all)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polanec (2016) [28]</td>
<td>Vienna</td>
<td>Medical University of Vienna, Konförsentät Vienna</td>
<td>2011.6–2015.9</td>
<td>65</td>
<td>65.3 a 62.3–87.4</td>
<td>10.8 a 4.2–74.5</td>
<td>NR NR</td>
<td>NR NR No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rastinehad (2015) [29]</td>
<td>USA</td>
<td>Icahn School of Medicine at Mount Sinai, Fox Chase Cancer Center, National Institutes of Health, Hofstra North Shore LIJ School of Medicine</td>
<td>2012.2–2014.11</td>
<td>312</td>
<td>202</td>
<td>65.1 60.3–70.3 b</td>
<td>7.3 5.0–11.4 b</td>
<td>NR NR NR No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rosenkrantz (2016) [30]</td>
<td>USA</td>
<td>NYU Langone Medical Center University “Federico II”, Ospedale S. Maria delle Grazie</td>
<td>2013.9–2015.2</td>
<td>343</td>
<td>134</td>
<td>64 NR</td>
<td>5.8 NR</td>
<td>7 (3 + 4) 6–9</td>
<td>NR NR No</td>
<td></td>
</tr>
<tr>
<td>Stanzione (2016) [31]</td>
<td>Italy</td>
<td>University “Federico II”, Ospedale S. Maria delle Grazie</td>
<td>2013.9–2015.2</td>
<td>343</td>
<td>134</td>
<td>64 NR</td>
<td>5.8 NR</td>
<td>7 (3 + 4) 6–9</td>
<td>NR NR No</td>
<td></td>
</tr>
<tr>
<td>Tan (2016) [32]</td>
<td>USA, Taiwan</td>
<td>UCLA, Cathay General Hospital, Hoag Hospital</td>
<td>2013.3–2016.12</td>
<td>106</td>
<td>63</td>
<td>66.5 43–79</td>
<td>7.9 5.6–10.6</td>
<td>7 (3 + 4) 6–10 ≥0 Yes (some)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tewes (2016) [33]</td>
<td>Germany</td>
<td>Hannover Medical School, Klinikum der Region Hannover</td>
<td>2012.12–2014.12</td>
<td>54</td>
<td>31</td>
<td>69.6 a NR</td>
<td>8.7 a NR</td>
<td>6 6–9 ≥0 No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Washino (2017) [34]</td>
<td>Japan</td>
<td>Jichi Medical University Saitama Medical Center</td>
<td>2010.6–2014.4</td>
<td>288</td>
<td>159</td>
<td>69 64–74 b</td>
<td>7.5 5.5–11 b</td>
<td>NR NR 0 No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Woo (2016) [35]</td>
<td>South Korea</td>
<td>Seoul National University College of Medicine</td>
<td>2011.1–2013.12</td>
<td>105</td>
<td>105</td>
<td>69 49–79</td>
<td>8.22 0.9–44.2</td>
<td>7 6–7 ≥1 Yes (all)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhao (2016) [36]</td>
<td>China</td>
<td>Peking University First Hospital</td>
<td>2010.11–2013.12</td>
<td>372</td>
<td>185</td>
<td>68.5 a NR</td>
<td>15 a NR</td>
<td>7 6–10 NR No</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MRI = magnetic resonance imaging; NR = not reported; PCa = prostate cancer; PSA = prostate-specific antigen.

* Mean.

b Interquartile range.
<table>
<thead>
<tr>
<th>Study</th>
<th>First author</th>
<th>Design</th>
<th>Consecutive enrollment</th>
<th>Reference standard</th>
<th>MRI–reference standard interval</th>
<th>No. of readers</th>
<th>Experience (yr)</th>
<th>Blinding</th>
<th>Vendor</th>
<th>Model</th>
<th>Endorectal coil</th>
<th>PI-RADSv2 application</th>
<th>Cutoff values</th>
<th>Localization</th>
<th>Type of analysis</th>
<th>Outcome assessed</th>
<th>Definition of csPCa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auer (2016) [16]</td>
<td>Retrospective</td>
<td>Yes</td>
<td>RP</td>
<td>NR</td>
<td>1-35</td>
<td>1,5</td>
<td>Siemens</td>
<td>Skyra</td>
<td>No Strict</td>
<td>4</td>
<td>Whole</td>
<td>Lesion</td>
<td>Any + csPCa</td>
<td>GS ≥ 7 (4 + 3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baldisserotto (2016) [17]</td>
<td>Retrospective</td>
<td>Yes</td>
<td>STRUSGB + targeted MRI–TRUS biopsy (cognitive) or RP</td>
<td>NR</td>
<td>10/1</td>
<td>2</td>
<td>Independent</td>
<td>&gt;10</td>
<td>Yes</td>
<td>3</td>
<td>GE</td>
<td>Signa HDxT</td>
<td>No Strict</td>
<td>3, 4</td>
<td>Whole</td>
<td>Patient</td>
<td>csPCa</td>
</tr>
<tr>
<td>De Visschere (2017) [18]</td>
<td>Retrospective</td>
<td>NR</td>
<td>STRUSGB or RP</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>3</td>
<td>Siemens</td>
<td>Trio</td>
<td>No Strict</td>
<td>4</td>
<td>Whole</td>
<td>Patient</td>
<td>csPCa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>El-Samei (2016) [19]</td>
<td>Retrospective</td>
<td>Yes</td>
<td>Targeted MRI–TRUS biopsy</td>
<td>≤2 wk</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>1.5</td>
<td>Phillips</td>
<td>Gyroscan</td>
<td>Yes</td>
<td>Strict</td>
<td>4</td>
<td>Whole</td>
<td>Lesion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kasel-Seibert (2016) [21]</td>
<td>Retrospective</td>
<td>NR</td>
<td>Targeted MRGB</td>
<td>≤7 d</td>
<td>2</td>
<td>Independent</td>
<td>10/1</td>
<td>Yes</td>
<td>1.5</td>
<td>Siemens</td>
<td>Avanto</td>
<td>Yes</td>
<td>4</td>
<td>Whole</td>
<td>Patient</td>
<td>csPCa</td>
<td></td>
</tr>
<tr>
<td>Lin (2016) [22]</td>
<td>Retrospective</td>
<td>NR</td>
<td>RP</td>
<td>≤6 mo</td>
<td>2</td>
<td>Independent</td>
<td>12/5</td>
<td>Yes</td>
<td>1.5</td>
<td>Phillips</td>
<td>Achieva</td>
<td>Yes</td>
<td>4</td>
<td>Whole</td>
<td>csPCa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Martorana (2016) [23]</td>
<td>Retrospective</td>
<td>Yes</td>
<td>TBB + targeted MRI–TRUS biopsy</td>
<td>30–78</td>
<td>2</td>
<td>Consensus</td>
<td>4/4</td>
<td>Yes</td>
<td>1.5</td>
<td>Phillips</td>
<td>Achieva</td>
<td>Yes</td>
<td>4</td>
<td>Whole</td>
<td>Lesion</td>
<td>csPCa</td>
<td></td>
</tr>
<tr>
<td>Mertan (2016) [24]</td>
<td>Prospective</td>
<td>Yes</td>
<td>STRUSGB + targeted MRI–TRUS biopsy</td>
<td>NR</td>
<td>1</td>
<td>≥8</td>
<td>Yes</td>
<td>3</td>
<td>Phillips</td>
<td>Achieva</td>
<td>Yes</td>
<td>4</td>
<td>FZ, TZ</td>
<td>csPCa</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Muller (2015) [25]</td>
<td>Retrospective</td>
<td>Yes</td>
<td>STRUSGB + targeted MRI–TRUS biopsy</td>
<td>≤6 wk</td>
<td>5</td>
<td>Independent</td>
<td>12/7/1/1/0,5</td>
<td>Yes</td>
<td>3</td>
<td>Phillips</td>
<td>Achieva</td>
<td>Yes</td>
<td>4</td>
<td>FZ, TZ</td>
<td>csPCa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Park (2016) [26]</td>
<td>Retrospective</td>
<td>Yes</td>
<td>RP</td>
<td>21–48</td>
<td>14/3</td>
<td>Yes</td>
<td>3</td>
<td>Phillips</td>
<td>Achieva</td>
<td>No Strict</td>
<td>4</td>
<td>Whole</td>
<td>Patient</td>
<td>csPCa</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Park (2016) [27]</td>
<td>Retrospective</td>
<td>NR</td>
<td>Targeted MRGB</td>
<td>NR</td>
<td>NR</td>
<td>9/4</td>
<td>Independent</td>
<td>2</td>
<td>Yes</td>
<td>3</td>
<td>GE, Siemens, Philips, DiscoveryMR750, Achieva, Trio</td>
<td>No Strict</td>
<td>4</td>
<td>Whole</td>
<td>Patient</td>
<td>csPCa</td>
<td></td>
</tr>
<tr>
<td>Polanec (2016) [28]</td>
<td>Retrospective</td>
<td>Yes</td>
<td>Targeted MRGB</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>Yes</td>
<td>3</td>
<td>Siemens</td>
<td>Trio</td>
<td>No Strict</td>
<td>4</td>
<td>FZ, TZ</td>
<td>Patient</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ratineiba (2015) [29]</td>
<td>Prospective</td>
<td>NR</td>
<td>STRUSGB + targeted MRI–TRUS biopsy</td>
<td>NR</td>
<td>NR</td>
<td>Consensus</td>
<td>NR</td>
<td>Yes</td>
<td>3</td>
<td>Siemens</td>
<td>Vario</td>
<td>Yes</td>
<td>Calculated</td>
<td>3, 4</td>
<td>Whole</td>
<td>Patient</td>
<td>Any + csPCa</td>
</tr>
<tr>
<td>Stanzione (2016) [31]</td>
<td>Prospective</td>
<td>NR</td>
<td>STRUSGB + targeted TRUSGB</td>
<td>20–30 d</td>
<td>14/10</td>
<td>2</td>
<td>Independent</td>
<td>16/10</td>
<td>Yes</td>
<td>3</td>
<td>Siemens</td>
<td>Trio</td>
<td>No Strict</td>
<td>3</td>
<td>Whole</td>
<td>Patient</td>
<td>Any + csPCa</td>
</tr>
<tr>
<td>Tan (2016) [32]</td>
<td>Retrospective</td>
<td>Yes</td>
<td>Targeted MRGB</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>3</td>
<td>Siemens</td>
<td>Trio, Verio</td>
<td>No Strict</td>
<td>3</td>
<td>Whole</td>
<td>Patient</td>
<td>csPCa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tewes (2016) [33]</td>
<td>Retrospective</td>
<td>Yes</td>
<td>Targeted MRGB</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>Yes</td>
<td>3</td>
<td>Siemens</td>
<td>Skyra</td>
<td>No Strict</td>
<td>3</td>
<td>FZ, TZ</td>
<td>Patient</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Washino (2017) [34]</td>
<td>Retrospective</td>
<td>NR</td>
<td>Targeted MRI–TRUS biopsy (cognitive)</td>
<td>0.5–1.6 mo</td>
<td>1</td>
<td>14</td>
<td>Yes</td>
<td>1.5 or 3</td>
<td>Toshiba</td>
<td>Excelart Vantage/Vantage Titan 3T</td>
<td>No Strict</td>
<td>3</td>
<td>Whole</td>
<td>Patient</td>
<td>csPCa</td>
<td>GS ≥ 7 or maximum core length ≥ 4 mm</td>
<td></td>
</tr>
</tbody>
</table>
reports based on PI-RADSv1 or an in-house scoring system [29]. Regarding the reference standard domain, eight studies had a high risk of bias. Seven were based on either only systematic biopsy or targeted biopsy [18,19,21,28,30,32,33]; in one study, targeted biopsy was performed, but on lesions that were suspicious on ultrasonography, and not on MRI [31]. Those in which radical prostatectomy or a systematic plus targeted biopsy (MRI guided, MRI–transrectal ultrasound fusion, or cognitive) was used as the reference standard were considered to have a low risk of bias. In 10 studies, the definition of clinically significant cancer did not abide by those described in the PI-RADSv2 guidelines, and therefore showed high concern for applicability [16,21,22,24,29,30,32,34–36]. Regarding the flow and timing domain, two studies had a high risk of bias as patients did not receive the same reference standard [17,18].

3.4. Diagnostic accuracy of PI-RADSv2

The sensitivity and specificity of individual studies were 73–100% and 8–100%, respectively. The Q-test revealed that substantial heterogeneity was present ($p < 0.001$). The Higgins $I^2$ statistics demonstrated substantial heterogeneity in terms of the sensitivity ($I^2 = 85.55\%$) and considerable heterogeneity in terms of the specificity ($I^2 = 95.30\%$). The coupled forest plot of the sensitivity and specificity demonstrated the absence of a threshold effect (Fig. 3). The Spearman correlation coefficient between the sensitivity and the false-positive rate was $0.45$ (95% confidence interval [CI] 0.023–0.738), also indicating the lack of a threshold effect.

For all 21 studies combined, the pooled sensitivity was $0.89$ (95% CI 0.86–0.92) with specificity of $0.73$ (95% CI 0.60–0.83; Fig. 3). In the HSROC curve, there was a large difference between the 95% confidence region and the 95% prediction region, thus indicating heterogeneity between the studies (Fig. 4). The area under the HSROC curve was $0.91$ (95% CI 0.88–0.93). According to the Deeks’ funnel plot, the likelihood of publication bias was low, with a $p$ value of 0.75 for the slope coefficient (Fig. 5).

For the six studies that provided a head-to-head comparison between PI-RADSv1 and PI-RADSv2, PI-RADSv2 demonstrated higher pooled sensitivity of $0.95$ (95% CI 0.85–0.98) compared with $0.88$ (95% CI 0.80–0.93) for PI-RADSv1 ($p = 0.04$). However, the pooled specificity did not show a significant difference between the two versions of PI-RADS: $0.73$ (95% CI 0.47–0.89) for v2 and $0.75$ (95% CI 0.36–0.94) for v1 ($p = 0.90$).

3.5. Heterogeneity exploration

As we found a considerable heterogeneity among the included studies, meta-regression analyses were performed (Supplementary Table 1). Among several potential variables, proportion of patients with PCa, magnetic field strength, and reference standard were significant factors affecting the heterogeneity ($p < 0.01$ for all). However, among these three, only specificity according to the
proportion of patients with PCa showed a statistically significant and clinically meaningful difference: 0.65 (95% CI 0.52–0.78) in studies ≤50% of patients with PCa versus 0.86 (0.75–0.97) in studies with >50% of patients with PCa. Otherwise, no clinically meaningful differences were seen: for magnet strength (3 vs 1.5 T), sensitivity of 0.90 (95% CI 0.86–0.94) versus 0.89 (95% CI 0.81–0.97, p = 0.03) and specificity of 0.73 (95% CI 0.59–0.86) versus 0.72 (95% CI

Fig. 2 – Grouped bar charts show (A) risk of bias and (B) concerns for applicability of 21 included studies using QUADAS-2. QUADAS-2 = Quality Assessment of Diagnostic Accuracy Studies-2.

Fig. 3 – Coupled forest plots of pooled sensitivity and specificity. Numbers are pooled estimates with 95% CI in parentheses. Corresponding heterogeneity statistics are provided at bottom right corners. Horizontal lines indicate 95% CIs. CI = confidence intervals.
0.44–1.00, \( p = 0.81 \); for reference standard (radical prostatectomy vs biopsy), sensitivity of 0.89 (95% CI 0.83–0.95) versus 0.91 (95% CI 0.88–0.94, \( p < 0.01 \)) and specificity of 0.65 (95% CI 0.37–0.94) versus 0.73 (95% CI 0.58–0.87, \( p = 0.48 \)). Other variables, including the cutoff value, use of endorectal coil, and type of analysis (per patient vs per lesion), were not significant factors (\( p = 0.32–0.70 \)).

3.6. Subgroup analysis

As there were four studies that used both ≥3 and ≥4 as cutoff values [17,22,23,29], or determined both any cancer and clinically significant cancer as outcomes [16,21,24,29,32], multiple subgroup analyses were performed in order to assess various clinical settings (Supplementary Table 2). Regarding cutoff values, the pooled sensitivity was 0.89 (95% CI 0.84–0.92) with specificity of 0.74 (95% CI 0.58–0.85) for 17 studies using ≥4 [16–31,35], whereas these were 0.95 (95% CI 0.89–0.97) and 0.47 (95% CI 0.21–0.74) in eight studies using ≥3 [17,22,23,29,32–34,36]. When we stratified studies according to the outcome assessed, the following results were yielded: (1) cutoff of ≥4 for determining any PCa, sensitivity of 0.89 (95% CI 0.83–0.93) with specificity of 0.80 (95% CI 0.62–0.90); (2) cutoff of ≥3 for determining any PCa, sensitivity of 0.96 (95% CI 0.93–0.98) with specificity of 0.49 (0.29–0.70); (3) for determining csPCa regardless of cutoff values, sensitivity of 0.89 (95% CI 0.84–0.92) with specificity of 0.64 (95% CI 0.46–0.78); (4) cutoff of ≥4 for determining csPCa, sensitivity of 0.90 (95% CI 0.85–0.94) with specificity of 0.62 (95% CI 0.45–0.77); and (5) cutoff of ≥3 for determining csPCa, sensitivity of 0.96 (95% CI 0.87–0.99) with specificity of 0.29 (0.05–0.77). When the studies using a cutoff value of ≥4 were separately assessed according to the type of analysis, per-patient analysis in eight studies [17,18,20,26,27,29,31,35] yielded pooled sensitivity of 0.89 (95% CI 0.81–0.93) with specificity of 0.76 (95% CI 0.60–0.88), whereas per-lesion analysis in nine studies [16,19,21–25,28,30] yielded pooled sensitivity of 0.87 (95% CI 0.83–0.91) with specificity of 0.70 (95% CI 0.44–0.88).

Based on the localization of PCa, the pooled sensitivity was 0.93 (95% CI 0.87–0.96) with specificity of 0.68 (95% CI 0.43–0.86) in seven studies analyzing PZ cancers [20,24,25,28,30,31,33]. In identical studies, except for that by Stanzione et al [31], which analyzed TZ cancers, the pooled sensitivity and specificity were 0.88 (95% CI 0.77–0.94) and 0.75 (95% CI 0.59–0.86), respectively.

Studies including patients without previous biopsies yielded sensitivity of 0.82 (95% CI 0.72–0.90) and specificity of 0.75 (95% CI 0.65–0.83), whereas the diagnostic performance values were 0.87 (95% CI 0.80–0.92) and 0.71 (95% CI 0.42–0.89) in studies including patients with a history of previous biopsy.

3.7. Discussion

In our meta-analysis, we assessed the diagnostic accuracy of PI-RADSv2 for detecting PCa. The pooled sensitivity and specificity of all 21 studies were 0.89 (95% CI 0.86–0.92) and 0.73 (95% CI 0.60–0.83), respectively. When comparing our data with the only two existing meta-analyses using mpMRI for detecting PCa, a trend toward higher sensitivity and lower specificity can be inferred. In the study by de Rooij...
et al [37], which evaluated seven studies using a combination of T2WI, DWI, and DCE-MRI, the pooled sensitivity and specificity were 0.74 (95% CI 0.66–0.81) and 0.88 (95% CI 0.82–0.92), respectively. In a more recent meta-analysis by Hamoen et al [6] which analyzed 14 studies using PI-RADSv1, the pooled sensitivity and specificity were 0.78 (95% CI 0.70–0.84) and 0.79 (95% CI 0.68–0.86), respectively. However, the comparison between the three studies merely provided an indirect comparison. In order to address this issue, we separately assessed a subgroup of studies using both PI-RADSv1 and PI-RADSv2. In a head-to-head comparison between them, PI-RADSv2 demonstrated higher pooled sensitivity (0.95) compared with PI-RADSv1 (0.88, p = 0.04) without a statistically significant difference in specificity (0.73 vs 0.75, p = 0.90). This increase in sensitivity compared with its predecessor may imply that the revisions undertaken during the development of PI-RADSv2, including the introduction of dominant sequences according to zonal anatomy, limited contribution of DCE-MRI secondary to DWI and T2WI, and specific guidelines for deriving an integrated overall score, were, in fact, on the right track. Especially, we speculate that the use of dominant sequences, that is, DWI for the PZ and DCE-MRI for the TZ, may have been crucial for the improved sensitivity without a loss in specificity, as suggested by Baur et al [10].

Considering that one of the main intentions for the generation of PI-RADS was to standardize reporting of mpMRI in order to decrease variability and bring about widespread acceptance and implementation in daily practice, it was promising to find that nearly all (20 of 21) studies used PI-RADSv2 strictly according to published guidelines [11]. Only one study formed PI-RADSv2 scores from existing clinical radiological reports that were based on PI-RADSv1 or an in-house scoring system [29]. This is an improvement when compared with prior studies conducted using PI-RADSv1, where investigators used varying methods in determining the overall score (overall five-point score or sum of the scores from each modality) [6]. Still, there is a need for further clarification regarding the cutoff value for detecting PCa. In the studies included in our meta-analysis, cutoff values were predefined in only six studies, while the majority (15/21) were exploratory in nature, testing multiple criteria. When using a cutoff value of >4, sensitivity (0.89) and specificity (0.74) were generally good, whereas using ≥3 yielded excellent sensitivity (0.95) and poor specificity (0.47). These results may be taken into consideration when generating the next updated PI-RADS. For instance, using the former may be adequate for general use of PI-RADS, whereas the latter could be proposed to be indicated when a higher cancer detection rate is clinically required (ie, persistently high PSA level despite a previously negative biopsy).

In the current study, subgroup analyses were performed to account for differences in outcomes (any cancer vs clinically significant cancer). There was no significant difference for using either outcome irrespective of whether the criteria of ≥3 or ≥4 were used. However, the definition of clinically significant cancer was different among the 13 studies. Only three studies defined csPCa strictly according to the PI-RADSv2 guidelines (Gleason score ≥7 [3 + 4], volume >0.5 ml, or extraprostatic extension) [11]. Most others used one or two of the three criteria. Including only the former three studies may have provided more robust results; yet it was not only pragmatic to include all available studies, but this approach would present a general overview of the existing literature, as it is the first meta-analysis of studies currently dealing with PI-RADSv2.

In this meta-analysis, we looked into the technical aspects of MRI. Meta-regression analyses revealed that the use of endorectal coil was not a statistically significant factor. Furthermore, although magnet strength showed statistically significant differences between 3 and 1.5 T, this did not reveal to be clinically meaningful (sensitivity of 0.90 vs 0.89, p = 0.03, respectively). Although there had been debate over these two issues in the past, both 3 and 1.5 T are now well established, and the overall benefit of using an endorectal coil is not evident [38,39]. The PI-RADSv2 guidelines currently recommend either usage, and the results of our study provide additional evidence to support this.

Regarding the methods of analysis in the studies, there was significant heterogeneity regarding reference standard and type of analysis. Radical prostatectomy was the reference standard in five studies, while the majority were based on a combination of systematic and targeted biopsies. The possibility of PCa despite negative biopsy results in the latter group should be kept in mind. In addition, approximately half of the studies each reported outcomes in a per-patient (n = 11) and per-lesion (n = 10) Manner. Per-lesion analysis is known to take into account the performance of localizing the disease; however, this was not shown to be a significant factor in the meta-regression analysis.

Our meta-analysis had some limitations. Nearly all studies were retrospective in study design, resulting in a high risk of bias for patient selection. It is possible that pooling data from predominantly retrospective studies may have led to increased diagnostic sensitivity [40]. In addition, not only was performing a meta-analysis using only three prospective studies technically unfeasible, but the derived results would not be representative of the existing literature on PI-RADSv2 as well. Furthermore, we used validated methods for the systematic review and reported the data using standard reporting guidelines, including PRISMA and the guidelines of the Handbook for Diagnostic Test Accuracy Reviews published by the Cochrane Collaboration [12,41]. Another limitation is considerable heterogeneity in our pooled analysis, which affected the general applicability of our summary estimates. To explore the heterogeneity of our data, we performed meta-regression and multiple subgroup analyses. According to the analyses, the proportion of patients with PCa, the magnetic field strength, and the reference standard were significant factors affecting the heterogeneity. Especially, the reference standard included various methods, including radical prostatectomy and a combination of systematic and targeted biopsies (ie, MRI guided, MRI-transrectal ultrasound fusion, or cognitive). Furthermore, the fact that various definitions were
used for clinically significant cancer needs to be emphasized. Our analyses using meta-regression and subgroup analyses may explain some of the heterogeneity, but a portion remains unexplained. Another important limitation is the small number of included studies for head-to-head comparison between PI-RADSv1 and PI-RADSv2. However, we were able to elucidate statistically significant difference in the sensitivity between the two versions using only six studies, which have been published until now.

4. Conclusions

PI-RADSv2 shows good performance for the detection of PCa with pooled sensitivity of 0.89 and specificity of 0.73. PI-RADSv2 has higher pooled sensitivity compared with PI-RADSv1 without significantly different specificity.

Author contributions: Sang Yoon Kim had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Woo, Suh, S.Y. Kim.

Acquisition of data: Woo, Suh, S.Y. Kim.

Analysis and interpretation of data: Woo, Suh, S.Y. Kim.

Drafting of the manuscript: Woo.

Critical revision of the manuscript for important intellectual content: Suh, S.Y. Kim, Cho, S.H. Kim.

Statistical analysis: Suh.

Obtaining funding: None.

Administrative, technical, or material support: None.


Other: None.

Financial disclosures: Sang Yoon Kim certifies that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (eg, employment/affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received, or pending), are the following: None.

Funding/Support and role of the sponsor: None.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.eururo.2017.01.042.

References


Park JJ, Park BK. Role of PI-RADSv2 with multiparametric MRI in determining who needs active surveillance or definitive treatment according to PRIAS. J Magn Reson Imaging 2017;45:1753–9.


