Title Page

Title: Interobserver Agreement for the Standardized Reporting System PSMA-RADS 1.0 on ^{18}F-DCFPyL PET/CT Imaging

Running Title: PSMA-RADS 1.0 Interobserver Agreement

Authors: Rudolf A. Werner[1,2], Ralph A. Bundschuh[3], Lena Bundschuh[3], Mehrbod S. Javadi[1], Jeffrey P. Leal[1], Takahiro Higuchi[2,4], Kenneth J. Pienta[5], Andreas K. Buck[2], Martin G. Pomper[1], Michael A. Gorin[1,5], Constantin Lapa[2,*], Steven P. Rowe[1,5,*]

[1]The Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[2]Department of Nuclear Medicine/Comprehensive Heart Failure Center, University Hospital Würzburg, Germany

[3]Department of Nuclear Medicine, University Medical Center Bonn, Germany

[4]Department of Bio Medical Imaging, National Cardiovascular and Cerebral Research Center, Suita, Japan

[5]The James Buchanan Brady Urological Institute and Department of Urology, Johns Hopkins University School of Medicine, Baltimore, MD, USA

* equally contributed

Correspondence:
Steven P. Rowe, M.D., Ph.D.
Division of Nuclear Medicine and Molecular Imaging
The Russell H. Morgan Department of Radiology and Radiological Science
Johns Hopkins University School of Medicine
601 N. Caroline St.
Baltimore, MD 21287
Phone: (410) 502-1520
E-mail: srowe8@jhmi.edu

Submission Type: Original article

Word Count: 3,359

References: 29

Figures: 5

ABSTRACT

Objectives: Recently, the standardized reporting and data system for prostate-specific membrane antigen (PSMA)-targeted positron emission tomography (PET) imaging studies, termed PSMA-RADS version 1.0, was introduced. We aimed to determine the interobserver agreement for applying PSMA-RADS to imaging interpretation of ^{18}F-DCFPyL PET examinations in a prospective setting mimicking the typical clinical work-flow at a prostate cancer referral center.

Methods: Four readers (two experienced readers (ER, > 3 years of PSMA-targeted PET interpretation experience) and two inexperienced readers (IR, < 1 year of experience)), who had all read the initial publication on PSMA-RADS 1.0, assessed 50 ^{18}F-DCFPyL PET/computed tomography (CT) studies independently. Per scan, a maximum of 5 target lesions were selected by the observers and a PSMA-RADS score for every target lesion was recorded. No specific pre-existing conditions were placed on the selection of the target lesions, although PSMA-RADS 1.0 suggests that readers focus on the most highly avid or largest lesions. An overall scan impression based on PSMA-RADS was indicated and interobserver agreement rates on a target lesion-based, on an organ-based, and on an overall PSMA-RADS score-based level were computed.

Results: The number of target lesions identified by each observer were as follows: ER 1, 123; ER 2, 134; IR 1, 123; and IR 2, 120. Among those selected target lesions, 125 were chosen by at least two individual observers (all four readers selected the same target lesion in 58/125 (46.4%) instances, three readers in 40/125 (32%) and two observers in 27/125 (21.6%) instances). The interobserver agreement for PSMA-RADS scoring among identical target lesions was good (intraclass correlation coefficient (ICC) for four, three and two identical target lesions, ≥0.60, respectively). For lymph nodes, an excellent interobserver agreement was derived (ICC=0.79). The interobserver agreement for an overall scan impression based on PSMA-RADS was also excellent (ICC=0.84), with a significant

difference for ER (ICC=0.97) vs. IR (ICC=0.74, P=0.005).

Conclusions: PSMA-RADS demonstrates a high concordance rate in this study, even among readers with different levels of experience. This suggests that PSMA-RADS can be effectively used for communication with clinicians and can be implemented in the collection of data for large prospective trials.

Keywords: ^{18}F-DCFPyL, PSMA-RADS, interreader, interobserver, PSMA, prostate cancer, RADS, reporting and data system

INTRODUCTION

Radiotracers targeting prostate-specific membrane antigen (PSMA), such as the urea-based small molecule ^{18}F-DCFPyL, have demonstrated excellent performance characteristics in identifying sites of disease in subjects with prostate cancer (PCa) (*1-3*). However, in patients with extensive tumor burden (*4*) or for lesion detection in preoperative lymph node staging (*5*), clinical interpretors have to consider certain pitfalls, e.g. uptake in benign lesions or in nonprostatic malignancies (*6-10*). To aid in the interpretation of PSMA-targeted PET imaging studies, mulitple structured reporting systems have been proposed. These include the *Prostate Cancer Molecular Imaging Standardized Evaluation* and the *PSMA-reporting and data system (PSMA-RADS*, version 1.0) (*11-14*). Such frameworks help convey to the reader the level of certainty that an equivocal finding or a finding without a cross sectional imaging correlate is a site of disease. Striving for a readily applicable system for a clinical observer, PSMA-RADS is simple, easy to memorize and utilize, and exclusively based on imaging findings (i.e., the site and intensity of radiotracer uptake). Both individual target lesions (maximum five per scan) and the overall impression of the imaging study should receive a PSMA-RADS score. Such scores are on a 5-point scale that reflects the confidence of the interpreting imaging specialist that a given lesion represents a site of PCa (from 1 = definitively benign to 5 = high degree of certainty that PCa is present). PSMA-RADS 1.0 may facilitate the collection of data for larger clinical trials, can serve as a guide for nuclear medicine physician in interpreting PSMA-targeted PET scans, and can enable efficient communication with referring clinicians (*13*).

To validate the utility of PSMA-RADS, further confirmatory work on this proposed standardized reporting system is needed and the interobserver agreement among different interpretors has to be addressed. As such, we undertook to determine the interobserver reliability of PSMA-RADS in a prospective setting in which readers with varying experience levels evaluated 50 ^{18}F-DCFPyL PET/computed tomography (CT) scans randomly selected from a large trial evaluating the clinical utility of the radiotracer. All observers had read the original PSMA-RADS publication but were blinded to all information about the patients and

were provided no other instructions, thus simulating some elements of a real-world, busy clinical PCa practice.

MATERIALS AND METHODS

In total, 50 patients with histologically proven PCa who had undergone ^{18}F-DCFPyL PET/CT imaging were included in this evaluation. All patients were originally imaged as part of an institutional review board-approved protocol (ClinicalTrials.gov identifier NCT02825875) and all patients signed written informed consent. ^{18}F-DCFPyL was used according to an Food and Drug Administration Investigational New Drug application (IND 121064).

Imaging Procedure.

As per our standard practice, patients were asked to be *nil per os* (with the exception of water and medications) for at least four hours prior to radiotracer injection. ^{18}F-DCFPyL was synthesized as previously described (*15*). Integrated PET/CT using either a Discovery RX 64-slice PET/CT (General Electric, Waukesha, Wisconsin, USA) or a Biograph mCT 128-slice PET/CT (Siemens, Erlangen, Germany) operating in 3D emission mode with CT attenuation correction was performed in all patients. ^{18}F-DCFPyL ≤333 MBq (≤9 mCi) was administered intravenously and after an uptake time of approximately 60 minutes, acquisitions from the mid-thigh to the vertex of the skull were conducted, covering six to eight bed positions (depending on patient height and the scanner) with patients in the supine position. A detailed description can be found in (*7*).

Imaging Interpretation.

PET images were analyzed using XD3 Software (Mirada Medical, Oxford, UK). PET, CT, and hybrid PET/CT imaging overlay were assessed in all 50 patients. Two experienced readers (a dual board-certified nuclear medicine physician/radiologist (ER1) as well as a board-certified nuclear medicine physician (ER2) with >3 years of experience in reading

PSMA-targeted PET scans, respectively) and two inexperienced readers (a recently board-certified nuclear medicine physician (IR1) and a resident (IR2), <1 year experience in reading PSMA-targeted PET scans), blinded to the clinical status of the patients (other than knowing that the patients had been imaged due to a history of PCa), performed an evaluation of all scans independently. Except for ER1, the remaining three readers had no previous experience with reading ^{18}F-labeled PSMA-targeted PET images (i.e., those observers had clinical experience solely in interpreting ^{68}Ga-PSMA-11 or ^{68}Ga-PSMA Imaging and Therapy (I&T) PET scans). Prior to beginning the blinded independent reads, the inexperienced readers underwent a training session with five cases to gain familiarity with the workstation and the XD3 Software (Mirada Medical, Oxford, UK) which was used to display the scans.

PSMA-RADS-1A lesions are benign with no abnormal radiotracer uptake, PSMA-RADS-1B are benign lesions (often characterized by biopsy or pathognomonic imaging) that have abnormal radiotracer uptake. Often, characterizing a lesion as PSMA-RADS-1B involves previous conventional imaging or histologic diagnosis; as such, PSMA-RADS-1A and -1B were subsumed under PSMA-RADS-1 in the present blinded analysis. No other changes to the PSMA-RADS system were implemented in this study. A complete summary of the PSMA-RADS scoring system (from PSMA-RADS-1 to -5) can be found in (*13*).

In accordance to the specifications of PSMA-RADS 1.0, a maximum of five target lesions were selected by the readers. PSMA-RADS suggests that target lesions be those that are largest in size and/or have the most intense radiotracer uptake, although ultimately target lesion selection is left to the discretion of the interpreting imaging specialist. Further, a maximum of three lesions per organ can be included. The following organ compartments were defined: lymph nodes (LN), skeleton, prostate/local recurrence, soft tissue (other than LN), liver, thyroid, and lung (*16*). A PSMA-RADS Score had to be assigned to every target lesion. Additionally, all involved organ compartments were identified by the readers and an overall scan score was assigned. The overall PSMA-RADS score was defined analogous to somatostatin receptor RADS (i.e., the highest PSMA-RADS score of any of the individual target lesions) (*17*). Moreover, the following general parameters were assessed by each

observer in a binary fashion: overall scan result (positive in case of suspicious radiotracer uptake above background), organ involvement, and LN involvement. Additionally, the number of organs affected, the number of organ metastases, the number of LN regions, and the number of LN had to be indicated on a 5-point scale (from 1 to ≥5 organ metastases, LNs, or number of organs/LN areas affected). The following LN areas were defined: cervical, thoracic/axillary, retroperitoneal, (pre)sacral, and pelvic (*16*). Moreover, the concordance between both ERs and IRs was evaluated in an interobserver setting for the Overall PSMA-RADS Score.

Statistical Analysis.

Continious data are presented as mean ± standard deviation. The categorical variables are presented as frequency (percentage). The degrees of agreement were assessed using intraclass correlation coefficients (ICC) and their 95% confidence intervals based on a mean-rating, single-measure, consistency model. According to Chicchetti, an ICC of less than 0.4 indicates poor interobserver agreement, 0.4 – 0.59 equates to fair agreement, 0.6 – 0.74 equates to good agreement, and and 0.75 – 1 corresponds to excellent interobserver agreement (*18*). Statistical analysis was performed using MedCalc Statistical Software (version 18.2.1, MedCalc Software bvba, Ostend, Belgium; http://www.medcalc.org; 2018). The statistical significance level was set at $P<0.05$.

RESULTS

Details about patients' characteristics are provided in Tab. 1.

General Parameters.

For the three parameters that had to be evaluated in a binary fashion (overall scan result, organ involvement, and LN involvement), the interobserver agreement was excellent (ICC,

0.75, 0.80 and 0.78, respectively) (*18*). Except for the number of organs affected (good interobserver agreement, ICC, 0.74), all general parameters that were evaluated on a 5-point scale demonstrated excellent agreement (number of LN areas affected, ICC 0.79; number of organ metastases, ICC, 0.92; number of LN metastases, ICC, 0.90). Tab. 2 summarizes all results for those general scan parameters and Fig. 1 displays the distribution for number of organ and LN metastases for all 4 readers.

Target lesion- and Compartment-based Interobserver Agreement.

In total, the following number of target lesions were recorded by each reader: ER 1, 123; ER 2, 134; IR 1, 123; and IR 2, 120. Among those selected target lesions, 125 were chosen by at least two inidividual observers. The majority of the lesions were assigned to either LN, 64/125 (51.2%) or skeleton, 39/125 (31.2%) (Tab. 3).

Identical target lesion included by four readers. The identical target lesion was included by all four readers in 58/125 (46.4%) instances, with the majority of those findings being either LN (26/58, 44.8%) or bone lesions (19/58, 32.8%). In 29/58 (50%) of those target lesions, all four readers designated the identical PSMA-RADS score, with another 17/58 (29.3%) having agreement on the PSMA-RADS score by three readers. The remaining lesions (12/58, 20.7%) had agreement by two readers on the PSMA-RADS score. The ICC was 0.60 (0.48 – 0.71). On an organ-based compartment level for all four readers selecting the same LN, the interobserver agreement rate was 0.79 (0.66 – 0.89). Fig. 2 illustrates the PSMA-RADS Score for four identical target lesions among all readers.

Identical target lesion included by three readers. In 40/125 (32%) of the cases, three readers identified an identical target lesion. LN comprised 22/40 (55%) of these target lesions with 12/40 (30%) being bone findings. In 21/40 (52.5%) all three readers agreed on the same PSMA-RADS score (two readers, 15/40 (37.5%) and no concordance in the remaining 4/40 (10%)). The ICC was 0.60 (0.43 – 0.75). Similar to the situation of four identical target lesion selections, the interobserver agreement was 0.66 for LN (0.44 – 0.83).

Identical target lesion included by two readers. In 27 of the 125 identical target lesions (21.6%), a minimum of two readers selected the same finding. LN (16/27, 59.3%) and bone lesions (8/27, 29.6%) were seen in the majority of the cases. In approximately half of the cases (15/27 (55.6%)), both readers agreed on the PSMA-RADS score (no concordance in the remaining 12/27 (44.4%)). The ICC was 0.62 (0.32 – 0.81) for two identical target lesions (LN, ICC=0.57 (0.12 – 0.83)).

Taken together, the ICC for four, three, and two identical chosen target lesions can be described as good. The number of investigated identical bone lesions by all four, three, or two readers was too small for a reliable assessment of ICCs. Tab. 3 summarizes the compartment-based and target lesion interobserver agreement findings. Tab. 4 provides a distribution of the different PSMA-RADS scores for those target lesions that had been included by all four readers.

Overall PSMA-RADS.

In the majority of the cases, the readers described the scan impression with an overall PSMA-RADS score of 4 or 5. The ICC was 0.84 (0.77 – 0.90; i.e. excellent agreement). Tab. 4 gives an overview of the distribution of the different overall PSMA-RADS scores for all four readers. Fig. 3 illustrates the overall PSMA-RADS distribution among different readers.

Experienced vs. Inexperienced Readers.

Compared to ERs serving as a gold standard, the ICC of the ERs for an overall PSMA-RADS score level was 0.97 (0.94 – 0.98), while for the IRs, the ICC was 0.74 (0.58 – 0.84). A statistical significant difference could be reached for the ICC of the ERs vs. the ICC of the IRs (P=0.005). These findings were further corroborated on a target-based level investigating all the identical target lesion which were included by all four readers. The ICC for the ERs was 0.80 (0.68 – 0.88) and was statistically significant different from the ICC for IRs, 0.53

(0.32- 0.60), P=0.013. Figs. 4 and 5 provide examples of lesions in which reader experience may have played a role in PSMA-RADS scoring.

DISCUSSION

In light of the growing availability of ^{68}Ga- or ^{18}F-labeled PSMA-targeted imaging agents (*19-22*), the number of molecular imaging specialists that routinely interpret PET scans with these compounds outside of controlled clinical trials is currently expanding (*23*). However, numerous studies have reported on pitfalls while reading PSMA-targeted PET studies, e.g. in Paget's disease, sarcoidosis or in nervous tissue such as ganglia (*7-10*). Any systematic approach to the interpretation of PSMA-targeted PET scans should therefore build in a measure of uncertainty as to the presence of PCa. The recently reported system PSMA-RADS version 1.0, incorporates such uncertainty with recommended follow-up for indeterminate lesions (*12*). Further, such a system should also facilitate communication of important findings between image interpreters and referring clinicians, should be useful for collecting data in multi-center prospective studies, and should allow for the eventual implementation of machine learning algorithms based on the system. For all of these applications, high inter-observer reproducibility is necessary.

The ICC for the Overall PSMA-RADS Score for both ERs (0.97) was consistent with excellent interobserver agreement, while the two IRs still agreed well (0.74) on an overall PSMA-RADS score level (all four interpreters, 0.84, Fig. 3). This is in line with previous reports in which experienced readers demonstrated an almost-perfect reproducibility on ^{68}Ga-PSMA-11 PET/CT for specified lesions (low-experienced observers, substantial agreement) (*16*). Notably, these results are in contradistinction to other standardized reporting systems, such as prostate imaging (PI)-RADS version 2 for prostate MRI (moderate interobserver agreement among experienced radiologists with a Fleiss' $k < 0.6$) (*24*). In a similar vein, a significant variation was present in both the PI-RADS distribution between

radiologists and, more importantly, in the detection of suspected clinically significant cancer by PI-RADS using multiparametric MRI (*25*).

On an overall scan impression level, the majority of the PET studies were assigned PSMA-RADS-4 or -5 scores by all observers (Table 4). We hypothesize that this observation derives from the high specificity and sensitivity of PSMA-targeted radiotracers. While PI-RADS highly depends on the experience of the reading radiologists (*25*), PSMA-RADS seems to be readily applicable even for less experienced readers (ICC, 0.74). These findings were further corroborated on a target lesion level (Fig. 2). Despite the fact that PSMA-RADS provides little specific information on the selection of target lesions, a minimum of three readers (i.e. minimum one IR) designated the same PSMA-RADS score within the context of all four readers selecting the same target lesion with an agreement rate of >79% (Tab. 3). Moreover, on an organ compartment-based level, the ICC for LN lesions based on PSMA-RADS was 0.79, which is almost identical to a previous assessment for the interobserver agreement for LN (Fleiss' k, 0.80) (*16*).

A nuance of the current study is that the ERs gained experience with subtly different PSMA-targeted radiotracers. There is a current trend towards increased use of ^{18}F-labeled PSMA-targeted imaging agents for PCa molecular imaging, although ^{68}Ga-PSMA-11 has been by far the most commonly used radiotracer to date (*26*). In head-to-head comparisons between ^{68}Ga- and ^{18}F-labeled compounds, a higher detection rate for sites of disease as well as an increased tumor-to-background ratio were demonstrated with a radiofluorinated agent (*27,28*). Some of the differences in interpretation between ER 1 and ER 2 might be related to their relative familiarities with these different PSMA-targeted radiotracers. A common example which has been classified differently by the ^{18}F-trained reader compared to the ^{68}Ga-trained readers, is given in Fig. 4: While ER1 called uptake in a right iliac LN lesion PSMA-RADS-4 (i.e. PCa highly likely to be present), two other readers (ER2 and IR1, both trained with ^{68}Ga-PSMA PET imaging agents) classified this lesion as PSMA-RADS-3A (i.e. a suspicious but indeterminate LN) (*13*). The ^{18}F-trained reader may have a higher

confidence in lesion interpretation on ^{18}F-DCFPyL PET scans, most likely due to the higher sensitivity in the detection rate of small lesions using ^{18}F-labeled radiotracers compared to ^{68}Ga-PSMA PET imaging agents (*27,28*).

Further corroborating the need for a standardized framework system (*11,12*), one of the IR classified moderate radiotracer uptake in mediastinal and hilar LN as PSMA-RADS-4 (Fig. 5), while ER1 called it PSMA-RADS-2 (i.e. likely benign). Even though the IR had potentially misinterpreted the low-level uptake in the LNs (longitudinal follow-up imaging showed no change in these LNs), this did not impact the overall scan score. Thus, PSMA-RADS may contribute to a self-learning effect: PSMA-RADS-4 lesions may be "downgraded" to PSMA-RADS-2 when subsequent imaging confirms stability, which in turn would increase the understanding of the IR to differentiate between typical and atypical sites of PCa metastases.

This study has several limitations. First, false-positive findings, in particular on a target lesion level, cannot be ruled out, as histopathological assessment of the target lesions (many of which are small and not targetable on conventional imaging) would not be feasible. Second, the readers were blinded to clinical status and potential corroborative imaging, potentially lowering inter-observer agreement; however, the cases in this study were randomly selected and the readers blinded to ancillary information in order to create a "worst case scenario" reflection of a busy real-world clinical practice to best test the applicability of PSMA-RADS. Although, in many situations, clinical information would be available to readers, we wished to ascertain the robustness of PSMA-RADS as an imaging-finding-driven construct. Nonetheless, future studies must clarify if providing clinical information has an important impact on the agreement rate of multiple observers and should also include stratification by serum prostate specific antigen levels. Given the small number of identical bone lesions, ICC could not be provided for bone metastases. However, the readers in this study may have identified different target lesions in some patients with extensive skeletal involvement. Lastly, a larger trial including more scans and readers could further corroborate our preliminary findings. Nonetheless, the agreement rate of the overall PSMA-RADS score was excellent among all observers and this is a promising initial result.

CONCLUSION

In the present prospective study investigating the interobserver agreement of the novel structured reporting system PSMA-RADS version 1.0, a high concordance rate, even among readers with different experience, was observed. Thus, PSMA-RADS may be a useful framework for interpreting PSMA-targeted imaging studies, which in turn paves the way for implementing PSMA-RADS in the collection of data for larger prospective trials.

ACKNOWLEDGMENTS

The Prostate Cancer Foundation Young Investigator Award and National Institutes of Health grants CA134675, CA183031, CA184228, and EB024495. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 701983.

REFERENCES

1. Giesel FL, Will L, Kesch C, et al. Biochemical recurrence of prostate cancer: initial results with [(18)F]PSMA-1007 PET/CT. *J Nucl Med.* 2018;59:632-635.

2. Calais J, Czernin J, Cao M, et al. (68)Ga-PSMA-11 PET/CT Mapping of Prostate Cancer Biochemical Recurrence After Radical Prostatectomy in 270 Patients with a PSA Level of Less Than 1.0 ng/mL: Impact on Salvage Radiotherapy Planning. *J Nucl Med.* 2018;59:230-237.

3. Szabo Z, Mena E, Rowe SP, et al. Initial Evaluation of [(18)F]DCFPyL for Prostate-Specific Membrane Antigen (PSMA)-Targeted PET Imaging of Prostate Cancer. *Mol Imaging Biol.* 2015;17:565-574.

4. Schmuck S, von Klot CA, Henkenberens C, et al. Initial Experience with Volumetric (68)Ga-PSMA I&T PET/CT for Assessment of Whole-Body Tumor Burden as a Quantitative Imaging Biomarker in Patients with Prostate Cancer. *J Nucl Med.* 2017;58:1962-1968.

5. Maurer T, Gschwend JE, Rauscher I, et al. Diagnostic Efficacy of (68)Gallium-PSMA Positron Emission Tomography Compared to Conventional Imaging for Lymph Node Staging of 130 Consecutive Patients with Intermediate to High Risk Prostate Cancer. *J Urol.* 2016;195:1436-1443.

6. Rowe SP, Deville C, Paller C, et al. Uptake of (18)F-DCFPyL in Paget's Disease of Bone, an Important Potential Pitfall in Clinical Interpretation of PSMA PET Studies. *Tomography.* 2015;1:81-84.

7. Werner RA, Sheikhbahaei S, Jones KM, et al. Patterns of uptake of prostate-specific membrane antigen (PSMA)-targeted (18)F-DCFPyL in peripheral ganglia. *Ann Nucl Med.* 2017;31:696-702.

8. Rischpler C, Beck TI, Okamoto S, et al. (68)Ga-PSMA-HBED-CC uptake in cervical, coeliac and sacral ganglia as an important pitfall in prostate cancer PET imaging. *J Nucl Med.* 2018.

9. Sheikhbahaei S, Afshar-Oromieh A, Eiber M, et al. Pearls and pitfalls in clinical interpretation of prostate-specific membrane antigen (PSMA)-targeted PET imaging. *Eur J Nucl Med Mol Imaging.* 2017;44:2117-2136.

10. Hofman MS, Hicks RJ, Maurer T, Eiber M. Prostate-specific Membrane Antigen PET: Clinical Utility in Prostate Cancer, Normal Patterns, Pearls, and Pitfalls. *Radiographics.* 2018;38:200-217.

11. Eiber M, Herrmann K, Calais J, et al. Prostate Cancer Molecular Imaging Standardized Evaluation (PROMISE): Proposed miTNM Classification for the Interpretation of PSMA-Ligand PET/CT. *J Nucl Med.* 2018;59:469-478.

12. Rowe SP, Pienta KJ, Pomper MG, Gorin MA. PSMA-RADS Version 1.0: A Step Towards Standardizing the Interpretation and Reporting of PSMA-targeted PET Imaging Studies. *Eur Urol.* 2018;73:485-487.

13. Rowe SP, Pienta KJ, Pomper MG, Gorin MA. Proposal for a Structured Reporting System for Prostate-Specific Membrane Antigen-Targeted PET Imaging: PSMA-RADS Version 1.0. *J Nucl Med.* 2018;59:479-485.

14. Fanti S, Minozzi S, Morigi JJ, et al. Development of standardized image interpretation for 68Ga-PSMA PET/CT to detect prostate cancer recurrent lesions. *Eur J Nucl Med Mol Imaging.* 2017;44:1622-1635.

15. Ravert HT, Holt DP, Chen Y, et al. An improved synthesis of the radiolabeled prostate-specific membrane antigen inhibitor, [(18) F]DCFPyL. *J Labelled Comp Radiopharm.* 2016;59:439-450.

16. Fendler WP, Calais J, Allen-Auerbach M, et al. (68)Ga-PSMA-11 PET/CT Interobserver Agreement for Prostate Cancer Assessments: An International Multicenter Prospective Study. *J Nucl Med.* 2017;58:1617-1623.

17. Werner RA, Solnes L, Javadi M, et al. SSTR-RADS Version 1.0 as a Reporting System for SSTR-PET Imaging and Selection of Potential PRRT Candidates: A Proposed Standardization Framework. *J Nucl Med.* 2018;59:1085-1091.

18. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment.* 1994;6:284-290.

19. Salas Fragomeni RA, Amir T, Sheikhbahaei S, et al. Imaging of Non-Prostate Cancers Using PSMA-Targeted Radiotracers: Rationale, Current State of the Field, and a Call to Arms. *J Nucl Med.* 2018.

20. Giesel FL, Hadaschik B, Cardinale J, et al. F-18 labelled PSMA-1007: biodistribution, radiation dosimetry and histopathological validation of tumor lesions in prostate cancer patients. *Eur J Nucl Med Mol Imaging.* 2017;44:678-688.

21. Paddubny K, Freitag MT, Kratochwil C, et al. Fluorine-18 Prostate-specific Membrane Antigen-1007 Positron Emission Tomography/Computed Tomography and Multiparametric Magnetic Resonance Imaging in Diagnostics of Local Recurrence in a Prostate Cancer Patient After Recent Radical Prostatectomy. *Clin Genitourin Cancer.* 2018;16:103-105.

22. Werner RA, Andree C, Javadi MS, et al. A Voice From the Past: Rediscovering the Virchow Node With Prostate-specific Membrane Antigen-targeted (18)F-DCFPyL Positron Emission Tomography Imaging. *Urology.* 2018;117:18-21.

23. Fendler WP, Eiber M, Beheshti M, et al. (68)Ga-PSMA PET/CT: Joint EANM and SNMMI procedure guideline for prostate cancer imaging: version 1.0. *Eur J Nucl Med Mol Imaging.* 2017;44:1014-1024.

24. Rosenkrantz AB, Ginocchio LA, Cornfeld D, et al. Interobserver Reproducibility of the PI-RADS Version 2 Lexicon: A Multicenter Study of Six Experienced Prostate Radiologists. *Radiology.* 2016;280:793-804.

25. Sonn GA, Fan RE, Ghanouni P, et al. Prostate Magnetic Resonance Imaging Interpretation Varies Substantially Across Radiologists. *Eur Urol Focus.* 2017.

26. Kesch C, Kratochwil C, Mier W, Kopka K, Giesel FL. (68)Ga or (18)F for Prostate Cancer Imaging? *J Nucl Med.* 2017;58:687-688.

27. Dietlein M, Kobe C, Kuhnert G, et al. Comparison of [(18)F]DCFPyL and [(68)Ga]Ga-PSMA-HBED-CC for PSMA-PET Imaging in Patients with Relapsed Prostate Cancer. *Mol Imaging Biol.* 2015;17:575-584.

28. Dietlein F, Kobe C, Neubauer S, et al. PSA-Stratified Performance of (18)F- and (68)Ga-PSMA PET in Patients with Biochemical Recurrence of Prostate Cancer. *J Nucl Med.* 2017;58:947-952.

TABLES

Parameter		
Age (median ± SD, in years)		65 ± 8
Race	White	38/50 (76%)
	Black	9/50 (18%)
	Asian/Other	3/50 (6%)
Indication for Scan	Staging	24/50 (48%)
	Biochemical Reccurence	9/50 (18%)
	Biochemical Persistence after Primary Surgery	6/50 (12%)
	Primary Diagnosis	5/50 (10%)
	Potential withdrawal of androgen deprivation therapy	3/50 (6%)
	Other	3/50 (6%)
Gleason Score (GS)	Overall (median ± SD, available in n=39)	8 ± 1
	GS 6	1/39 (2.6%)
	GS 7	15/39 (38.4%)
	GS 8	7/39 (17.9%)
	GS 9	15/39 (38.5%)
	GS 10	1/39 (2.6%)
PSA level (ng/ml)	Overall (median (range))	3.2 (0.02 - 48)
Prior therapies	in total	41/50 (82%)
	Surgery	29/41 (70.7%)
	Hormonal Therapy	21/41 (51.2%)
	RTx	18/41 (43.9%)
	CTx	6/41 (14.6%)

Table 1. Detailed patients' characteristics. SD = standard deviation, CTx = chemotherapy, PSA = prostate specific antigen, RTx = radiation therapy.

	Parameter	ICC	95% CI		0	1	2	3	4	5
Binary Fashion	Overall Scan Result (negative=0, positive=1)	0.75	0.64 – 0.83	ER1	13/50 (26%)	37/50 (74%)				
				ER2	13/50 (26%)	37/50 (74%)				
				IR1	6/50 (12%)	44/50 (88%)				
				IR2	13/50 (26%)	37/50 (74%)				
	Organ involvement (no=0, yes=1)	0.80	0.71 – 0.88	ER1	28/50 (56%)	22/50 (44%)				
				ER2	27/50 (54%)	23/50 (46%)				
				IR1	20/50 (40%)	30/50 (60%)				
				IR2	29/50 (58%)	21/50 (42%)				
	LN involvement (no=0, yes=1)	0.78	0.69 – 0.86	ER1	25/50 (50%)	25/50 (50%)				
				ER2	27/50 (54%)	23/50 (46%)				
				IR1	21/50 (42%)	29/50 (58%)				
				IR2	24/50 (48%)	26/50 (52%)				
	No. of affected organs	0.74	0.62 – 0.83	ER1	28/50 (56%)	17/50 (34%)	3/50 (6%)	2/50 (4%)		
				ER2	28/50 (56%)	18/50 (36%)	4/50 (8%)			
				IR1	20/50 (40%)	20/50 (40%)	9/50 (18%)	1/50 (2%)		
				IR2	29/50 (58%)	17/50 (34%)	4/50 (8%)			
	No. of organ metatases	0.92	0.89 – 0.95	ER1	28/50 (56%)	10/50 (20%)	1/50 (2%)	3/50 (6%)	2/50 (4%)	6/50 (12%)
				ER2	32/50 (64%)	6/50 (12%)	3/50 (6%)	1/50 (2%)	2/50 (4%)	6/50 (12%)
				IR1	23/50 (46%)	13/50 (26%)	3/50 (6%)	2/50 (4%)	1/50 (2%)	8/50 (16%)

5-point assessment*				IR2	29/50 (58%)	8/50 (16%)	3/50 (6%)	3/50 (6%)	2/50 (4%)	5/50 (10%)
	No. of affected LN areas	**0.79**	**0.70 – 0.86**	ER1	25/50 (50%)	11/50 (22%)	7/50 (14%)	5/50 (10%)	2/50 (4%)	
				ER2	27/50 (54%)	8/50 (16%)	7/50 (14%)	6/50 (12%)	2/50 (4%)	
				IR1	21/50 (42%)	19/50 (38%)	9/50 (18%)	1/50 (2%)		
				IR2	24/50 (48%)	11/50 (22%)	10/50 (20%)	3/50 (6%)	2/50 (4%)	
	No. of LN metatases	**0.90**	**0.85 – 0.94**	ER1	25/50 (50%)	7/50 (14%)	4/50 (8%)	2/50 (4%)		12/50 (24%)
				ER2	28/50 (56%)	4/50 (8%)	4/50 (8%)	2/50 (4%)	2/50 (4%)	10/50 (20%)
				IR1	25/50 (50%)	8/50 (16%)	2/50 (4%)	3/50 (6%)	3/50 (6%)	9/50 (18%)
				IR2	24/50 (48%)	6/50 (12%)	7/50 (14%)	2/50 (4%)	3/50 (6%)	8/50 (16%)

Table 2. Overview of general parameters assessed by all 4 readers (ER, experienced reader, IR, inexperienced reader). For the parameters evaluated in a binary fashion, the description for 0 and 1 is indicated in brackets (under "Parameters"). * The 5-point assessment was structured as follows: from 1 to ≥ 5 organ metastases, lymph nodes (LNs), or number of organs/LN areas affected. ICC = intraclass coefficients, 95%CI = 95% confidence intervals.

	Compartment-based Distribution							Agreement Rate based on PSMA-RADS		ICC based on PSMA-RADS	
	LN	Bone	prostate/ local recurrence	lung	soft tissue#	thyroid gland	liver	for all identical TL	for minimum 3 out of 4 (2 out of 3 TL)ß	for all identical TL	for LN
all identical TL (n=125)*	64/125 (51.2)	39/125 (31.2)	11/125 (8.8)	5/125 (4.0)	3/125 (2.4)	2/125 (1.6)	1/125 (0.8)	n/a	n/a	n/a	n/a
4 identical TL (n=58/125, 46.4%)	26/58 (44.8)	19/58 (32.8)	8/58 (13.8)	3/58 (5.2)	1/58 (1.7)	1/58 (1.7)		29/58 (50.0)	46/58 (79.3)	0.60 (0.48 – 0.71)	0.79 (0.66 – 0.89)
3 identical TL (n=40/125, 32%)	22/40 (55.0)	12/40 (30.0)	3/40 (7.5)	2/40 (5.0)		1/40 (2.5)		21/40 (52.5)	36/40 (90.0)	0.60 (0.43 – 0.75)	0.66 (0.44 – 0.83)
2 identical TL (n=27/125, 21.6%)	16/27 (59.3)	8/27 (29.6)			2/27 (7.4)		1/27 (3.7)	15/27 (55.6)	n/a	0.62 (0.32 – 0.81)	0.57 (0.12 – 0.83)

Table 3. Overview of Target Lesions (TL). Distribution among different compartments is indicated for all investigated TL; for identical TL, which had been identified by all 4 readers (4 identical TL); for identical TL, which had been identified by 3 readers (3 identical TL) and for identical TL, which had been identified by 2 readers (2 identical TL). According to the designated PSMA-RADS score by the different readers, the agreement rate and the intraclass correlation coefficients (ICC) are indicated. *chosen by a minimum of two readers. #other than LN. ß in the context of four readers selecting the same target lesion, a minimum of three readers designated the same PSMA-RADS score (in case of 3 investigated TL, a minimum of two readers). Values in brackets display percentage for compartment-based distribution and agreement rates (for ICC, 95% confidence intervals).

	PSMA-RADS	1#	2	3A	3B	3C	3D	4	5
4 identical TL	ER1		1/58 (1.7)			4/58 (6.9)		22/58 (37.9)	31/58 (53.5)
	ER2		1/58 (1.7)		1/58 (1.7)	2/58 (3.4)		27/58 (46.6)	27/58 (46.6)
	IR1		3/58 (5.2)	2/58 (3.4)	2/58 (3.4)	4/58 (6.9)		12/58 (20.7)	35/58 (60.3)
	IR2	1/58 (1.7)	1/58 (1.7)			3/58 (5.2)		28/58 (48.3)	25/58 (43.1)
Overall PSMA-RADS	ER1	10/50 (20)	2/50 (4)			1/50 (2)		15/50 (30)	22/50 (44)
	ER2	9/50 (18)	2/50 (4)			3/50 (6)	1/50 (2)	16/50 (32)	19/50 (38)
	IR1	6/50 (12)	5/50 (10)	2/50 (4)		2/50 (4)	3/50 (6)	9/50 (18)	23/50 (46)
	IR2	10/50 (20)	3/50 (6)		1/50 (2)			15/50 (30)	21/50 (42)

Table 4. Distribution of PSMA-RADS Score for 4 identical Target Lesions (TL, n=58) and for the Overall PSMA-RADS Score (scans, n=50) among all four readers. ER = Experienced Reader, IR = Inexperienced Reader. #PSMA-RADS-1A and -1B were subsumed under PSMA-RADS score 1, as described in the text. Values in brackets indicate percentage.

FIGURE AND FIGURE LEGENDS

Figure 1. Distribution for number of organ and lymph node (LN) metastases for all 4 readers (ER, experienced reader, IR, inexperienced reader). The 5-point assessment was structured as follows: from 1 to ≥ 5 organ (A) or LN (B) affected.

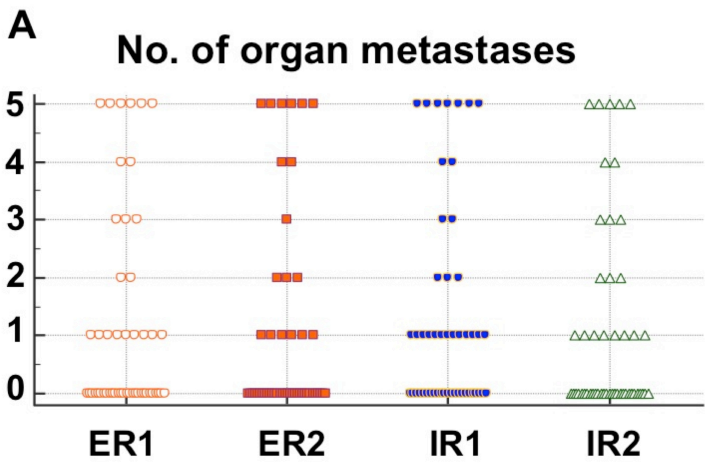
Figure 2. Overview of Target Lesion (TL) assessment (identical target lesion included by all four readers). PSMA-RADS-1A and -1B were subsumed under PSMA-RADS-1, as described in the text.

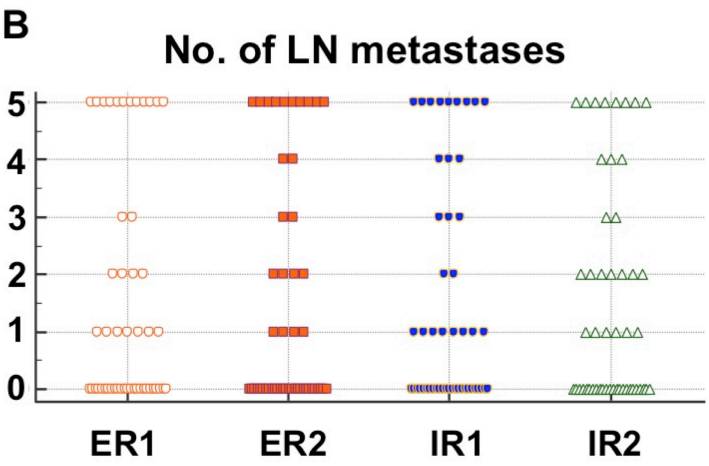
Figure 3. Overview of Overall-PSMA RADS Scoring for all four readers (ER, experienced reader, IR, inexperienced reader). PSMA-RADS-1A and -1B were subsumed under PSMA-RADS-1, as described in the text.

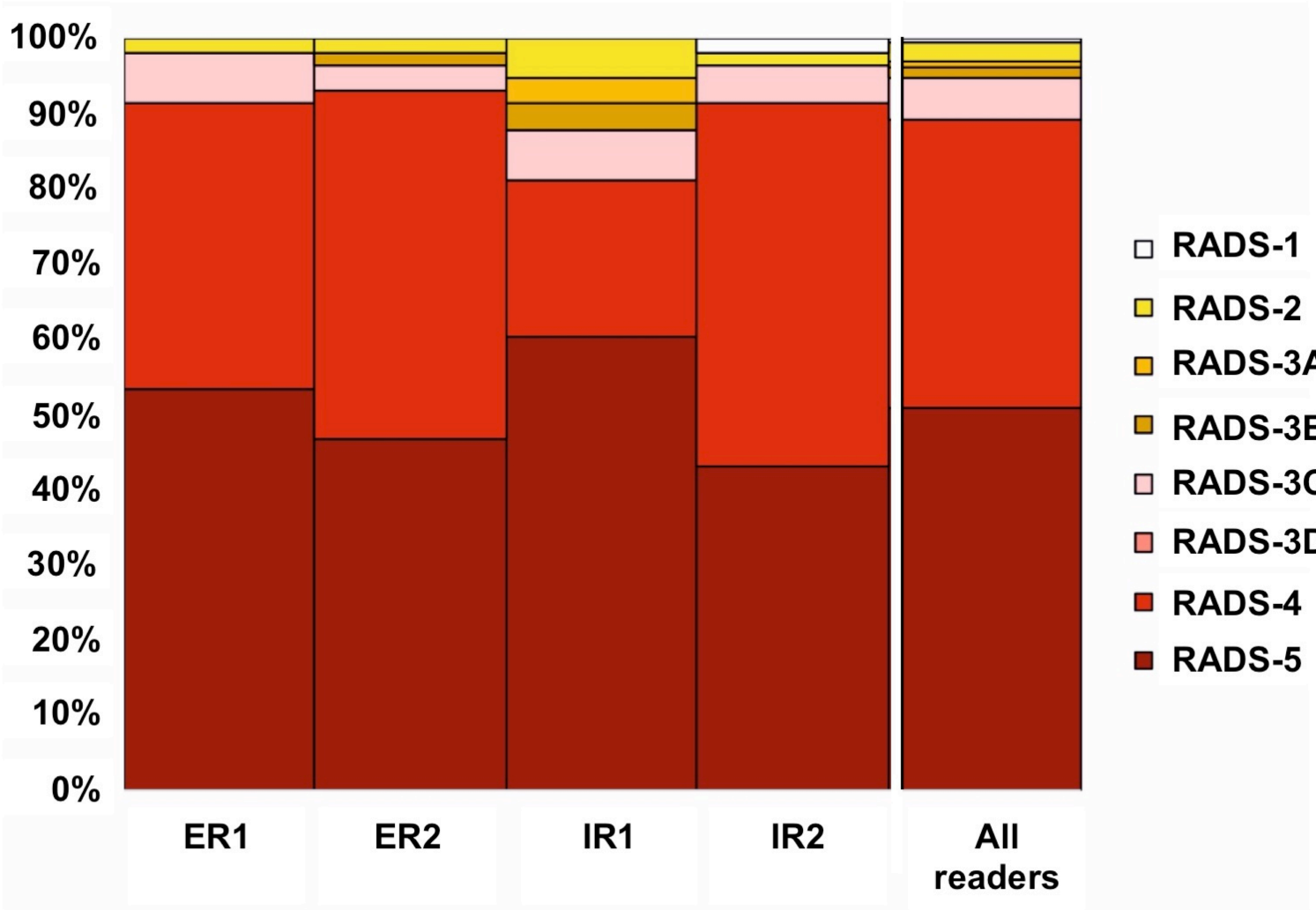
Figure 4. Example of different PSMA-RADS classifications of ^{68}Ga- and ^{18}F-labeled trained readers. Images from a 60 year old male patient undergoing ^{18}F-DCFPyL PET/CT for primary diagnostic assessment (prostate-specific antigen level at date of scan, 13.5 and no previous therapies). (A) Whole body maximum intensity projection demonstrate multiple sites of suspicious radiotracer uptake (e.g. 3rd right rib, red arrowhead and 6th right rib, red arrow). On (B) axial CT, (C) axial ^{18}F-DCFPyL PET and (D) axial ^{18}F-DCFPyL PET/CT, mild radiotracer uptake in a right iliac lymph node is seen (red arrow). While the experienced reader trained on ^{18}F-DCFPyL PET called this lesion PSMA-RADS-4, two remaining readers trained on ^{68}Ga-PSMA PET imaging agents classified this lesion as PSMA-RADS-3A (i.e. a suspicious but indeterminate LN) (*12*). One might speculate that the ^{18}F-trained reader has a higher confidence in lesion interpretation on ^{18}F-DCFPyL PET scans, most likely due to the higher sensitivity in the detection rate of small lesions using ^{18}F-labeled radiotracers compared to ^{68}Ga-PSMA PET imaging agents (*27*). All four readers classified the overall scan impression as PSMA-RADS-5, as the CT in (E) revealed findings corresponding to a 6th right rib metastasis, with discernible radiotracer uptake on (F) axial ^{18}F-DCFPyL PET and (G)

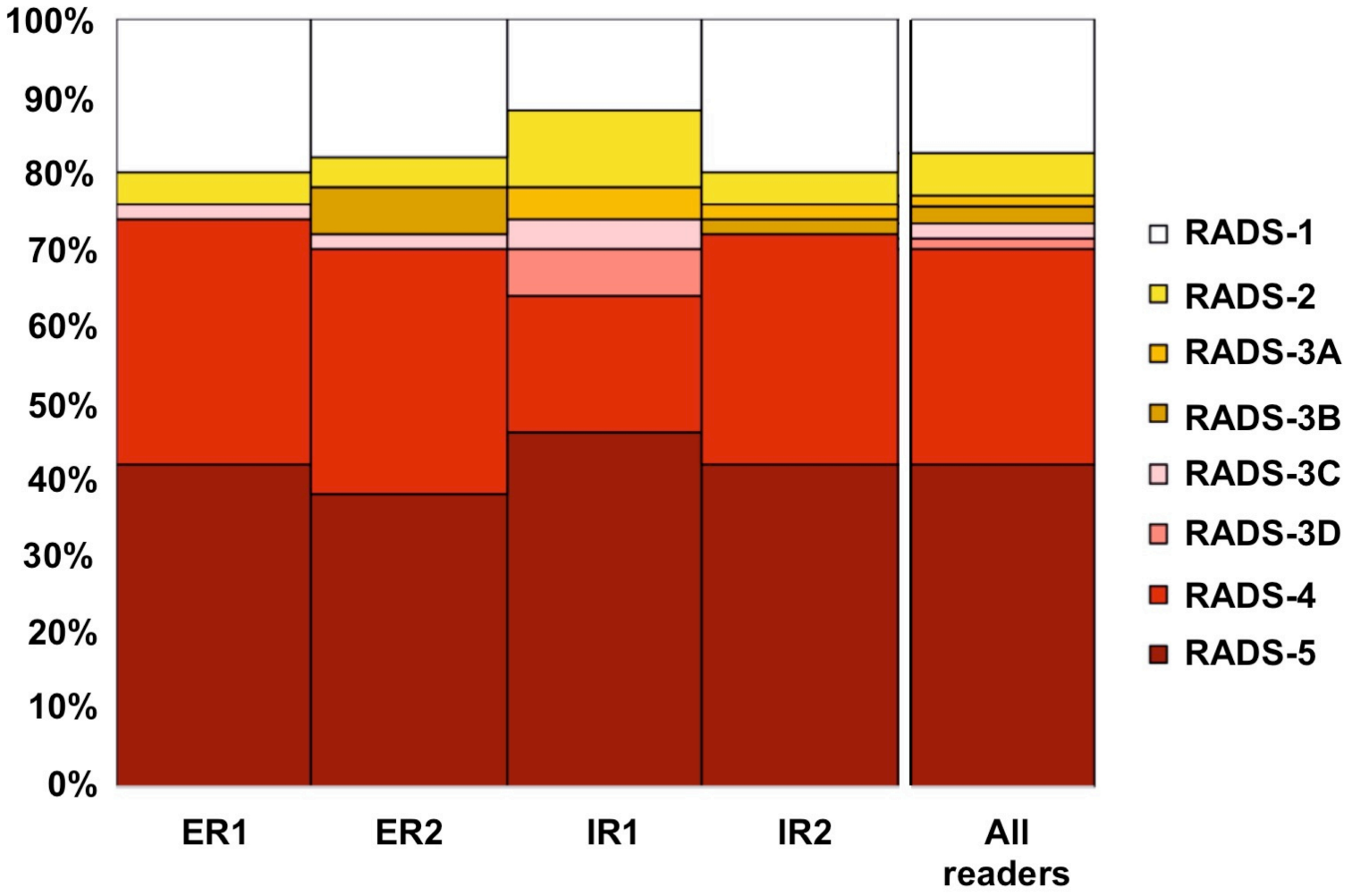
axial ^{18}F-DCFPyL PET/CT (red double thin arrows). Magnification of this 6th rib suspicious site of uptake provided in (H) axial CT, (I) axial ^{18}F-DCFPyL PET and (J) axial ^{18}F-DCFPyL PET/CT further suggested this to be a malignant lesion at this uptake site (red arrowhead).

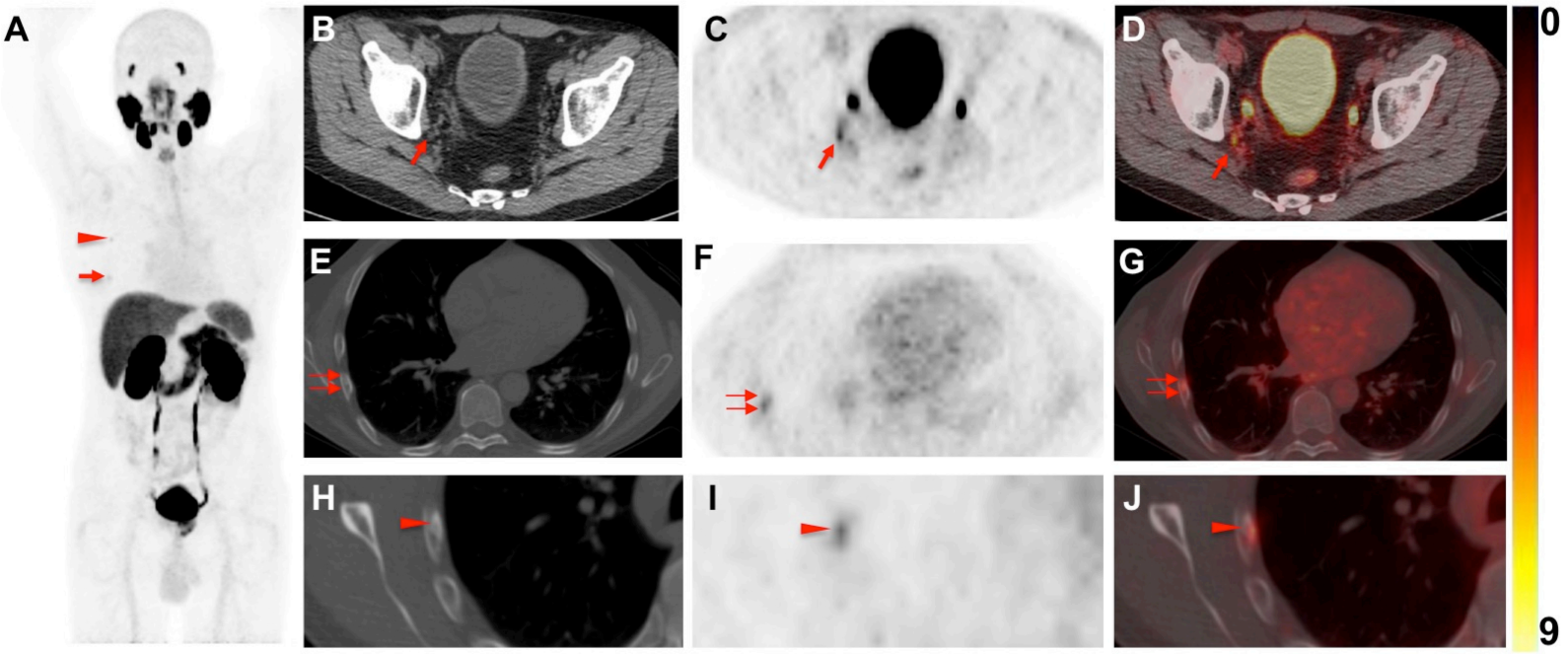
Figure 5. Example of different PSMA-RADS classifications of experienced vs. inexperienced readers. Images from a 76 year old male patient undergoing staging for metastatic PCa (prostate-specific antigen level at date of scan, 0.63 with prior prostatectomy). (A) ^{18}F-DCFPyL whole body maximum intensity projection demonstrates radiotracer uptake in right hilar and subcarinal lymph nodes (LNs, red arrows), a lung lesion (red arrowhead), and the right iliac bone (double thin red arrows). (B) Axial CT, (C) axial ^{18}F-DCFPyL PET and (D) axial ^{18}F-DCFPyL PET/CT demonstrate mild to moderate radiotracer uptake in a subcarinal LN (red arrow). An IR called this finding PSMA-RADS-4, while an ER classified it PSMA-RADS-2 (i.e. likely benign, due to low-level uptake in a soft tissue site atypical for metastatic prostate cancer). The hilar and subcarinal LNs remained unchanged on follow-up imaging, suggesting these are benign in nature. All four readers classified the overall scan impression as PSMA-RADS-5: (E) axial CT, (F) axial ^{18}F-DCFPyL PET and (G) axial ^{18}F-DCFPyL PET/CT revealed intense radiotracer uptake in the right iliac bone (double thin red arrows). Apart from that, the lung lesion (^{18}F-DCFPyL whole body maximum intensity projection in A, red arrowhead) were classified as PSMA-RADS-2 by an IR. (H) Axial CT, (I) axial ^{18}F-DCFPyL PET and (J) axial ^{18}F-DCFPyL PET/CT of this lesion further confirmed the suspicion of a benign lesion (red arrowhead, most likely peripheral interstitial thickening). Follow-up imaging also corroborated this impression.

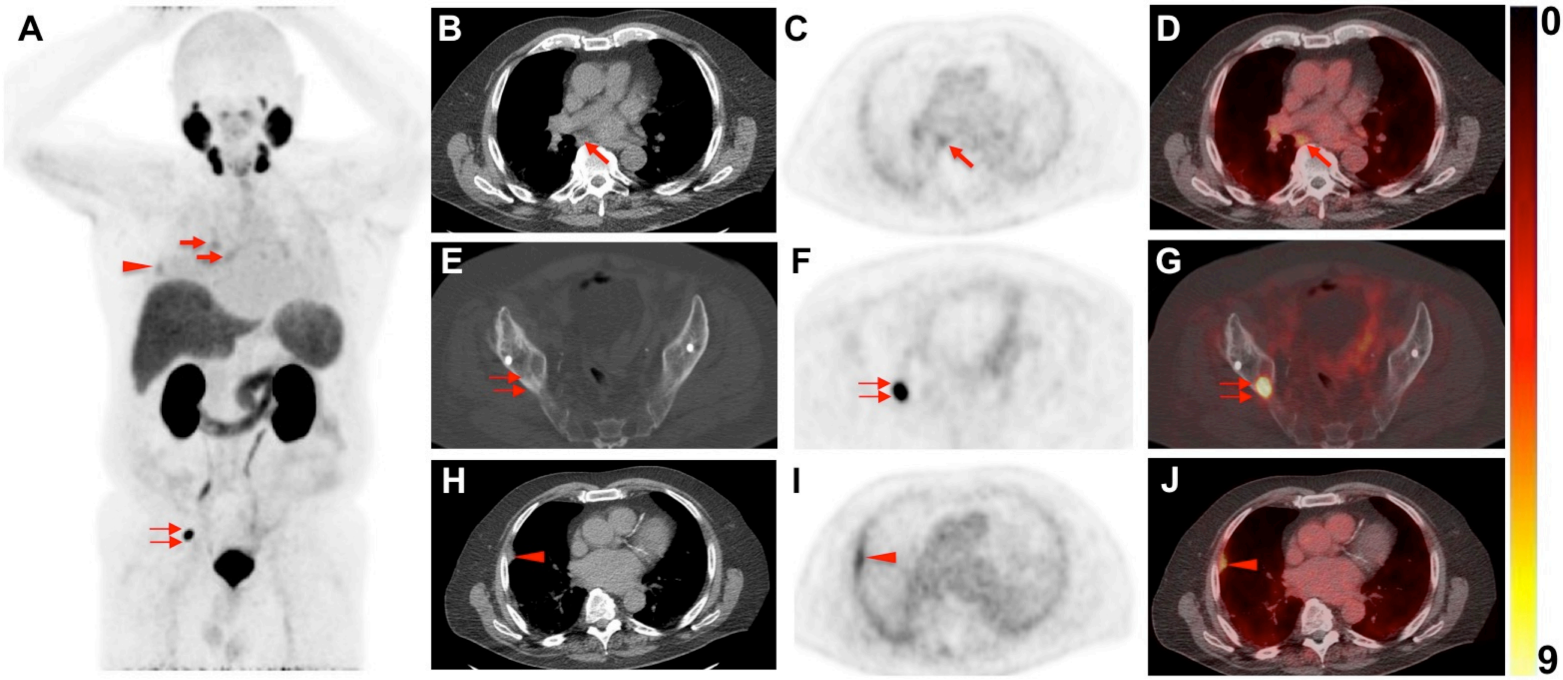

A
No. of organ metastases
5
4
3
2
1
0
ER1
ER2
IR1
IR2


B
No. of LN metastases
5
4
3
2
1
0
ER1
ER2
IR1
IR2


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
ER1
ER2
IR1
IR2
All readers
RADS-1
RADS-2
RADS-3A
RADS-3B
RADS-3C
RADS-3D
RADS-4
RADS-5


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
ER1
ER2
IR1
IR2
All readers
RADS-1
RADS-2
RADS-3A
RADS-3B
RADS-3C
RADS-3D
RADS-4
RADS-5


A
B
C
D
E
F
G
H
I
J
0
9


A
B
C
D
E
F
G
H
I
J
0
9