Prostate Imaging Reporting and Data System and Likert Scoring System: Multiparametric MR Imaging Validation Study to Screen Patients for Initial Biopsy

**Raphaëlle Renard-Penna, MD**
**Pierre Mozer, MD, PhD**
**François Cornud, MD**
**Nicolas Barry-Delongchamps, MD**
**Eric Bruguière, MD**
**Daniel Portalez, MD**
**Bernard Malavaud, MD, PhD**

**Purpose:** To compare the diagnostic performance of the magnetic resonance (MR) imaging–based Prostate Imaging Reporting and Data System (PI-RADS) and a Likert scale in the detection of prostate cancer in a cohort of patients undergoing initial prostate biopsy.

**Materials and Methods:** This institutional review board–approved two-center prospective study included 118 patients with normal digital rectal examination (DRE) results but elevated prostate-specific antigen (PSA) levels (4–20 ng/mL) who were referred for initial prostate biopsies and had one suspicious (Likert scale score, ≥3) focus at prebiopsy 1.5-T multiparametric MR imaging performed with T2-weighted, diffusion-weighted [DW], and dynamic contrast material–enhanced imaging. Targeted core biopsies and random systematic core biopsies were performed. The elementary unit for analysis was the core. Relationships were assessed by using the Mann-Whitney U test. Yates corrected and Pearson χ² tests were used to evaluate categoric variables. A training set was randomly drawn to construct the receiver operating characteristic curves for the summed PI-RADS scores and for the Likert scale scores. The thresholds to recommend biopsy were obtained from the Youden J statistics and were tested in the remaining validation set in terms of predictive characteristics. Interobserver variability was analyzed by using weighed κ statistics in a random set of 50 patients.

**Results:** Higher T2-weighted, DW, and dynamic contrast-enhanced imaging PI-RADS scores were observed in areas that yielded cancer-positive cores. The percentage of positive cores increased with the sum of scores aggregated in five classes as follows: For summed PI-RADS scores of 3–5, the percentage of positive cores was 2.3%; for scores of 6–8, it was 5.8%; for scores of 9 or 10, it was 24.7%; for scores of 11 or 12, it was 51.8%; and for scores of 13–15, it was 72.1% (P for trend, <.0001). For the threshold of summed PI-RADS scores of 9 or greater, sensitivity was 86.6%, specificity was 82.4%, the positive predictive value was 52.4%, the negative predictive value was 96.5%, and accuracy was 83.2%. The respective data for Likert scale scores of 3 or greater were 93.8%, 73.6%, 44.3%, 98.1%, and 73.3%. Good interobserver agreement was observed for the Likert scale (κ = 0.80) and the summed PI-RADS (κ = 0.73) scoring systems.

**Conclusion:** PI-RADS provided the site-specific stratified risk of cancer-positive cores in biopsy-naive men with normal DRE results and elevated PSA levels. There was no significant difference between summed PI-RADS scores of 9 or greater and Likert scale scores of 3 or greater in the detection of cancer in the peripheral zone.

1From the Departments of Radiology (R.R.) and Urology (P.M.), Hôpital Pitié-Salpétrière, Paris, France; Department of Radiology, Hôpital Cochin, Paris, France (F.C., N.B.); and Departments of Radiology (E.B., D.P.) and Urology (B.M.), Institut Universitaire du Cancer, 1 avenue Irène Joliot-Curie, 31059 Toulouse Cedex 9, France. From the 2013 RSNA Annual Meeting. Received February 6, 2014; revision requested March 19; revision received June 19; accepted July 8; final version accepted October 22. Address correspondence to B.M. (e-mail: malavaud.bernard@iuct-oncopole.fr).

© RSNA, 2015
Widespread prostate-specific antigen (PSA) testing and random sampling of the prostate gland by means of transrectal ultrasonographically (US)-guided biopsy (1) were instrumental in establishing prostate cancer as the most prevalent cancer in male patients in the United States, with more than 240,000 new cases expected in 2013 (2). As acknowledged by the American Urological Association and the European Association of Urology (3,4) this paradigm also resulted in the incidental detection of a silent reservoir of well-differentiated small-volume cancers of marginal clinical importance (5). Because multiparametric magnetic resonance (MR) imaging allows noninvasive evaluation of the anatomy, angiogenesis, and cell density of the prostate gland (6,7), it could shift the paradigm from random sampling of the gland to targeted biopsies directed at clinically important cancers (8).

One major impediment to the promotion of multiparametric MR imaging is the lack of standardization in the expression of results (9). The European Society of Urogenital Radiology (ESUR) proposed use of the MR Prostate Imaging Reporting and Data System (PI-RADS) (10), while the PREDICT (Prostate Diagnostic Imaging Consensus Meeting) panel recommended use of the five-point Likert scale (9). Both systems reflect experts’ opinions, as evidence-based recommendations could not yet be formulated.

Another limitation is that cancer detection is traditionally assessed on a patient-by-patient basis, where the absence of cancer can be suspected but not demonstrated in patients with negative transrectal US-guided biopsy results, while minute correlations between pathologic and multiparametric MR imaging findings are possible in other patients who undergo resection and whole-mount step-section pathologic examination, amounting to a classic verification bias. To deal with this problem, we took advantage of a commercial MR imaging–transrectal US fusion technology (11–13) that provides records of the spatial distribution of the cores within the prostate volume in the three-dimensional (3D) transrectal US and multiparametric MR imaging archives (11).

Because most studies included patients with repeat biopsies who were likely to differ in terms of cancer prevalence, volume, and Gleason score from patients undergoing initial biopsy (14–16), we sought to compare the performance of the PI-RADS and a Likert scale in cancer detection in a cohort of patients undergoing initial prostate biopsy.

Materials and Methods

Patients

This prospective study, which was approved by the Comité de Protection des Personnes Sud-Ouest et Outre Mer, Bordeaux, France (CPP-DC2011/37), was conducted at two academic institutions. Informed consent was obtained from all patients.

Inclusion criteria were as follows: From June 2011 to December 2012, patients who were referred for initial biopsies for elevated PSA levels (4–20 ng/mL) without findings suggestive of cancer at digital rectal examination (DRE) and who were offered multiparametric MR imaging. That is the diagnostic

Implication for Patient Care

The results of the present study therefore support the role of multiparametric MR imaging in the initial evaluation of clinically suspected prostate cancer, where it could facilitate the detection of clinically important cancer foci.

Abbreviations:

- ADC = apparent diffusion coefficient
- AUC = area under the ROC curve
- CI = confidence interval
- DICOM = Digital Imaging and Communications in Medicine
- DRE = digital rectal examination
- DW = diffusion weighted
- ESUR = European Society of Urogenital Radiology
- PI-RADS = Prostate Imaging Reporting and Data System
- PSA = prostate-specific antigen
- ROC = receiver operating characteristic
- 3D = three-dimensional

Author contributions:

Guarantors of integrity of entire study, R.R., F.C., D.P., B.M.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; manuscript final version approval, all authors; literature research, R.R., P.M., F.C., N.B., E.B., B.M.; clinical studies, R.R., P.M., F.C., E.B., D.P., B.M.; experimental studies, P.M., F.C., N.B., E.B., D.P.; statistical analysis, F.C., E.B., B.M.; and manuscript editing, R.R., P.M., F.C., E.B., D.P., B.M.

Conflicts of interest are listed at the end of this article.
pathway proposed after negative biopsy results to patients in whom there is persistent clinical suspicion of prostate cancer (4).

Patients with a single focus of suspicion at multiparametric MR imaging (defined by a score ≥ 3 on the five-point Likert scale [9]) were offered protocol biopsies consisting of two or three targeted cores plus 12 random systematic cores obtained at the same session by using a deformable transrectal US–MR imaging image fusion system for biopsy guidance. Patients with suspicious findings at DRE, in whom the American Urological Association and the European Association of Urology guidelines recommend that additional cores should be obtained from the DRE suspect area (3,4), were excluded from the study.

**Multiparametric MR Imaging**

The MR imaging pulse sequences used followed the ESUR prostate MR imaging guidelines (10) and the guidelines of a European consensus meeting on multiparametric MR imaging (17). Table 1 provides an ESUR- and START Consortium–compliant (18) description of the protocol. All patients were imaged in the supine position with a 1.5-T system, without an endorectal coil. Fast spin-echo T2-weighted images were first acquired in three planes. DW imaging with multiple b values and apparent diffusion coefficient (ADC) mapping were performed in the same planes as the T2-weighted sequences. Fat-saturated T1-weighted fast-field-echo images (echo-planar imaging) with a temporal resolution of 8.5–15 seconds were acquired before and after a bolus injection (20 mL/18 sec) of gadoterate meglumine (Dotarem; Guerbet, Aulnay-Sous-Bois, France).

**Prebiopsy MR Image Analysis**

The objective was to validate inclusion (single lesion with a Likert scale score ≥ 3) before protocol biopsies. Multiparametric MR imaging data were analyzed at workstations (institution 1: MR Workspace, Syngovia Siemens Healthcare, Erlangen, Germany; institution 2: Extended WorkSpace, Phillips Healthcare, Erlangen, Germany; institution 2: Extended WorkSpace, Phillips Healthcare, Erlangen, Germany; institution 2: Extended WorkSpace, Phillips Healthcare, Erlangen, Germany).
Healthcare, Best, the Netherlands) by one senior radiologist (R.R. or D.P., both with > 10 years of experience in prostate MR imaging) who was blinded to the PSA data.

Pulse sequence images were analyzed independently and were scored according to the Likert scoring system (10). Focal signal decrease was first evaluated from base to apex on axial T2-weighted transverse sections, taking into account the zonal anatomy of the gland. Suspicious foci were then evaluated on DW sequences as areas showing restricted diffusion on axial transverse ADC maps and high signal intensity on high-b-value images. Focal enhancement after contrast medium injection was viewed on axial transverse dynamic contrast material–enhanced images. Regions of interest were manually drawn around areas of suspicion on the dynamic contrast-enhanced images to obtain perfusion curves.

MR Imaging/3D Real-time US Fusion–guided Biopsy

Elastic fusion registration and MR imaging–guided transrectal biopsies were performed by using a commercially available U.S. Food and Drug Administration–approved MR imaging/3D transrectal US fusion–guided system (Urostation; Koelis, LaTronche, France) shown to allow precise 3D transrectal US targeting of a predefined location, recording of the core spatial location, and, ultimately, fusion of MR imaging–Digital Imaging and Communications in Medicine (DICOM) archives and transrectal US images (11,19–21).

Briefly, MR imaging data were loaded in the Koelis workstation to achieve segmentation of the prostate volume (19). Regions of interest were superimposed on areas that were suspicious at multiparametric MR imaging. Immediate prostate 3D transrectal US reconstruction was performed from one axial and two oblique acquisitions by using a motorized 3D end-fire endorectal probe (HD9; Philips Medical Systems, Best, the Netherlands). Three-dimensional transrectal US and MR imaging reconstructions were finally registered by means of an algorithm that controls the prostate deformations and changes in position at the time of transrectal US and biopsy (11).

Given the rigid attachment of the biopsy guide to the probe, the position of a presumptive core could be modeled from the position of the probe. Repeat 3D transrectal US acquisitions then allowed modeling of the spatial distribution of a simulated core in relation to the target within the prostate volume. This function was systematically used for targeted cores to optimize the position of the probe before firing the biopsy gun.

Two targeted cores were obtained, followed by 12 sextant-random systematic laterally directed cores with no systematic transition zone cores. A third targeted core was obtained when the first two did not provide an adequate sample. For all cores, 3D transrectal US acquisition was repeated with the needle in situ to register its precise location within the prostate volume. At the end of the procedure, all cores, numbered consecutively, were referenced to the MR imaging and 3D transrectal US archives (Fig 1).

Procedures were performed in conjunction by a senior radiologist in charge of target definition and prostate contouring (R.R. or D.P.) and a senior urologist (B.M. or P.M., both with > 20 years of experience in transrectal US–guided biopsy and > 200 fusion biopsies) who obtained the targeted and random biopsy cores. Cores were numbered and referred in microcassettes for pathologic examination by dedicated uropathologists who evaluated on a core-by-core basis the presence and length of cancer, as well as the Gleason primary and secondary grades.

Postprocedure Analysis

To acknowledge the diversity of multiparametric MR imaging features within the volumes sampled by random cores and the potential variations in accuracy for targeted cores, multiparametric MR imaging scores were assessed after the procedure on a core-by-core basis. To that purpose, the Koelis workstation was used to register the spatial distribution of the cores within the multiparametric MR imaging DICOM archives (19). The multiparametric MR imaging features of the volumes sampled by the individual cores were then characterized at the corresponding workstation according to the PI-RADS system and the Likert scale (9,10) by a radiologist who was blinded to the pathologic results (R.R., D.P., E.B., or F.C., all with > 10 years of experience in transrectal US–guided biopsy) and the PI-RADS score at multiparametric MR imaging, 14) that was targeted by two targeted (red) cores in complement to 12 random (green) systematic cores in a 61-ml prostate.
of experience in prostate MR imaging). Likert scale scores were defined as follows: A score of 1 indicated clinically important disease was highly unlikely to be present; a score of 2, clinically important cancer was unlikely to be present; a score of 3, the presence of clinically important cancer was equivocal; a score of 4, clinically important cancer was likely to be present; and a score of 5, clinically important disease was highly likely to be present. Cores were also annotated in terms of location in one of the 16 sectors in the European consensus panel diagram (17).

**Interobserver Agreement in Likert and PI-RADS Scores**

Data in a random set of 50 patients (25 per institution) were drawn and scored independently by two radiologists (R.R. and D.P.) to evaluate interobserver variability.

**Statistical Analysis**

We present means and standard deviations or 95% confidence intervals (CIs) for continuous variables and percentages for categoric variables.

The elementary unit for analysis was the core, annotated with pathologic and multiparametric MR imaging characteristics assessed after the procedure. The relationship between scores and biopsy results was then assessed by using the nonparametric Mann-Whitney U test. Yates corrected and Pearson χ² tests were used to evaluate the association between categoric variables. All P values were two sided. The level for statistical significance was set at P < .05.

A training set of two-thirds of the study population was randomly drawn to construct the empiric receiver operating characteristic (ROC) curves of the multiparametric MR imaging score systems. The thresholds to recommend a biopsy were estimated from the Youden J statistics (sensitivity + specificity − 1) and were tested in terms of diagnostic performance in the remaining third of the population.

The characteristics of positive cores were compared according to the above-defined thresholds. Adverse pathologic features in a core were defined according to Harnden et al (22) (cancer length < 3 mm and no Gleason pattern 4 or 5 vs cancer length ≥ 3 mm or Gleason pattern 4 or 5). Given the ordinal nature of the PI-RADS and Likert score systems, the weighed version of the κ statistics was used to evaluate interobserver variability (23). κ Scores were expressed according to the verbal categories of Byrt (24). Because κ scores not only reflect agreement but are also affected by the distribution of data across the scoring categories, the maximum possible κ score was calculated for each sequence (25). The proportion of agreement between the two readers was also presented.

**Results**

**Population**

There were 446 patients with PSA levels of 4–20 ng/mL who underwent multiparametric MR imaging. Of those, 245 (54.9%) had findings with a Likert scale score of 3 or higher. One hundred eighteen (26.3%) patients with one focus with a Likert scale score of 3 or higher were offered protocol biopsies. No patients declined biopsies.

Cancer was demonstrated in 69 (58.5%) of 118 patients. No differences in PSA level were demonstrated in relation to cancer diagnosis (PSA: 7.3 ng/mL ± 2.9 vs 7.7 ng/mL ± 3.5 in patients without and those with cancer, respectively, P = .59). Core characteristics in the whole database (n = 1731) and according to biopsy technique are presented in Table 2.

**Cancer Detection at Multiparametric MR Imaging**

In keeping with the primary biopsy setting of the present study, the majority of cores sampled the peripheral zone of the gland (1695 [97.9%] of 1731). Positive cores (n = 310) showed higher scores at T2-weighted, DW, and dynamic contrast-enhanced imaging than did negative cores (n = 1421) (T2-weighted imaging: 3.9 ± 0.9 vs 2.3 ± 0.9; DW imaging: 4.1 ± 1.2 vs 1.8 ± 1.1; and dynamic contrast-enhanced imaging: 3.9 ± 1.3 vs 1.9 ± 1.1; P < .00001 for all).

Conversely, the percentage of positive cores increased with T2-weighted imaging, DW imaging, and dynamic contrast-enhanced imaging scores (Fig 2). A continuous increase in the proportion of positive cores was also observed when the sums of the scores at T2-weighted, DW, and dynamic contrast-enhanced imaging (summed PI-RADS scores) were organized in five classes (mean and 95% CIs: 18 of 793 [2.3%; 95% CI: 1.2%, 3.3%] for summed PI-RADS scores of 3–5; 24 of 417 [5.8%; 95% CI: 3.5%, 8.0%] for summed scores of 6–8; 44 of 178 [24.7%; 95% CI: 18.3%, 31.1%] for summed scores of 9 or 10; 59 of 114 [51.8%; 95% CI: 42.4%, 61.1%] for summed scores of 11 or 12; and 165 of 229 [72.1%; 95% CI: 66.2%, 77.9%] for summed scores of 13–15 [P for trend, <.0001]) (Fig 3). The cancer yield also increased in proportion to Likert scale score (Likert scale score of 1: 16 of 730 [2.2%; 95% CI: 1.1%, 3.3%]; Likert scale score of 2: eight of 337 [2.4%; 95% CI: 0.1%, 4.0%]; Likert scale score of 3: 39 of 292 [13.4%; 95% CI: 9.4%, 17.3%]; Likert scale score of 4: 80 of 147 [54.4%; 95% CI: 46.3%, 62.6%]; and Likert scale score of 5: 165 of 229 [74.2%; 95% CI: 68.5%, 80.0%] [P for trend, <.0001]).

The balance between sensitivity and specificity for different thresholds was analyzed by using the ROC curves obtained in a randomly selected training set of 1119 cores amounting to 66% of the total cohort (Fig 4). Both systems showed high areas under the ROC curve (AUCs): 0.89 (95% CI: 0.87, 0.91) and 0.90 (95% CI: 0.88, 0.92), suggesting clinically relevant predictive characteristics (26). The Youden J statistics indicated threshold values of 9 for the summed PI-RADS scores and 3 for the Likert scale scores.

The characteristics of these thresholds were then validated in the 612 remaining cores (Table 3). Both systems exhibited excellent negative predictive values, but the summed PI-RADS scores showed a trend toward higher positive predictive value, at 52.4% (97
Table 2

Core Characteristics in the Entire Database and according to Biopsy Technique

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>All Cores (n = 1731)</th>
<th>Random Systematic Cores (n = 1414)</th>
<th>Targeted Cores (n = 317)</th>
<th>PValue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiparametric MR imaging feature</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PI-RADS score at T2-weighted imaging (from 1 to 5)</td>
<td>...</td>
<td>2.3 ± 1.1</td>
<td>3.6 ± 0.9</td>
<td>.00001</td>
</tr>
<tr>
<td>PI-RADS score at DW imaging (from 1 to 5)</td>
<td>...</td>
<td>1.9 ± 1.2</td>
<td>3.5 ± 1.3</td>
<td>&lt;.00001</td>
</tr>
<tr>
<td>PI-RADS score at dynamic contrast-enhanced imaging (from 1 to 5)</td>
<td>...</td>
<td>2.1 ± 1.2</td>
<td>3.4 ± 1.4</td>
<td>&lt;.00001</td>
</tr>
<tr>
<td>Summed PI-RADS score (from 3 to 15)</td>
<td>...</td>
<td>6.3 ± 3.3</td>
<td>10.6 ± 3.3</td>
<td>&lt;.00001</td>
</tr>
<tr>
<td>Likert score (from 1 to 5)</td>
<td>...</td>
<td>2.0 ± 1.3</td>
<td>3.6 ± 1.2</td>
<td>&lt;.00001</td>
</tr>
<tr>
<td>Pathologic features</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Core length (mm)</td>
<td>11.2 ± 3.7 (1–18)</td>
<td>11.8 ± 3.2 (1–18)</td>
<td>12.5 ± 3.3 (1–18)</td>
<td>.02</td>
</tr>
<tr>
<td>No. of cancerous cores*</td>
<td>310 (17.9)</td>
<td>187 (13.2)</td>
<td>123 (38.8)</td>
<td>&lt;.000011</td>
</tr>
<tr>
<td>Cancer length (mm)</td>
<td>6.2 ± 4.0 (1–18)</td>
<td>5.4 ± 3.7</td>
<td>7.5 ± 4.1</td>
<td>&lt;.00001</td>
</tr>
<tr>
<td>Percentage of core positive for cancer</td>
<td>52.7 ± 31.2</td>
<td>41.1 ± 30.8</td>
<td>61.2 ± 30.0</td>
<td>.00008</td>
</tr>
<tr>
<td>Gleason score*</td>
<td></td>
<td></td>
<td></td>
<td>.025</td>
</tr>
<tr>
<td>3 + 3</td>
<td>144</td>
<td>99</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>3 + 4</td>
<td>86</td>
<td>47</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>4 + 3</td>
<td>38</td>
<td>22</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>4 + 4</td>
<td>42</td>
<td>19</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>Clinically important cancer*‡</td>
<td></td>
<td></td>
<td></td>
<td>.0034*</td>
</tr>
<tr>
<td>No</td>
<td>50/310 (16.1)</td>
<td>40/187 (21.4)</td>
<td>10/123 (8.1)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>260/310 (83.9)</td>
<td>147/187 (78.6)</td>
<td>113/123 (91.9)</td>
<td></td>
</tr>
</tbody>
</table>

Note.—Unless otherwise specified, data are means ± standard deviations, with ranges in parentheses. Unless otherwise specified, P values were calculated with the Mann-Whitney U test.

* Data are numbers of cores, with percentages in parentheses.
*† Calculated with the Yates corrected χ² test.
* § Odds ratio = 4.1 (95% CI: 3.2, 5.5).
*† Calculated with the Pearson χ² test.
*§ According to Harnden et al (22), cancers that are not clinically important are < 3 mm long and do not have a Gleason score of 4 or 5; cancers that are clinically important are ≥ 3 mm long and/or have a Gleason score of 4 or 5.
* Odds ratio = 3.1 (95% CI: 1.5, 6.4).

Figure 2: Graphs show percentages of cores positive for cancer according to PI-RADS scores at T2-weighted MR imaging (T2-W) (left), DW imaging (DWI) (center), and dynamic contrast-enhanced (DCE) imaging (right). Parentheses = number of cores obtained for each category.
Figure 3: Graphs show percentages of cores positive for cancer (red) and cores showing aggressive features (according to Harnden et al [22]) (blue), according to, A, the summed PI-RADS scores in five increments and, B, the five categories of the Likert scale. Parentheses = number of cores obtained for each category.

Relationship between Multiparametric MR Imaging Findings and Adverse Features in Positive Cores

Higher T2-weighted, dynamic contrast-enhanced, and DW imaging scores were observed in cores with summed Gleason scores of 7 or higher than in cores with summed Gleason scores of 6. Incremental scores also correlated with the length of cancer in the cores. As a consequence, significantly higher scores were demonstrated in cores that showed aggressive features according to the Harnden definition (Table 5), as compared with cores that showed incidental findings. Conversely, the percentage of positive cores also increased with summed PI-RADS and Likert scores (Fig 3).

Interobserver Agreement of Likert and PI-RADS Scores

As shown in Table 6, good interobserver agreement (24) was observed for the Likert scale ($\kappa = 0.80$ [95% CI: 0.69, 0.91]) and the summed PI-RADS ($\kappa = 0.73$ [95% CI: 0.61, 0.83]) scoring systems. Regarding the three MR imaging sequences, good levels of agreement were achieved for PI-RADS T2-weighted and dynamic contrast-enhanced scores ($\kappa = 0.61$ [95% CI: 0.43, 0.78] and 0.71 [95% CI: 0.60, 0.83], respectively) while only fair consistency was observed for DW imaging scores ($\kappa = 0.53$ [95% CI: 0.38, 0.69]).

Discussion

In our study, the PI-RADS/ESUR scoring system for multiparametric MR imaging was validated to stratify the likelihood of cancer detection in a given location at initial biopsy. The cancer yield of biopsies increased proportionally with T2-weighted, DW, and dynamic contrast-enhanced imaging scores, thereby confirming the relevance of the PI-RADS system in the stratification of cancer suspicion. Whatever sequence was used, the first two scoring increments (scores 1 and 2) were frequent (941 [54.5%], 1170 [67.6%], and 1128 [65.3%] of 1731 at T2-weighted, DW, and dynamic contrast-enhanced imaging, respectively) but entailed a low risk of cancer (at T2-weighted imaging: 21 [2.2%] of 941; at DW imaging: 36 [3.1%] of 1170; and at dynamic contrast-enhanced imaging: 57 [5.1%] of 1128). This was also true for both the summed PI-RADS scores and the Likert
cores. The Likert scale rates the composite degree of suspicion of cancer, while with PI-RADS, the degree of suspicion of cancer with each pulse sequence type was scored independently. The structured and semiquantitative process of the PI-RADS may facilitate the training of radiologists and encourage the use of multiparametric MR imaging in prostate cancer diagnosis. In a cohort of patients scheduled for radical prostatectomy, Rosenkrantz et al (27) compared the Likert and PI-RADS scales in terms of accuracy among three senior radiologists. While both scales proved equally useful in the peripheral zone for detection of cancers larger than 3 mm in maximal diameter at prostatectomy (accuracy range: 88.2%–87.1% and 88.5%–89.6% for Likert and PI-RADS, respectively), the Likert scale showed better accuracy in the transition zone (accuracy range: 87.1%–92.6% and 70.0%–87.6%, respectively). This is consistent with previous reports.

**Table 3**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Training Set (n = 1119)</th>
<th>Validation Set (n = 612)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Summed PI-RADS Scores</td>
<td>Likert Scale Scores</td>
</tr>
<tr>
<td>AUC</td>
<td>0.89 ± 0.02</td>
<td>0.89 ± 0.01</td>
</tr>
<tr>
<td>Youden J-test–selected threshold</td>
<td>≥9</td>
<td>≥3</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>86.4 (80.8, 90.8) [171/198]</td>
<td>91.4 (86.6, 94.9) [181/198]</td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>82.1 (79.4, 84.5) [756/921]</td>
<td>73.3 (70.3, 76.1) [675/921]</td>
</tr>
<tr>
<td>Positive predictive value (%)</td>
<td>50.9 (45.4, 56.3) [171/336]</td>
<td>42.4 (37.7, 47.2) [181/427]</td>
</tr>
<tr>
<td>Negative predictive value (%)</td>
<td>96.6 (95.0, 97.7) [756/783]</td>
<td>97.5 (96.0, 98.5) [675/692]</td>
</tr>
<tr>
<td>Overall accuracy (%)</td>
<td>82.8 (80.5, 85.0) [927/1119]</td>
<td>76.5 (73.9, 78.9) [856/1119]</td>
</tr>
</tbody>
</table>

Note.—In the training set, there were 198 (17.7%) positive cores, 913 (81.6%) systematic cores, and 206 (18.4%) targeted cores. In the validation set, there were 112 (18.3%) positive cores, 501 (81.9%) systematic cores, and 111 (18.1%) targeted cores. Data in parentheses are 95% CIs, and data in brackets are raw data.

* No differences between the AUCs of the ESUR score and those of the Likert scale score were demonstrated (P = .78, C statistics).
that showed that dynamic contrast-enhanced imaging (28,29) and DW imaging (29) carried no incremental value in the transition zone, which is characterized by foci of benign prostatic hyperplasia, dense fibrotic tissue, and increased vascularity. Assigning the same weight in interpretation to T2-weighted and DW or dynamic contrast-enhanced imaging might therefore be ill adapted to the specifics of this zone.

In view of the results, on the basis of initial biopsy results, the sensitivity (86.6% [97 of 112] and 93.8% [105 of 112] for summed PI-RADS scores ≥ 9 and Likert scale scores ≥ 3, respectively) and specificity (82.4% [412 of 500] and 73.6% [368 of 500], respectively) values observed in the validation set are likely to be valuable in identifying suspicious locations, while allowing confident prediction of the absence of cancer (negative predictive value) would help to control unnecessary procedures. Comparable characteristics were reported for repeat biopsy (19), supporting multiparametric MR imaging irrespective of the context of initial or repeat biopsy. Moreover, multiparametric MR imaging characteristics, as evaluated by using the PI-RADS and Likert scoring systems, also correlated with aggressive features in the cores, suggesting that diagnostic strategies based on multiparametric MR imaging findings could also be instrumental in controlling the current state of overdetection of incidental cancers.

However, insufficient evidence on the consistency of scoring between readers could impair the diffusion of the PI-RADS system. To our knowledge, a single prior study (30) that used in-bore 3.0-T MR imaging-guided biopsy as the reference standard addressed this question and showed that good interobserver agreement for cancer detection was observed between three senior radiologists. This level of performance was confirmed for 1.5-T MR imaging in the present study, where both the summed PI-RADS scores and the Likert scale scores achieved a good level of agreement between two senior radiologists (κ = 0.73

Table 4

Detection Yields of Threshold Summed PI-RADS and Likert Scale Scores as Defined after the Procedure in DICOM Archives

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Summed PI-RADS Scores</th>
<th>PValue*</th>
<th>Likert Scale Scores</th>
<th>PValue*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distribution</td>
<td>&lt;9</td>
<td>≥9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of cores positive for cancer</td>
<td>1210 (69.9)</td>
<td>521 (30.1)</td>
<td>&lt;.00001 (29.5 [20.1, 41.9])</td>
<td>1067 (61.6)</td>
</tr>
<tr>
<td>No. of cores with clinically important cancer</td>
<td>42/1210 (3.5)</td>
<td>268/521 (51.4)</td>
<td>&lt;.00001 (35.4 [23.3, 53.9])</td>
<td>16/1067 (1.5)</td>
</tr>
</tbody>
</table>

Note.—Unless otherwise specified, data in parentheses are percentages.

* Calculated with the Mann-Whitney nonparametric test.
† According to Harnden et al (22), clinically important cancers have a Gleason score of 4 or 5.
‡ According to Harnden et al (22), cancers that are not clinically important are <3 mm long and do not have a Gleason score of 4 or 5.

Table 5

Multiparametric MR Imaging Characteristics of 310 Positive Cores as a Function of Presence or Absence of Aggressive Features

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Gleason Score of 6 (n = 144)</th>
<th>PValue*</th>
<th>Correlation with Cancer Length in Core</th>
<th>PValue</th>
<th>Not Clinically Important (n = 50)</th>
<th>Clinically Important (n = 260)</th>
<th>PValue*</th>
</tr>
</thead>
<tbody>
<tr>
<td>PI-RADS score at T2-weighted imaging</td>
<td>3.6 ± 0.9</td>
<td>4.2 ± 0.9</td>
<td>&lt;.0001</td>
<td>0.30</td>
<td>&lt;.0001</td>
<td>3.4 ± 1.2</td>
<td>4.0 ± 0.9</td>
</tr>
<tr>
<td>(from 1 to 5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PI-RADS score at DW imaging (from 1 to 5)</td>
<td>3.7 ± 1.3</td>
<td>4.4 ± 1.0</td>
<td>&lt;.0001</td>
<td>0.33</td>
<td>&lt;.0001</td>
<td>3.4 ± 1.5</td>
<td>4.2 ± 1.1</td>
</tr>
<tr>
<td>PI-RADS score at dynamic contrast-enhanced</td>
<td>3.7 ± 1.3</td>
<td>4.1 ± 1.3</td>
<td>.0004</td>
<td>0.40</td>
<td>&lt;.0001</td>
<td>3.2 ± 1.6</td>
<td>4.1 ± 1.2</td>
</tr>
<tr>
<td>imaging (from 1 to 5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summed PI-RADS score (from 3 to 15)</td>
<td>11.1 ± 3.3</td>
<td>12.7 ± 2.9</td>
<td>.0009</td>
<td>0.42</td>
<td>&lt;.0001</td>
<td>10.2 ± 3.9</td>
<td>12.3 ± 2.9</td>
</tr>
<tr>
<td>Likert score (from 1 to 5)</td>
<td>3.9 ± 1.2</td>
<td>4.5 ± 0.9</td>
<td>&lt;.0001</td>
<td>0.39</td>
<td>&lt;.0001</td>
<td>3.6 ± 1.3</td>
<td>4.3 ± 1.0</td>
</tr>
</tbody>
</table>

* Calculated with the Mann-Whitney nonparametric test.
† Calculated with the Spearman r test.
‡ According to Hamden et al (22), cancers that are not clinically important are <3 mm long and do not have a Gleason score of 4 or 5; cancers that are clinically important are ≥3 mm long and/or have a Gleason score of 4 or 5.
and 0.80, respectively). As expected, the Likert scale that was part of the inclusion process (Likert ≥ 3) showed better consistency in terms of κ score and proportion of agreement than did the summed PI-RADS scores.

While good interobserver agreement was observed for T2-weighted imaging (κ = 0.61; agreement, 68%) and dynamic contrast-enhanced imaging (κ = 0.71; agreement, 62%), only fair results were observed for DW imaging (κ = 0.53; agreement, 58%). This might reflect imprecisions in the criteria used for DW image scoring. For instance, ADCs of the normal peripheral zone differ substantially between patients and, in patients with cancer, correlate with the ADCs of high-grade tumors (31). They are also affected by the typical heterogeneity in density of prostate cancer (32) and by parameters related to the MR imaging unit. As a consequence, although it is a quantitative parameter, the use of fixed thresholds was not encouraged in the PI-RADS scoring system, which recommends evaluating reduction in ADC by parameters related to the MR imaging unit. As a consequence, although it is a quantitative parameter, the use of fixed thresholds was not encouraged in the PI-RADS scoring system, which recommends evaluating reduction in ADC and hyperintensity in areas that show reduced ADCs to when segmentation algorithms are used for interpretation. Some subjectivity also applies to the present article: institution receives money from patent WO2009071766 A8. Other relationships: disclosed no relevant relationships. E.B. disclosed no relevant relationships. B.M. disclosed no relevant relationships.

Disclosures of Conflicts of Interest: R.R. disclosed no relevant relationships. P.M. Activities related to the present article: institution received grant ANR-11-TECS-0017 from Projet Prosho. Activities not related to the present article: institution receives money from patent WO2009071766 A8. Other relationships: disclosed no relevant relationships. E.B. Activities related to the present article: disclosed no relevant relationships. F.C. Activities related to the present article: disclosed no relevant relationships. N.B. Activities related to the present article: disclosed no relevant relationships. Janssen, and Astellas. Other relationships: disclosed no relevant relationships. W.O. Activities related to the present article: disclosed no relevant relationships. W.O. Activities related to the present article: disclosed no relevant relationships. E.B. disclosed no relevant relationships. B.M. disclosed no relevant relationships.

References


14. Tan N, Lane BR, Li J, Moussa AS, Soriano M, Jones JS. Prostate cancers diagnosed at repeat biopsies are smaller and less likely to be high grade. J Urol 2008;180(4):1325–1329; discussion 1329.


