Analysis of interreader agreement in structured reports of pelvic multiparametric magnetic resonance imaging using the METastasis Reporting and Data System for Prostate Cancer guidelines

Xiang Liu
Tingting Xie
Zhaonan Sun
Ying Guo
Xiaodong Zhang
Xiaoying Wang

PURPOSE

To evaluate interreader agreement on pelvic multiparametric magnetic resonance imaging (mp-MRI) interpretation among radiologists using a structured reporting tool based on the METastasis Reporting and Data System for Prostate Cancer (MET-RADS-P) guidelines.

METHODS

A structured report for follow-up pelvic mpMRI for advanced prostate cancer (APC) patients was formulated based on MET-RADS-P guidelines. In total, 163 paired pelvic mpMRI examinations were performed from December 2017 to February 2021 on 105 patients with APC. These were retrospectively reviewed by two senior and two junior radiologists for metastatic lesion detection and were categorized by these readers using primary/secondary response assessment categories (RACs), with and without the structured report. Interreader agreement regarding metastasis detection and RAC scores was evaluated with Cohen's kappa and weighted Cohen's kappa statistics (K), respectively.

RESULTS

The two senior radiologists showed higher agreement with the reference standard for metastasis detection using the structured report (S1: $K = 0.83$; S2: $K = 0.73$) compared with the conventional report (S1: $K = 0.72$; S2: $K = 0.61$). Junior radiologists showed similar results (J1: 0.66 vs. 0.59; J2: 0.65 vs. 0.57). The overall agreement between the two senior radiologists was excellent for the primary RAC pattern using the structured reports ($K = 0.81$) and was substantial for secondary RAC categorization ($K = 0.75$). The interreader agreement of the two junior radiologists was substantial for both primary and secondary RAC values ($K = 0.76, 0.68$).

CONCLUSION

Good interreader agreement was found for the follow-up assessment of APC patients between radiologists, where the pelvic mpMRI was reported using MET-RADS-P guidelines. This improvement applied to both metastatic lesion detection and qualitative RAC assessment.

KEYWORDS

Advanced prostate cancer, interreader agreement, METastasis Reporting and Data System for Prostate Cancer, mpMRI, response assessment category

From the Department of Radiology (X.L., Z.S., Y.G., X.Z., X.W. ✉wangxiaoying@bjmu.edu.cn), Peking University First Hospital, Beijing, China; Department of Radiology (T.X.), Peking University Shenzhen Hospital, Shenzhen, China.

Imaging to describe the metastatic status of patients is the cornerstone for managing biomarker development and therapeutic clinical tests.[1] The imaging of biomarkers can provide information on disease distribution, likely prognosis, therapy-induced changes, and response duration.[2]

Whole-body magnetic resonance imaging (MRI) is now an imaging tool that enables tumor detection and therapy evaluations in patients with advanced prostate cancer (APC). The METastasis Reporting and Data System for Prostate Cancer (MET-RADS-P) is a recently

published practical guide that provides the minimum standards for whole-body MRI scans for image acquisition, interpretation, and reporting of baseline and follow-up APC patients, and it enables the documentation of response heterogeneity using response assessment categories (RACs) at the regional level.[2,3] More importantly, the MET-RADS-P score has been confirmed to be a prognostic imaging biomarker, as it stratifies the cancer-specific survival of patients with castration-resistant prostate cancer (PCa).[4]

One of the important purposes of the MET-RADS-P guide is to ensure the uniformity of imaging interpretations. To date, limited data are available on the interreader agreement of MRI examination reports when using MET-RADS-P guidelines. Pricolo et al.[5] found excellent interobserver agreement for the RAC assessment of bone between a senior radiologist and resident radiologist when using the MET-RADS-P guidelines, but results were mixed for other body regions that relied on limited paired whole-body MRI examinations. Therefore, further improvement of interreader agreement needs to be addressed. Consistency can often be improved through training, while another solution is to use a structured reporting tool.

Structured reports are popular in clinical radiology workflows and have shown great potential in improving practical workflows by providing professional, well-defined, and consistent report templates. Dimarco et al.[6] confirmed that structured reports improved the interreader agreement of pancreatic ductal adenocarcinoma staging compared with free-text reports. Therefore, this study hypothesized that a structured report could also improve the interreader agreement on the metastatic evaluation of PCa, using MET-RADS-P guidelines.

Given that PCa initially and predominantly metastasizes to pelvic lymph nodes and bone, and that extrapelvic metastases in the absence of pelvic involvement are rare,[7,8] a routine pelvic examination is adequate for the metastasis detection and response evaluation for PCa patients.[9,10] In this setting, the researchers tailored a structured report for follow-up pelvic multiparametric MRI (mpMRI), taking the MET-RADS-P template for reference. This was used with a cohort of patients with APC at the study site, mimicking a typical clinical workflow. The purpose of this study was to evaluate the interreader agreement of pelvic mpMRI interpretation among radiologists using a structured reporting tool based on MET-RADS-P guidelines.

Methods

Study participants

This retrospective study was approved by the Peking University First Hospital Institutional Review Board, and informed consent was obtained from all patients in written form (2021-060).

The inclusion criteria for patients in this study included a histologic diagnosis of PCa, with metastatic lesions presented in previous and ongoing follow-up pelvic mpMRI examinations at the institution. Only patients that had a complete pelvic mpMRI dataset before and after systemic therapy were included. The study excluded patients who had an incomplete pelvic mpMRI protocol (n = 7), poor image quality (n = 5), and absent clinical information (n = 11).

In total, 163 pairs of pelvic mpMRI examinations were gathered for analysis. These were performed on 105 patients with APC who had undergone at least two examinations between December 2017 and February 2021 for follow-up assessment after cancer therapy. All patients underwent baseline scanning before therapy. Among them, 58 patients had one follow-up examination (116 scans total, 58 examination pairs), 36 patients had two follow-up examinations (108 scans total, 72 examination pairs), and 11 patients had three follow-up examinations (44 scans total, 33 examination pairs). Pre-MRI clinical information [age, prostate-specific antigen (PSA) values, and therapy method] was collected for all patients.

Imaging technique

All pelvic mpMRI images were acquired on two 3.0 T MRI scanners (Discovery, GE Healthcare; Intera, Philips Healthcare) using an acquisition protocol that complies with the MET-RADS-P standard. The imaging protocol consisted of multiplanar T1-/T2-weighted imaging and diffusion-weighted imaging with b values of 800–1,000 s/mm² along

with reconstructed apparent diffusion coefficient maps. The T1-weighted imaging was obtained using the Dixon technique with in-phase and out-of-phase and three-dimensional dynamic contrast-enhanced MRI.[3] For patients who had previously undergone prostatic biopsies, mpMRI examinations were performed at least four weeks after the latest biopsy.

MET-RADS-P system

The MET-RADS-P system assigned the presence of clearly identified disease to 14 predefined regions of the body (the primary disease site, seven skeletal and three nodal regions, and lung, liver, and other soft tissue sites); this was used at baseline and follow-up assessments according to the morphological and signal characteristics on all acquired images. For each anatomic region of metastasis, a qualitative response assessment on a scale of RAC 1 to 5 (1: highly likely to be responding; 2: likely to be responding, 3: stable; 4: likely to be progressing; 5: highly likely to be progressing) was recorded and compared with the baseline study.[3]

Structured report template

A structured report for follow-up pelvic mpMRI for patients with APC was formulated in line with the MET-RADS-P guidelines by two urinary radiologists (with 4 and 15 years of experience in urinary radiology, respectively) (Figure 1). The structured report template consists of four sections: 1) clinical evaluation: a statement regarding the patient's clinical performance, prior treatment methods, current pathological status, and prior/current PSA level; 2) imaging technique: details of the pelvic mpMRI technique, including the imaging protocol and quality [notably, obvious deviations in techniques and artefacts should be recorded with their causes (e.g., metal implant artefacts, patient movement)]; 3) key radiological findings: the presence of metastasis and the RAC scores for each pelvic region (including primary disease, skeletal pelvis, lymph nodes, seminal vesicles, rectum, and bladder) based on the baseline and follow-up examination; and 4) diagnostic impression: an overall diagnostic impression.

Image interpretation

All examinations were retrospectively and independently reviewed, interpreted, and scored according to MET-RADS-P guidelines by two senior radiologists (all with six years of experience in urinary radiology) and two junior radiologists (all with three years of

1. Clinical evaluation

1.1 Pathological result

Time _____   Result _____

1.2 Prior and current PSA level

Time	PSA level

1.3 Prior and current therapy

○ radiotherapy _____
○ chemotherapy _____
○ endocrine therapy _____
○ Other _____

1.4 Prior MRI

Primary disease _____
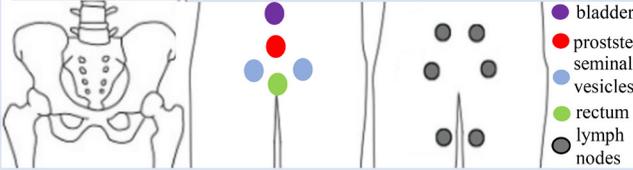Metastatic state _____

2. Technique evaluation

Imaging protocol ☐ T2WI ☐ T1WI ☐ DWI ☐ ADC

Imaging quality ☐ incomplete mpMRI protocol
☐ Unqualified scanning parameter
☐ Image artifact ○ Metal artifact
○ Motion artifact

3. Radiological findings

☐ Primary disease   ☐ Skeletal pelvis
☐ Involved ○ Yes ○ No   ☐ Involved ○ Yes ○ No
☐ RAC-1 ___ ☐ RAC-2 ___   ☐ RAC-1 ___ ☐ RAC-2 ___

☐ Lymph nodes   ☐ Seminal vesicles
☐ Involved ○ Yes ○ No   ☐ Involved ○ Yes ○ No
☐ RAC-1 ___ ☐ RAC-2 ___   ☐ RAC-1 ___ ☐ RAC-2 ___

☐ Rectum   ☐ Bladder
☐ Involved ○ Yes ○ No   ☐ Involved ○ Yes ○ No
☐ RAC-1 ___ ☐ RAC-2 ___   ☐ RAC-1 ___ ☐ RAC-2 ___

● bladder
● prostste
● seminal vesicles
● rectum
● lymph nodes

4. Diagnostic impression

Figure 1. The structured report template of follow-up pelvic multiparametric magnetic resonance imaging based on the METastasis Reporting and Data System for Prostate Cancer guidelines.

experience in urinary radiology). To reproduce the typical pelvic mpMRI interpretation workflow as much as possible, the four radiologists could obtain access to retrospective MRI examinations on a PACS workstation. For follow-up assessment after therapy, the reports of prior examinations and all clinical information were made available to the radiologists.

Each radiologist read the same pair of MRI examinations twice with and without the structured report template (a structured and conventional report, respectively), with a one-month washout period (each radiologist read the assigned MRI scans without the structured report template for the first time and with the structured report template for the second time after one month). After image interpretation, the presence or absence of metastasis was noted for each anatomic region, and two RAC values between 1–5 (Figures 2, 3) were recorded for the primary and secondary metastatic regions in case of the heterogeneity of responses according to the MET-RADS-P guidelines. The primary RAC value is based on the predominant pattern (more than half of the lesions) of response within the region. The secondary RAC value represents the second most common response pattern within the regions (when assessing a single lesion in a region, the secondary RAC value is exempt). A radiology expert (with more than 15 years of reading experience) reviewed and evaluated all pelvic mpMRI examinations to indicate the reference standard.

Statistical analysis

After testing, the data was found to be not normally distributed. As such, clinical data (including the age and PSA level of the patient cohort) are represented as medians and interquartile ranges. The interreader agreement between the radiologists for region-based metastatic lesion detection was evaluated by Cohen's kappa statistics (K). The primary and secondary RAC scores for each region were evaluated using weighted Cohen's kappa statistics (K).[11,12] Interreader agreement was interpreted as none to slight ($K < 0.20$), fair (K: 0.21–0.40), moderate (K: 0.41–0.60), substantial (K: 0.61–0.80), or excellent (K: 0.81–1.00). Statistical analysis was carried out with SPSS software (version 23.0, IBM Corp., Armonk, NY, USA). Statistical significance was set at $P < 0.05$.
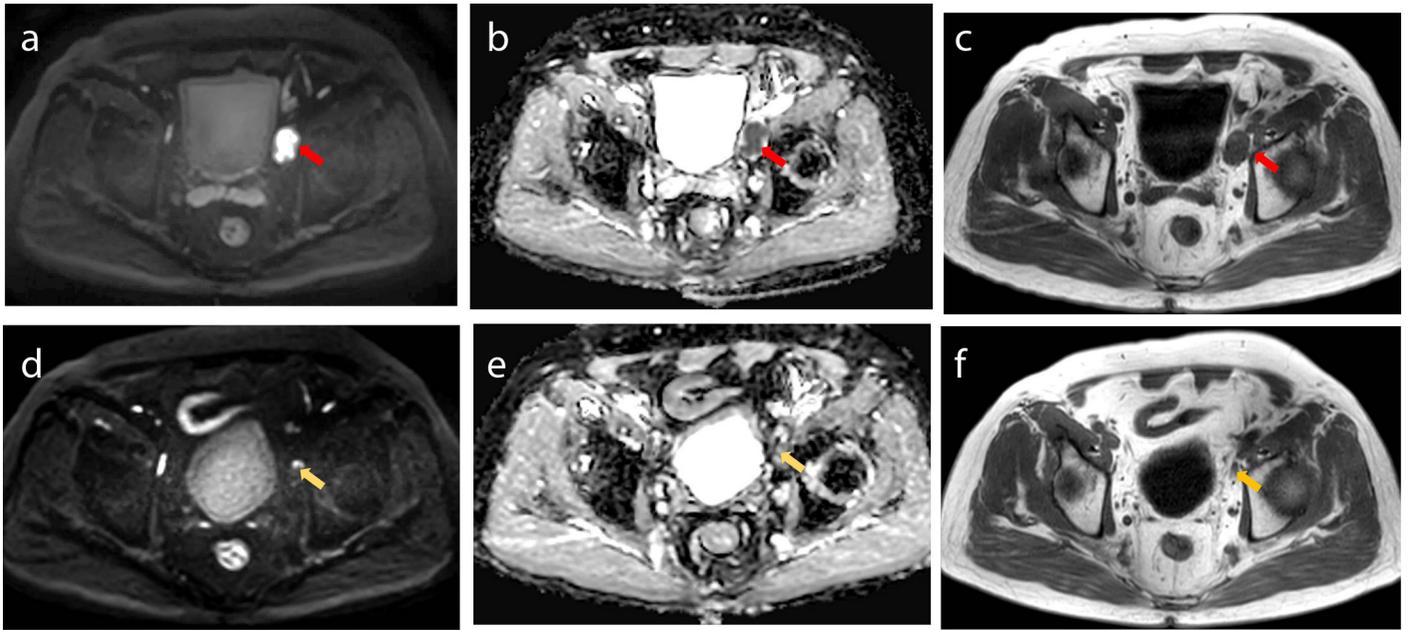
Figure 2. Example images of a 72-year-old patient with advanced prostate cancer in response to endocrine therapy, with primary response assessment category 1. **(a-c)** Axial diffusion-weighted images (DWI) **(a)**, apparent diffusion coefficient (ADC) maps **(b)**, and T1 weighted images (T1WI) **(c)** before therapy show a metastatic lymph node (red arrow) within the region of the left internal iliac artery. **(d-f)** The metastatic lymph node (yellow arrow) shrinks on DWI **(d)**, ADC maps **(e)**, and T1WI images **(f)** after endocrine therapy.
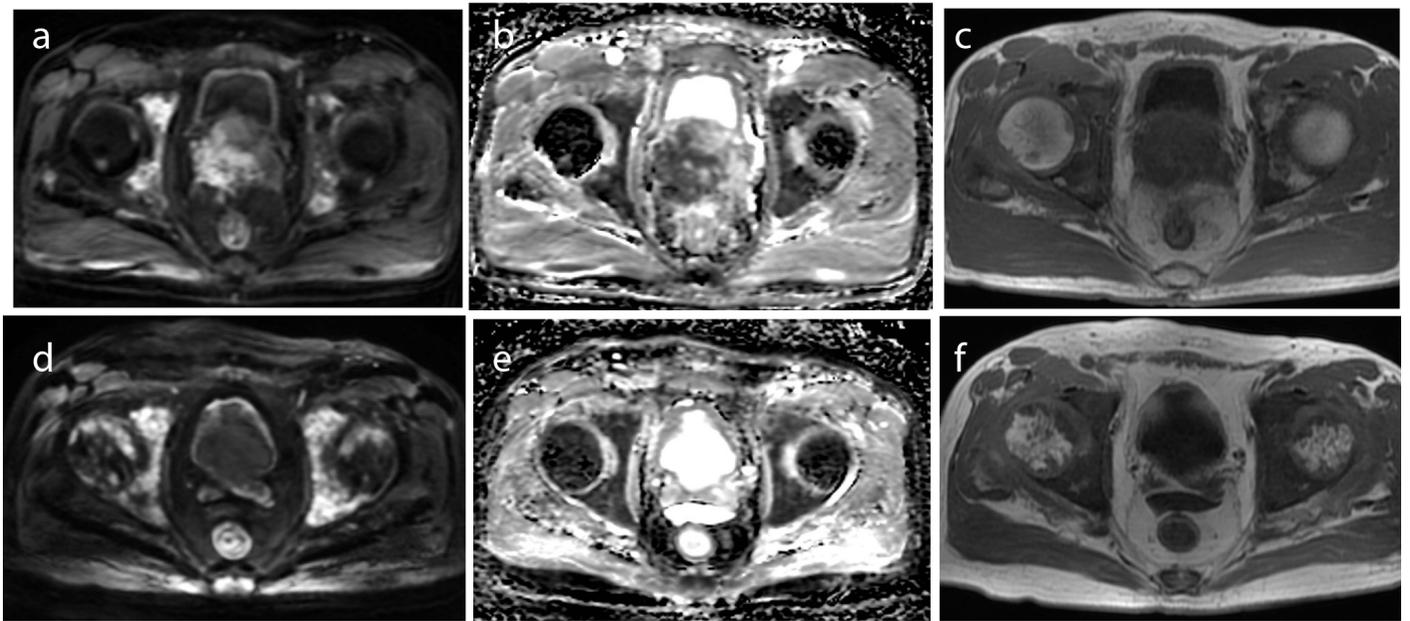
Figure 3. Example images of a 78-year-old advanced prostate cancer patient who is progressing, with primary response assessment category 5. **(a-c)** Axial diffusion-weighted images (DWI) **(a)**, apparent diffusion coefficient (ADC) maps **(b)**, and T1 weighted images (T1WI) **(c)** at the baseline pelvic show multiparametric magnetic resonance imaging (MRI) the presence of diffuse lesions in pelvic bones. **(d-f)** The lesions are more extensive in the follow-up MRI of DWI **(d)**, ADC maps €, and T1WI images **(f)**.

Table 1. Baseline characteristics of the study cohort (n = 105)

Parameters	
Age (y)	76 (68, 82)
PSA* level (IQR; ng/mL)	26.33 (3.56, 74.94)
Post-MRI therapy method (no. of patients)	
Radiotherapy	21
Chemotherapy	20
Endocrine therapy	20
Radical prostatectomy	24
Endocrine therapy + radiotherapy	10
Radical prostatectomy + radiotherapy	10
Gleason score# (no. of patients)	
4 + 3	13
4 + 4	10
4 + 5	11
Distributed site of pelvic metastases (no. of metastatic lesions)	
Rectum	24
Bladder	31
Lymph nodes	103
Skeletal pelvis	80
Seminal vesicles	37

*The PSA determinations were performed less than two weeks before baseline pelvic mpMRI scanning. #The Gleason score was only available for patients who underwent radical prostatectomy. PSA, prostate specific antigen; IQR, interquartile range; MRI, magnetic resonance imaging; mpMRI, multiparametric magnetic resonance imaging.

Results

Patient demographics

The clinical, radiological, and pathological characteristics of the study cohort at the time of inclusion are summarized in Table 1. For 34 of the 105 patients who had undergone radical prostatectomy, a Gleason score (GS) of 4 + 3 was the most common pattern (38%, n = 13), followed by a GS of 4 + 5 (33%, n = 11) and a GS of 4 + 4 (29%, n = 10). The remaining patients were treated with radiotherapy (n = 21), chemotherapy (n = 20), endocrine therapy (n = 20), and endocrine therapy combined with radiotherapy (n = 10).

The number and distribution of metastatic sites at the baseline MRI scanning are shown in Table 1. A total of 275 regions were inspected across the 163 mpMRI examinations, and lymph nodes were the most frequent regions of metastasis (n = 103), followed by bone (n = 80) and seminal vesicles (n = 37). Of the 163 mpMRI examinations, 90 were found to have multiple metastases, and 36 of 90 patients had both lymph node and skeletal pelvis metastases. A more detailed distribution of metastatic sites is shown in the Supplementary Material (Supplementary Table 1).

Detection of metastatic lesions

As all the patients included in this study had metastatic APC, lesion detection of the primary disease at the prostate site was not analyzed here.

As shown in Table 2, two senior radiologists reported the presence of metastasis in a total of 272 and 278 cases with the conventional report, and 263 and 281 with the structured report, respectively. When using the structured report, the two senior radiologists showed substantial to excellent agreement [K values: S1 vs. reference: 0.83 (0.79–0.88); S2 vs. reference: 0.73 (0.68–0.78)] regarding the reference standard for metastatic lesion detection within the five regions. This value was higher than that of the radiologists using the conventional report [K values: S1 vs. reference: 0.72 (0.67–0.77); S2 vs. reference: 0.61 (0.56–0.67)]. In addition, the interreader agreement between the two senior radiologists improved from substantial [K value of conventional report: 0.77 (0.72–0.81)] to excellent with the structured report [K value: 0.84 (0.79–0.88)].

The two junior radiologists reported the presence of metastasis in a total of 299 and 317 cases with the conventional report, and in 301 and 292 cases with the structured report, respectively. Similar to senior radiologists, structured reports improved the diagnostic accuracy of metastatic lesions and interreader agreement compared with conventional reports. The two junior radiologists showed substantial agreement [K values: J1 vs. reference: 0.66 (0.60–0.71); J2 vs. reference: 0.65 (0.60–0.71)] regarding the reference standard for metastasis detection using the structured report. This value was higher than that of the radiologists using the conventional report [K values: J1 vs. reference: 0.59 (0.53–0.65); J2 vs. reference: 0.57 (0.51–0.63)]. In addition, the interreader agreement between the two junior radiologists improved from moderate [K value of conventional report: 0.58 (0.52–0.64)] to substantial with the structured report [K value: 0.69 (0.64–0.74)]. A more detailed number and distribution of metastatic regions for the 163 mpMRI examinations are provided in the Supplementary Material (Supplementary Table 2).

Assessment of primary RAC categorization

Considering that the structured report of pelvic mpMRI based on MET-RADS-P guidelines performed better than the conventional report in lesion detection, the researchers further analyzed its effect on RAC categorization for the two senior radiologists and two junior radiologists.

As shown in Table 3, the two senior radiologists achieved high agreement with

Table 2. The diagnostic accuracy and interreader agreement of metastatic lesions

	Rectum	Bladder	Lymph nodes	Skeletal pelvis	Seminal vesicles	Overall
Conventional report (no. of metastatic lesions)						
S1	29	33	101	74	35	272
S2	35	37	93	75	38	278
J1	43	41	95	74	46	299
J2	49	46	98	72	52	317
S1 vs. reference (K, 95% CI)	0.62 (0.45–0.78)	0.57 (0.41–0.73)	0.69 (0.57–0.80)	0.75 (0.65–0.85)	0.68 (0.54–0.82)	0.72 (0.67–0.77)
S2 vs. reference (K, 95% CI)	0.49 (0.32–0.66)	0.43 (0.27–0.60)	0.62 (0.50–0.74)	0.62 (0.50–0.74)	0.57 (0.42–0.72)	0.61 (0.56–0.67)
J1 vs. reference (K, 95% CI)	0.47 (0.31–0.62)	0.43 (0.27–0.60)	0.56 (0.43–0.68)	0.61 (0.49–0.73)	0.57 (0.42–0.71)	0.59 (0.53–0.65)
J2 vs. reference (K, 95% CI)	0.40 (0.25–0.55)	0.45 (0.29–0.61)	0.62 (0.50–0.74)	0.58 (0.46–0.70)	0.47 (0.32–0.62)	0.57 (0.51–0.63)
S1 vs. S2 (K, 95% CI)	0.69 (0.55–0.83)	0.66 (0.52–0.80)	0.77 (0.67–0.87)	0.78 (0.71–0.89)	0.70 (0.57–0.83)	0.77 (0.72–0.81)
J1 vs. J2 (K, 95% CI)	0.52 (0.37–0.67)	0.51 (0.36–0.66)	0.55 (0.42–0.68)	0.63 (0.51–0.75)	0.48 (0.33–0.63)	0.58 (0.52–0.64)
Structured report (no. of metastatic lesions)						
S1	26	30	101	73	33	263
S2	31	36	103	76	35	281
J1	39	46	92	75	49	301
J2	37	39	94	81	41	292
S1 vs. reference (K, 95% CI)	0.67 (0.51–0.83)	0.78 (0.65–0.90)	0.84 (0.76–0.93)	0.87 (0.79–0.94)	0.78 (0.66–0.90)	0.83 (0.79–0.88)
S2 vs. reference (K, 95% CI)	0.63 (0.42–0.75)	0.64 (0.50–0.80)	0.76 (0.66–0.87)	0.71 (0.60–0.81)	0.68 (0.54–0.82)	0.73 (0.68–0.78)
J1 vs. reference (K, 95% CI)	0.48 (0.31–0.64)	0.55 (0.40–0.69)	0.69 (0.46–0.71)	0.82 (0.73–0.90)	0.62 (0.49–0.76)	0.66 (0.60–0.71)
J2 vs. reference (K, 95% CI)	0.48 (0.21–0.55)	0.53 (0.37–0.69)	0.68 (0.57–0.79)	0.72 (0.61–0.83)	0.60 (0.45–0.74)	0.65 (0.60–0.71)
S1 vs. S2 (K, 95% CI)	0.81 (0.69–0.93)	0.77 (0.65–0.89)	0.82 (0.72–0.91)	0.87 (0.72–0.91)	0.81 (0.70–0.93)	0.84 (0.79–0.88)
J1 vs. J2 (K, 95% CI)	0.67 (0.54–0.79)	0.55 (0.40–0.69)	0.70 (0.54–0.77)	0.75 (0.65–0.85)	0.68 (0.56–0.80)	0.69 (0.64–0.74)

S1 and S2 indicate the two senior radiologists; J1 and J2 indicate the two junior radiologists. Values in parenthesis represent a 95% confidence interval (CI).

Table 3. The interreader agreement of primary RAC categorization (K, 95% CI)

	Primary disease*	Rectum	Bladder	Lymph nodes	Skeletal pelvis	Seminal vesicles	Overall
S1 vs. reference	0.76 (0.63–0.82)	0.73 (0.70–0.91)	0.86 (0.74–0.95)	0.78 (0.69–0.86)	0.71 (0.62–0.83)	0.82 (0.69–0.95)	0.77 (0.60–0.94)
S2 vs. reference	0.78 (0.63–0.86)	0.82 (0.73–0.92)	0.85 (0.71–0.93)	0.79 (0.65–0.94)	0.74 (0.65–0.79)	0.79 (0.64–0.89)	0.76 (0.62–0.93)
J1 vs. reference	0.68 (0.56–0.76)	0.78 (0.69–0.87)	0.80 (0.74–0.92)	0.72 (0.61–0.84)	0.65 (0.53–0.71)	0.71 (0.57–0.84)	0.67 (0.53–0.85)
J2 vs. reference	0.61 (0.53–0.74)	0.71 (0.65–0.84)	0.73 (0.58–0.82)	0.74 (0.60–0.82)	0.63 (0.52–0.75)	0.74 (0.5–0.84)	0.69 (0.51–0.83)
S1 vs. S2	0.80 (0.63–0.91)	0.87 (0.68–0.96)	0.90 (0.73–1.00)	0.85 (0.62–0.94)	0.72 (0.61–0.83)	0.83 (0.61–0.92)	0.81 (0.70–0.96)
J1 vs. J2	0.71 (0.60–0.81)	0.78 (0.61–0.92)	0.79 (0.64–0.90)	0.81 (0.60–0.95)	0.68 (0.54–0.85)	0.76 (0.62–0.82)	0.76 (0.61–0.85)

*The response assessment category (RAC) evaluation of the primary disease was only performed for patients who had not undergone radical prostatectomy. S1 and S2 indicate the two senior radiologists; J1 and J2 indicate the two junior radiologists. Values in parenthesis represent a 95% confidence interval (CI).

Table 4. The interreader agreement of secondary RAC categorization (*K*, 95% CI)

	Primary disease*	Rectum	Bladder	Lymph nodes	Skeletal pelvis	Seminal vesicles	Overall
S1 vs. reference	0.66 (0.51–0.73)	0.71 (0.58–0.80)	0.78 (0.62–0.94)	0.72 (0.56–0.93)	0.64 (0.54–0.73)	0.76 (0.62–0.85)	0.71 (0.53–0.96)
S2 vs. reference	0.71 (0.60–0.82)	0.74 (0.51–0.93)	0.77 (0.63–0.89)	0.70 (0.57–0.91)	0.66 (0.52–0.78)	0.73 (0.52–0.88)	0.72 (0.54–0.93)
J1 vs. reference	0.53 (0.46–0.74)	0.64 (0.49–0.77)	0.65 (0.54–0.82)	0.69 (0.51–0.80)	0.59 (0.41–0.67)	0.62 (0.45–0.81)	0.58 (0.41–0.72)
J2 vs. reference	0.57 (0.44–0.67)	0.62 (0.50–0.75)	0.63 (0.46–0.81)	0.66 (0.52–0.79)	0.56 (0.45–0.76)	0.61 (0.49–0.75)	0.59 (0.43–0.74)
S1 vs. S2	0.78 (0.65–0.90)	0.75 (0.63–0.89)	0.79 (0.71–0.93)	0.82 (0.61–0.93)	0.73 (0.64–0.91)	0.77 (0.60–0.87)	0.75 (0.61–0.87)
J1 vs. J2	0.67 (0.56–0.82)	0.75 (0.61–0.88)	0.74 (0.59–0.86)	0.72 (0.60–0.83)	0.63 (0.56–0.73)	0.63 (0.47–0.74)	0.68 (0.53–0.86)

*The response assessment category (RAC) evaluation of the primary disease was only performed for patients who had not undergone a radical prostatectomy. S1 and S2 indicate the two senior radiologists; J1 and J2 indicate the two junior radiologists.
Values in parenthesis represent a 95% confidence interval (CI).

the reference standard for the primary RAC values [*K* values: S1 vs. reference: 0.77 (0.60–0.94); S2 vs. reference: 0.76 (0.62–0.93)]. The overall agreement between the two senior radiologists for the primary RAC pattern was excellent [*K* value: 0.81 (0.70–0.96)]. For the two junior radiologists, the agreement using the reference standard was substantial [*K* value: J1 vs. reference: 0.67 (0.53–0.85); J2 vs. reference: 0.69 (0.51–0.83)]. The overall interreader agreement between the two junior radiologists had a *K* value of 0.76 (0.61–0.85).

Assessment of secondary RAC categorization

As shown in Table 4, for the four radiologists, the agreement was substantial for S1/S2 and the reference standard [*K* values: S1 vs. reference: 0.71 (0.53–0.96); S2 vs. reference: 0.70 (0.54–0.93)], and moderate for J1/J2 and the reference standard [*K* values: J1 vs. reference: 0.58 (0.41–0.72); J2 vs. reference: 0.59 (0.43–0.74)]. The interreader agreement was substantial for both senior and junior radiologists [S1 vs. S2: 0.75 (0.61–0.87); J1 vs. J2: 0.72 (0.53–0.86)]. The primary and secondary RAC values for each region are summarized in the Supplementary Material (Supplementary Table 3).

Discussion

In this study, the researchers developed a structured report for pelvic mpMRI using MET-RADS-P guidelines and investigated its reproducibility among multiple radiologists on a large cohort of patients with APC who underwent pelvic mpMRI for follow-up evaluation. The results showed that both the senior and junior radiologists performed better when using the structured report than when using the conventional report for metastasis detection, and high interreader agreement

regarding lesion detection and RAC categorization was found when using the structured report. As expected, the level of interreader agreement was generally higher between senior radiologists than between junior radiologists.

As novel whole-body imaging techniques, whole-body MRI and positron emission tomography/computed tomography (PET/CT) are known for being more accurate for evaluating the treatment responses of patient with APC with bone disease, compared with bone scanning and CT.[13] Whole-body MRI has been noted to provide clear categorization of bone metastasis response and is suggested to be suitable for wide deployment in disease detection settings,[14,15] given its established diagnostic accuracy, wide availability, and multi-organ evaluation capabilities.[13,16] However, in terms of the follow-up treatment evaluation, whole-body MRI is probably not a better technique compared with PET techniques, which are ahead in this specific domain.[17,18] Compared with whole-body MRI, several studies have confirmed the advantages of prostate-specific membrane antigen (PSMA) PET/CT for evaluating disease progression and treatment responses.[19,20] Such research shows that PSMA PET/CT promises to become a powerful alternative to whole-body MRI, assuming the limitations of ionizing radiation exposure and spatial resolution are solved.[21]

The subjective criteria applied for assessing metastatic lesions using whole-body MRI may result in unsatisfactory interreader concordance. The MET-RADS-P guidelines were designed to minimize the inconsistencies caused by various reading criteria.[3] However, for radiologists, especially junior radiologists, the MET-RADS-P guidelines are

too complex to use effectively. The *K* value of the interreader agreement between two radiologists varied from 0.56–1.0 (primary RAC) and 0.44–0.93 (secondary RAC) among different regions when using the MET-RADS-P guidelines.[5] By creating a structured report of follow-up pelvic mpMRI according to the standardization requirements of MET-RADS-P and actual clinical work experience of the unit, the study found improved interreader agreement for both RAC assessments compared with conventional reports. This is crucial for the follow-up evaluation of APC patients, as follow-up pelvic mpMRI examinations are usually reviewed by different medical staff at different periods. In addition, in an analysis of body regions, both senior and junior radiologists showed the highest diagnostic accuracy for metastasis detection in the regions of the skeletal pelvis and lymph node with or without structured reports. This may be attributed to the pelvic lymph node and bone metastases being present in most of the enrolled APC patients and the metastases within the two regions often appearing in the form of multiple metastases.

The RAC value provides a qualitative response assessment category for each anatomic region by comparing the alterations of the metastatic lesions between baseline and follow-up examinations. This study found that the interreader agreement between the two junior radiologists for primary RAC in the skeletal region was slightly lower than that in other regions. The reason for this finding is probably due to the different assessment criteria for bones and soft tissues. For response assessments of soft tissues (prostate, bladder, rectum, lymph nodes, and seminal vesicles), the RAC assessment standard was based on the prescribed and established RECIST guidance.[14,22] For bone disease (skeletal pelvis),

the RAC values were summarized using the newly developed MET-RADS criteria,[3] which were more complicated and mainly relied on subjective morphological features, thus affecting the assessment performance and interobserver agreement for bone metastases. Additionally, the interreader agreement on the secondary RAC pattern was slightly lower than that on the primary RAC pattern for both senior and junior radiologists, which may be because the secondary RAC assessment requires readers to be able to identify the response differences present in a small subgroup of metastases. The comparison of the radiologists and reference standard indicated that the accuracy of RAC categorization for senior radiologists was higher than that of junior radiologists, especially for the secondary RAC evaluation. This suggests that the feasibility and accuracy of the structured report for pelvic mpMRI using MET-RADS-P guidelines can be affected by the reader's experience in clinical practice.

In this research, the overall agreement for both primary and secondary RAC assessment between senior radiologists was slightly higher than that between junior radiologists, which differs from the previous study conducted by Pricolo et al.[5] Their results showed high interreader agreement between two readers with different levels of expertise (a senior radiologist with nine years of experience vs. a resident radiologist after six months of training). This may be attributed to the fact that only two readers were involved in their study and the less experienced resident radiologist was trained by the senior radiologist. Compared with the current study's four independent readers, the results of Pricolo et al.'s[5] study may have been affected by selection bias.

The current study had some limitations. First, it was limited to the follow-up analysis of pelvic mpMRI examination for patients with APC by radiologists, while further analysis of the impact on clinical decision-making processes and patient outcomes is not performed here. Therefore, prospective clinical studies are necessary to further consolidate the results. Second, the soft tissue evaluation of the RAC system in this research is tailored to the pelvic region instead of the whole body, which is currently a speculative and tentative application. In addition, although the readers recruited for the study were four independent radiologists, all readers came from the same institution, which may lead them to adopt similar interpretation schemes to reduce a priori variability in clinical assessments. Multicentre studies may be helpful to address this limitation. The final limitation was the weak standard of reference used. A pathology reference standard or comparison with other techniques (such as PSMA PET/CT) would be superior to the expertise used here, as this is considered best practice. However, it was difficult to gain access to the necessary histological/PET information.

In conclusion, a good interreader agreement was found for the follow-up assessment of APC patients between radiologists who had different levels of expertise using the structured report for pelvic mpMRI based on MET-RADS-P guidelines. In particular, the agreement was excellent between senior radiologists in metastatic lesion detection and qualitative RAC assessment. The study shows that interreader agreement can be improved using MET-RADS-P guidelines and provides insights into its clinical significance for the clinical management of metastasis in a growing number of APC patients.

Conflict of interest disclosure

The authors declare no conflicts of interest.

References

1. Halabi S, Kelly WK, Ma H, et al. Meta-analysis evaluating the impact of site of metastasis on overall survival in men with castration-resistant prostate cancer. *J Clin Oncol.* 2016;34(14):1652-1659. [Crossref]

2. Padhani AR, Lecouvet FE, Tunariu N, et al. METastasis Reporting and Data System for Prostate Cancer: practical guidelines for acquisition, interpretation, and reporting of whole-body magnetic resonance imaging-based evaluations of multiorgan involvement in advanced prostate cancer. *Eur Urol.* 2017;71(1):81-92. [Crossref]

3. Padhani AR, Tunariu N. Metastasis Reporting and Data System for Prostate Cancer in Practice. *Magn Reson Imaging Clin N Am.* 2018;26(4):527-542. [Crossref]

4. Yoshida S, Takahara T, Ishii C, et al. METastasis Reporting and Data System for prostate cancer as a prognostic imaging marker in castration-resistant prostate cancer. *Clin Genitourin Cancer.* 2020;18(4):e391-e396. [Crossref]

5. Pricolo P, Ancona E, Summers P, et al. Whole-body magnetic resonance imaging (WB-MRI) reporting with the METastasis Reporting and Data System for Prostate Cancer (MET-RADS-P): inter-observer agreement between readers of different expertise levels. *Cancer Imaging.* 2020;20(1):77. [Crossref]

6. Dimarco M, Cannella R, Pellegrino S, et al. Impact of structured report on the quality of preoperative CT staging of pancreatic ductal adenocarcinoma: assessment of intra- and inter-reader variability. *Abdom Radiol (NY).* 2020;45(2):437-448. [Crossref]

7. Ottosson F, Baco E, Lauritzen PM, Rud E. The prevalence and locations of bone metastases using whole-body MRI in treatment-naïve intermediate- and high-risk prostate cancer. *Eur Radiol.* 2021;31(5):2747-2753. [Crossref]

8. Woo S, Kim SY, Kim SH, Cho JY. Journal club: identification of bone metastasis with routine prostate MRI: a study of patients with newly diagnosed prostate cancer. *AJR Am J Roentgenol.* 2016;206(6):1156-1163. [Crossref]

9. Schmidkonz C, Cordes M, Schmidt D, et al. ^{68}Ga-PSMA-11 PET/CT-derived metabolic parameters for determination of whole-body tumor burden and treatment response in prostate cancer. *Eur J Nucl Med Mol Imaging.* 2018;45(11):1862-1872. [Crossref]

10. Liu X, Han C, Wang H, et al. Fully automated pelvic bone segmentation in multiparameteric MRI using a 3D convolutional neural network. *Insights Imaging.* 2021;12(1):93. [Crossref]

11. Rengo M, Boru CE, Badia S, et al. Preoperative measurement of the hiatal surface with MDCT: impact on surgical planning. *Radiol Med.* 2021;126(12):1508-1517. [Crossref]

12. Miskin N, Gaviola GC, Huang RY, et al. Intra- and intersubspecialty variability in lumbar spine MRI interpretation: a multireader study comparing musculoskeletal radiologists and neuroradiologists. *Curr Probl Diagn Radiol.* 2020;49(3):182-187. [Crossref]

13. Padhani AR, Lecouvet FE, Tunariu N, et al. Rationale for modernising imaging in advanced prostate cancer. *Eur Urol Focus.* 2017;3(2-3):223-239. [Crossref]

14. Scher HI, Morris MJ, Stadler WM, et al. Trial Design and objectives for castration-resistant prostate cancer: updated recommendations from the prostate cancer clinical trials working group 3. *J Clin Oncol.* 2016;34(12):1402-1418. [Crossref]

15. Lecouvet FE. Whole-body MR imaging: musculoskeletal applications. *Radiology.* 2016;279(2):345-365. [Crossref]

16. Lecouvet FE, Talbot JN, Messiou C, et al. Monitoring the response of bone metastases to treatment with magnetic resonance imaging and nuclear medicine techniques: a review and position statement by the European Organisation for Research and Treatment of Cancer imaging group. *Eur J Cancer.* 2014;50(15):2519-2531. [Crossref]

17. Tsechelidis I, Vrachimis A. PSMA PET in imaging prostate cancer. *Front Oncol.* 2022;12:831429. [Crossref]

18. Orcajo-Rincon J, Muñoz-Langa J, Sepúlveda-Sánchez JM, et al. Review of imaging techniques for evaluating morphological and functional responses to the treatment of bone metastases in prostate and breast cancer. *Clin Transl Oncol.* 2022;24(7):1290-1310. [Crossref]

19. Zhou C, Tang Y, Deng Z, et al. Comparison of ^{68}Ga-PSMA PET/CT and multiparametric MRI for the detection of low- and intermediate-risk prostate cancer. *EJNMMI Res.* 2022;12(1):10. [Crossref]

20. Dyrberg E, Hendel HW, Huynh THV, et al. ^{68}Ga-PSMA-PET/CT in comparison with ^{18}F-fluoride-PET/CT and whole-body MRI for the detection of bone metastases in patients with prostate cancer: a prospective diagnostic accuracy study. *Eur Radiol.* 2019;29(3):1221-1230. [Crossref]

21. Mosavi F, Johansson S, Sandberg DT, Turesson I, Sörensen J, Ahlström H. Whole-body diffusion-weighted MRI compared with (18) F-NaF PET/CT for detection of bone metastases in patients with high-risk prostate carcinoma. *AJR Am J Roentgenol.* 2012;199(5):1114-1120. [Crossref]

22. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer.* 2009;45(2):228-247. [Crossref]

Supplementary Table 1. Detailed distribution of pelvic metastases at the time of inclusion	
Regions	Number of lesions
Single region involvement	
Rectum	6
Bladder	8
Lymph nodes	27
Skeletal pelvis	22
Seminal vesicles	10
Multiple region involvement	
Bladder + rectum	2
Lymph nodes + rectum	6
Lymph nodes + bladder	8
Bladder + skeletal pelvis	3
Rectum + skeletal pelvis	2
Bladder + seminal vesicles	2
Seminal vesicles + rectum	1
Lymph nodes + skeletal pelvis	36
Seminal vesicles + skeletal pelvis	3
Lymph nodes + seminal vesicles	8
Lymph nodes + bladder + rectum	1
Lymph nodes + rectum + skeletal pelvis	1
Lymph nodes + bladder+ skeletal pelvis	4
Lymph nodes + seminal vesicles+ rectum	2
Seminal vesicles + rectum + skeletal pelvis	1
Lymph nodes + bladder + seminal vesicles	2
Lymph nodes + seminal vesicles + skeletal pelvis	6
Lymph nodes + seminal vesicles + rectum + skeletal pelvis	1
Lymph nodes + bladder + seminal vesicles + skeletal pelvis	1

Supplementary Table 2. The number and distribution of metastatic regions for the 163 multiparametric magnetic resonance imaging examinations

Distribution	Reference standard	Conventional report				Structured report			
		S1	S2	J1	J2	S1	S2	J1	J2
Single region involvement									
Rectum	6	9	10	7	6	8	9	7	7
Bladder	8	11	11	7	8	11	12	12	9
Lymph nodes	27	31	29	26	24	33	28	14	24
Skeletal pelvis	22	20	21	16	16	18	13	16	16
Seminal vesicles	10	5	8	11	10	11	9	13	10
Multiple region involvement									
Bladder + rectum	2	2	3	1	1	4	4	3	3
Lymph nodes + rectum	6	4	2	8	8	4	4	8	7
Lymph nodes + bladder	8	6	8	8	6	4	4	10	4
Rectum + skeletal pelvis	2	2	2	7	3	1	3	2	2
Seminal vesicles + rectum	1	1	1	1	2	2	1	5	3
Bladder + skeletal pelvis	3	4	4	8	5	2	6	3	4
Bladder + seminal vesicles	2	3	2	3	2	1	2	3	3
Lymph nodes + skeletal pelvis	36	32	25	20	17	38	39	24	29
Seminal vesicles + skeletal pelvis	3	3	3	3	2	4	3	2	8
Lymph nodes + seminal vesicles	8	10	7	7	9	7	5	7	5
Lymph nodes + bladder + rectum	1	2	1	3	4	3	3	2	3
Bladder + rectum + skeletal pelvis	-	-	1	2	2	-	-	2	2
Bladder + seminal vesicles + rectum	-	1	1	1	2	-	-	-	1
Lymph nodes + bladder + skeletal pelvis	4	2	2	-	1	3	2	6	6
Lymph nodes + seminal vesicles + rectum	2	3	2	3	4	3	2	3	2
Seminal vesicles + rectum + skeletal pelvis	1	1	2	2	3	-	2	1	-
Bladder + seminal vesicles + skeletal pelvis	-	1	-	-	3	-	-	1	1
Lymph nodes + bladder + seminal vesicles	2	1	2	3	4	1	2	3	2
Lymph nodes + seminal vesicles + skeletal pelvis	6	5	5	6	6	3	5	9	5
Lymph nodes + seminal vesicles + rectum + skeletal pelvis	1	1	3	2	1	-	3	2	1
Lymph nodes + bladder + seminal vesicles + skeletal pelvis	1	-	1	2	2	1	1	1	-
Lymph nodes + bladder + rectum + skeletal pelvis	-	-	1	3	5	-	-	-	1
Lymph nodes + bladder + seminal vesicles + rectum	-	-	1	1	1	-	1	-	-
Lymph nodes + bladder + seminal vesicles + rectum + skeletal pelvis	-	-	-	-	1	-	-	-	-

S1 and S2 indicate the two senior radiologists; J1 and J2 indicate the two junior radiologists.

Supplementary Table 3. The distribution of primary and secondary RAC for each radiologist

Regions	Primary RAC					Secondary RAC				
	Reference standard	S1	S2	J1	J2	Reference standard	S1	S2	J1	J2
Primary disease*										
RAC-1	23	20	21	18	18	18	16	15	14	17
RAC-2	27	29	25	29	31	23	25	22	26	21
RAC-3	15	17	18	18	17	12	15	16	11	13
RAC-4	3	4	2	2	2	1	2	2	3	3
RAC-5	2	2	3	3	2	-	1	-	-	1
Bladder										
RAC-1	9	7	11	14	9	10	9	8	7	9
RAC-2	13	12	15	20	15	5	6	7	6	4
RAC-3	7	7	8	8	10	5	4	3	4	3
RAC-4	1	2	3	3	3	-	1	1	-	-
RAC-5	1	2	-	1	2	1	1	1	1	-
Rectum										
RAC-1	11	12	14	13	11	4	3	4	5	4
RAC-2	6	6	7	13	9	7	9	5	4	4
RAC-3	4	5	6	10	9	4	3	4	6	1
RAC-4	2	2	3	1	4	-	1	-	-	-
RAC-5	1	1	1	2	4	2	3	2	2	2
Lymph nodes										
RAC-1	42	34	39	31	26	32	29	27	22	26
RAC-2	35	34	31	33	28	33	32	30	26	28
RAC-3	15	20	21	15	21	23	23	23	20	21
RAC-4	6	7	6	8	1	3	3	2	2	1
RAC-5	5	6	6	5	2	2	2	3	2	2
Skeletal pelvis										
RAC-1	27	20	22	24	27	22	19	17	19	17
RAC-2	30	26	30	31	27	27	23	24	24	24
RAC-3	16	17	13	14	19	17	18	12	15	17
RAC-4	6	9	9	6	6	4	4	5	3	2
RAC-5	1	1	2	-	1	-	-	1	-	-
Seminal vesicles										
RAC-1	13	10	9	18	10	11	9	7	9	10
RAC-2	14	13	12	18	17	7	9	4	8	5
RAC-3	6	6	8	8	5	6	5	6	5	2
RAC-4	3	3	4	4	6	4	3	3	4	3
RAC-5	1	1	1	2	3	-	-	-	-	-

*The response assessment category (RAC) evaluation of the primary disease was only performed for the patients who had not undergone radical prostatectomy. S1 and S2 indicate the two senior radiologists; J1 and J2 indicate the two junior radiologists.