^{68}Ga-PSMA-11 PET/CT Interobserver Agreement for Prostate Cancer Assessments: An International Multicenter Prospective Study

Wolfgang Peter Fendler[1,2], Jeremie Calais[1,3], Martin Allen-Auerbach[1], Christina Bluemel[4], Nina Eberhardt[5], Louise Emmett[6], Pawan Gupta[1], Markus Hartenbach[7], Thomas A. Hope[8], Shozo Okamoto[9], Christian Helmut Pfob[10], Thorsten D. Pöppel[11], Christoph Rischpler[10], Sarah Schwarzenböck[12], Vanessa Stebner[11], Marcus Unterrainer[2], Helle D. Zacho[13], Tobias Maurer[14], Christian Gratzke[15], Alexander Crispin[16], Johannes Czernin[1], Ken Herrmann[1,11], and Matthias Eiber[1,10]

[1]Department of Molecular and Medical Pharmacology, David Geffen School of Medicine at UCLA, Los Angeles, California; [2]Department of Nuclear Medicine, Ludwig-Maximilians-University Munich, Munich, Germany; [3]Department of Nuclear Medicine, Bichat University Hospital, AP-HP, University of Paris VII, Paris, France; [4]Department of Nuclear Medicine, Julius-Maximilians-University of Würzburg, Würzburg, Germany; [5]Department of Nuclear Medicine, Ulm University, Ulm, Germany; [6]Department of Diagnostic Imaging, St. Vincent's Public Hospital, Sydney, Australia, and University of New South Wales, Sydney, New South Wales, Australia; [7]Division of Nuclear Medicine, Department of Biomedical Imaging and Image-guided Therapy, Medical University of Vienna, Vienna, Austria; [8]Department of Radiology and Biomedical Imaging, University of California, San Francisco, California; [9]Department of Nuclear Medicine, Hokkaido University Graduate School of Medicine, Sapporo, Japan; [10]Department of Nuclear Medicine, Klinikum Rechts der Isar, Technical University of Munich, Munich, Germany; [11]Department of Nuclear Medicine, Medical Faculty, University Duisburg-Essen, University Hospital Essen, Essen, Germany; [12]Department of Nuclear Medicine, Rostock University Medical Centre, Rostock, Germany; [13]Department of Nuclear Medicine, Aalborg University Hospital, Aalborg, Denmark; [14]Department of Urology, Klinikum Rechts der Isar, Technical University of Munich, Munich, Germany; [15]Department of Urology, Ludwig-Maximilians-University of Munich, Munich, Germany; and [16]Institute of Medical Informatics, Biometry, and Epidemiology, Ludwig-Maximilians-University Munich, Germany

See an invited perspective on this article on page 1615.

The interobserver agreement for ^{68}Ga-PSMA-11 PET/CT study interpretations in patients with prostate cancer is unknown. **Methods:** ^{68}Ga-PSMA-11 PET/CT was performed in 50 patients with prostate cancer for biochemical recurrence (n = 25), primary diagnosis (n = 10), biochemical persistence after primary therapy (n = 5), or staging of known metastatic disease (n = 10). Images were reviewed by 16 observers who used a standardized approach for interpretation of local (T), nodal (N), bone (Mb), or visceral (Mc) involvement. Observers were classified as having a low (<30 prior ^{68}Ga-PSMA-11 PET/CT studies; n = 5), intermediate (30–300 studies; n = 5), or high level of experience (>300 studies; n = 6). Histopathology (n = 25, 50%), post–external-beam radiation therapy prostate-specific antigen response (n = 15, 30%), or follow-up PET/CT (n = 10, 20%) served as a standard of reference. Observer groups were compared by overall agreement (% patients matching the standard of reference) and Fleiss' κ with mean and corresponding 95% confidence interval (CI). **Results:** Agreement among all observers was substantial for T (κ = 0.62; 95% CI, 0.59–0.64) and N (κ = 0.74; 95% CI, 0.71–0.76) staging and almost perfect for Mb (κ = 0.88; 95% CI, 0.86–0.91) staging. Level of experience positively correlated with agreement for T (κ = 0.73/0.66/0.50 for high/intermediate/low experience, respectively), N (κ = 0.80/0.76/0.64, respectively), and Mc staging (κ = 0.61/0.46/0.36, respectively). Interobserver agreement for Mb was almost perfect irrespective of prior experience (κ = 0.87/0.91/0.88, respectively). Observers with low experience, when compared with intermediate and high experience, demonstrated significantly lower median overall agreement (54% vs. 66% and 76%, P = 0.041) and specificity for T staging (73% vs. 88% and 93%, P = 0.032). **Conclusion:** The interpretation of ^{68}Ga-PSMA-11 PET/CT for prostate cancer staging is highly consistent among observers with high levels of experience, especially for nodal and bone assessments. Initial training on at least 30 patient cases is recommended to ensure acceptable performance.

Key Words: prostate cancer; agreement; reproducibility; PET/CT; PSMA; interobserver

J Nucl Med 2017; 58:1617–1623
DOI: 10.2967/jnumed.117.190827

The radioligand ^{68}Ga-PSMA-11 (Glu-NH-CO-NH-Lys-(Ahx)-[^{68}Ga(HBED-CC)]) binds with high affinity to prostate-specific membrane antigen (PSMA) (*1*). High PSMA expression together with little or no background uptake enables accurate imaging of prostate cancer by PET (*2,3*). Current evidence strongly suggests that ^{68}Ga-PSMA-11 PET/CT adds value to current diagnostic approaches (*4*). Large, mainly retrospective trials demonstrate superior detection rates and higher accuracy for the localization of biochemical recurrence when compared with morphologic imaging or choline PET/CT (*5–9*). A recent systematic review supports the use of ^{68}Ga-PSMA-11

Received Jan. 26, 2017; revision accepted Apr. 3, 2017.
For correspondence or reprints contact: Wolfgang Peter Fendler, University of California at Los Angeles, Ahmanson Translational Imaging Division, 10833 Le Conte Ave., 200 Medical Plaza, Ste. B114-61, Los Angeles, CA 90095-7370.
E-mail: wfendler@mednet.ucla.edu
Published online Apr. 13, 2017.

PET/CT in patients with biochemical recurrence and low prostate-specific antigen (PSA) values (<2 ng/mL) (*10*). Moreover, there is evidence for additional value for primary staging (*11–13*), stratification for PSMA-targeted radioligand therapy, and management of metastatic disease (*14–18*).

Multicenter trials to evaluate accuracy and impact on management of ^{68}Ga-PSMA-11 PET/CT are currently under way in Europe and the United States (e.g., NCT02940262, NCT02918357, NCT02919111).

Before widespread clinical adoption of PSMA-targeted PET imaging, its interobserver variability and agreement need to be established (*19,20*). This information has thus far not been available for ^{68}Ga-PSMA-11 PET/CT interpretations. To address this unmet need, we evaluated prospectively the interobserver agreement for ^{68}Ga-PSMA-11 PET/CT interpretations and compared findings among readers with various levels of experience.

MATERIALS AND METHODS

Patients and Standard of Reference (SOR)

From 2 institutional databases (Ludwig-Maximilians-University and Technical University Munich), 50 patients who underwent ^{68}Ga-PSMA-11 PET/CT for the following indications were selected retrospectively: biochemical recurrence (n = 25), primary diagnosis (n = 10), biochemical persistence after primary therapy (n = 5), or staging of known metastatic disease (n = 10). Patient characteristics are given in Table 1. Twenty-five of 50 patients (50%) had histologic verification of PET/CT-positive lesions. In the remaining patients, PSA response after external-beam radiation (n = 15) or ^{68}Ga-PSMA-11 PET/CT follow-up (n = 10) served as SOR. PET/CT-positive lesions were defined during a joint reading session by consensus of 2 expert readers, each with more than 1,000 prior clinical or research ^{68}Ga-PSMA-11 PET/CT interpretations. Expert readers had access to all clinical data. Cases were selected to represent clinical routine, ranging from negative cases (n = 6, 12%) to extensive disease (n = 10, 20%), with typical pitfalls. Pitfalls included ^{68}Ga-PSMA-11 PET/CT false-positive (unspecific bone uptake, n = 4; celiac ganglia, n = 2; inflammatory or post-inflammatory, n = 4; benign tumor, n = 2) and false-negative lesions (n = 8) to resemble a total of 20 challenges in 15 patients (Table 2).

The prospective study was approved by the Institutional Review Board at the Ludwig-Maximilians-University Munich, Munich, Germany, and registered in the ISRCTN registry (number ISRCTN13499475).

Image Acquisition and Reconstruction

Patient preparation and image acquisition were performed as previously described (*8,13*). In brief, ^{68}Ga-PSMA-11 was injected intravenously at a median dose of 182 MBq (interquartile range, 80 MBq) along with 20 mg of furosemide. A median tracer uptake period of 57 min (interquartile range, 14 min) was allowed before imaging with either a Siemens Biograph mCT (n = 22, 44%), Siemens True Point 64 (n = 22, 44%), or GE Discovery 690 (n = 6, 12%) scanner.

In all patients, a diagnostic CT scan (reference mAs, 200–240; 120 kV) was obtained in the portal venous phase 80 s after intravenous injection of contrast agent followed by the PET scan. All patients received diluted oral contrast.

PET images were reconstructed with an axial 168 × 168 matrix based on the TrueX algorithm (3 iterations, 21 subsets; Biograph 64) and a 256 × 256 matrix based on the TrueX algorithm (4 iterations, 8 subsets; Biograph mCT) or on the VUE Point FX algorithm (2 iterations, 36 subsets; Discovery 690).

Observers

Sixteen physicians from 13 centers located in Europe (n = 9), North America (n = 2), Asia (n = 1), and Australia (n = 1) were recruited prospectively as research participants based on their training (nuclear medicine physician or radiologist) and prior experience with PET/CT. The research participants, that is, the observers, reviewed 50 ^{68}Ga-PSMA-11 PET/CT datasets. Each dataset included diagnostic CT and attenuation-corrected PET images.

Observers reported the number of previous clinical ^{68}Ga-PSMA-11 PET/CT interpretations. On the basis of this information, observers were classified as having a low (<30 prior ^{68}Ga-PSMA PET/CT studies; n = 5), intermediate (30–300 studies; n = 5), or high level of experience (>300 studies; n = 6).

TABLE 1
Patient Characteristics

Characteristic (n = 50)	Median or absolute number
Age (y)	70 (49–83)
Indication, lesion validation, and PSA (ng/mL)	
Primary diagnosis	10 (20%)
Validation by histopathology	10 of 10 (100%)
PSA	31.8 (2.1–167.0)
Biochemical persistence after primary surgery	5 (10%)
Validation by histopathology	5 of 5 (100%)
PSA	1.1 (0.4–2.2)
Biochemical recurrence	25 (50%)
Validation by histopathology	10 of 25 (40%)
Validation by post-EBRT follow-up	15 of 25 (60%)
PSA	0.9 (0.2–26.3)
Staging of metastatic disease	10 (20%)
Validation by ^{68}Ga-PSMA-11 PET/CT follow-up	10 of 10 (100%)
PSA	71.9 (0.9–9237.0)
Tumor stage	
^{68}Ga-PSMA-11 PET/CT positive for prostate cancer	44 (88%)
Local tumor	9 (18%)
N positive	30 (60%)
Mb positive	15 (30%)
Mc positive	6 (12%)

EBRT = external-beam radiation therapy.
Data are median, with range in parentheses, or absolute number, with percentages in parentheses.

Guidelines for Visual Interpretation

A written guide (supplemental materials [available at http://jnm.snmjournals.org]), 4 teaching cases, an electronic case report form, and 1 test patient dataset with disclosed data entries were provided to each observer. In addition, observers were asked to learn about ^{68}Ga-PSMA-11 PET/CT pitfalls (*21*) and the typical nomenclature for lymph node regions (*22*) to achieve best possible agreement.

The following patient information was disclosed to each observer before image interpretation: indication (biochemical recurrence, primary diagnosis, biochemical persistence after primary therapy, staging of

TABLE 2
Notable Pitfalls for ^{68}Ga-PSMA-11 PET/CT Interpretation Included in Study

Pitfall	No. of patients (%)	False-positive rate (%) separate for patients	False-negative rate (%) separate for patients
Frequently false ^{68}Ga-PSMA-11 PET/CT–positive lesions			
Bone degenerative/posttraumatic/unspecific	4 (8%)	2 (13%), 2 (13%), 11 (69%), 13 (81%)	
Celiac ganglia	2 (4%)	1 (6%), 9 (56%)	
Mediastinal lymph node sarcoidosis	2 (4%)	1 (6%), 1 (6%)	
Vertebral hemangioma	1 (2%)	8 (50%)	
Pulmonary tuberculosis cavity	1 (2%)	NA: lung metastasis in other location	
Postinflammatory uptake in lymph nodes	1 (2%)	7 (44%)	
Benign thyroid nodule	1 (2%)	3 (19%)	
Frequently false ^{68}Ga-PSMA-11 PET/CT–negative lesions			
Lymph nodes metastases with low uptake	3 (6%)		4 (25%), 8 (50%), 13 (81%)
Metastases at untypical location (cartilage, penis)	2 (4%)		9 (56%), 16 (100%)
Hepatic metastases with low uptake	2 (4%)		11 (69%), 12 (75%)
Bone metastases with low uptake	1 (2%)		6 (38%)

NA = not applicable.
Absolute number and proportion are given for patients and false-positive/-negative rate.

metastatic disease), age (y), weight (kg), injected dose (MBq), uptake time (min), PET/CT device, and PSA level (ng/mL). Observers were masked to all other clinical data. Visual image interpretation for the presence or absence of malignant disease was reported for predefined categories (Supplemental Table 1).

Semiquantitative Measurements

Each observer recorded SUV_{max} for 1 diseased target region per T, N, Mb, and Mc category. The target region for SUV measurement was automatically identified in the electronic case report form.

Each observer measured background activity by defining SUV_{max} and SUV_{mean} using a 1.5-cm-diameter circular region of interest placed in the center of the aortic arch and the left gluteus muscle. To exclude variability among different image software used for interpretation, observers were asked to repeat tumor and background SUV for 1 test patient dataset to exclude deviation greater than 10%.

Statistical Analysis and Reference Standard

For binary data, agreement among observer groups was evaluated using Fleiss' κ (*23*). For nonbinary data with more than 10 observations, agreement among observer groups was evaluated by intraclass correlation coefficient (ICC) using a 2-way mixed model for absolute agreement (average measures) (*24*). Ninety-five percent confidence intervals (CIs) are reported for κ and ICC values. Interpretation of κ and ICC was based on a classification provided by Landis and Koch (*25*): 0.0, poor; 0.0–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; 0.81–1.00, almost-perfect reproducibility.

Overall agreement, defined as complete agreement of an observer for all categories (T, N, Mb, Mc), and sensitivity and specificity compared with the SOR, respectively, were calculated for each observer. Group median and range were reported for overall agreement, sensitivity, and specificity. Difference between 2 groups was assessed by Student *t* test. The significance level was 5%.

Discrepancies in semiquantitative measurements between observer groups and the SOR were expressed as mean difference (Δ) ± SD. Statistical analyses were performed using R software (R Core Team 2015, R Foundation for Statistical Computing) with the package "irr" (Gamer et al., version 0.84) for Fleiss' κ and SPSS (version 15.0; SPSS Inc.) for all other statistical analyses.

At least substantial agreement for visual and semiquantitative interpretation of all scans for the 3 major staging categories (T, N, Mb) was defined as acceptable performance.

RESULTS

Patient Characteristics

Table 1 summarizes the patient characteristics. ^{68}Ga-PSMA-11 PET/CT studies were interpreted as positive for prostate cancer presence in 44 of 50 (88%) patients by the reference readers: local tumor was present in 9 patients (18%); 30 patients (60%) had lymph node (N)–positive disease, whereas 15 (30%) and 6 (12%) were staged as bone (Mb) and organ (Mc) positive, respectively.

Image Interpretation: Interobserver Agreement

The interobserver agreement for visual image interpretation is shown in Figure 1A and Table 3. Highly experienced observers agreed substantially or almost perfectly for all categories (T, N, Mb, Mc). Intermediate- and low-experienced observers provided substantially or almost-perfectly reproducible assessments for the T, N, and Mb categories and N and Mb categories, respectively.

Interobserver agreement was analyzed separately for patients with biochemical recurrence or persistence after primary definitive treatment: high-experienced observers agreed substantially or almost perfectly for all categories (T, N, Mb, Mc) whereas intermediate- and low-experienced observers agreed substantially or almost perfectly only for the N and Mb categories, and Mb category, respectively.

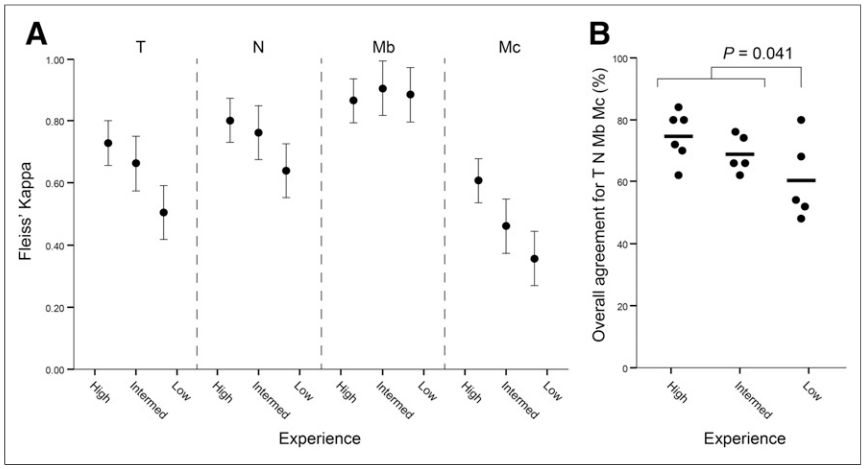

FIGURE 1. Interobserver agreement for visual image interpretation. (A) Fleiss' κ with corresponding 95% CI for T, N, Mb, and Mc staging is shown separately for observer groups. (B) Overall agreement for T, N, Mb, and Mc staging drawn separately for observer groups. High- and intermediate-experience groups had significantly higher agreement than low-experience group ($P = 0.041$).

Image Interpretation: Comparison to SOR

Median overall agreement with SOR for T, N, Mb, and Mc staging was 69% (range, 48–84) for the entire group of observers. High- or intermediate-experienced observers performed significantly better than low-experienced observers for T, N, Mb, and Mc staging (median, 76 or 66 vs. 54%, $P = 0.041$; Fig. 1B).

Table 4 summarizes sensitivity and specificity for the entire group and separated for low-, intermediate-, or high-experienced observers, each stratified by staging category. All observer groups were highly sensitive in detection of local tumor. However, on the basis of a higher rate for false-positive local findings, median specificity was significantly lower for observers with low versus intermediate or high experience (73% vs. 88% and 93%, $P = 0.032$). For lymph node and bone metastases performance compared with the SOR was almost identical (all $P > 0.05$). In assessing organ metastases, sensitivity was slightly higher for high-experienced observers (median, 58%) versus observers with intermediate or low experience (median, 50%).

Three patient examples for low degree of observer agreement are given in Figure 2. Notable sources for disagreement were among other false-negative findings due to low ^{68}Ga-PSMA-11 uptake and false-positive findings due to ^{68}Ga-PSMA-11 uptake in benign entities (Table 2). For instance, low ^{68}Ga-PSMA-11 uptake ($SUV_{max} < 5$) in lymph node metastases resulted in false-negative findings in 3 patients for 4 (25%), 8 (50%), and 13 (81%) observers, respectively (Fig. 2A). Degenerative or posttraumatic bone uptake resulted in false-positive Mb stage in 4 patients for 2 (13%), 2 (13%), 11 (69%), and 13 (81%) observers, respectively. Hepatic metastases resulted in false-negative Mc stage in 2 patients for 11 (69%) and 12 (75%) observers, respectively. Metastases to the thyroid cartilage and to the penis in 2 patients were missed by 9 (56%) and 16 (100%) observers, respectively, resulting in false-negative Mc stage. Celiac ganglia with high ^{68}Ga-PSMA-11 uptake in 2 patients resulted in false-positive N stage by 1 (6%) and 9 (56%) observers, respectively.

Semiquantitative Measurements

Interobserver agreement including mean Δ differences for SUV measurement is given in Table 5. Agreement was almost perfect for SUV_{max} of local tumor, lymph node, and bone metastases. Agreement was not associated with tumor lesion uptake (ICC, 1.00 for $SUV_{max} < 10$; 0.94 for $10 \leq SUV_{max} < 20$; 0.98 for $SUV_{max} \geq 20$). SUV_{max} and SUV_{mean} of mediastinal blood pool and muscle were highly reproducible. Figure 3 illustrates agreement among individual SUV measurements.

Overall, observers with high or intermediate experience fulfilled our criteria for acceptable performance, whereas observers with low experience did not, based on fair agreement for local staging.

TABLE 3
Interobserver Agreement for Visual Image Interpretation

Dataset	T	N	Mb	Mc
All patients ($n = 50$)				
High	0.73* (0.66–0.80)	0.80* (0.73–0.87)	0.87* (0.79–0.94)	0.61* (0.54–0.68)
Intermediate	0.66* (0.58–0.75)	0.76* (0.67–0.85)	0.91* (0.82–0.99)	0.46 (0.37–0.55)
Low	0.50 (0.42–0.59)	0.64* (0.55–0.73)	0.88* (0.80–0.97)	0.36 (0.27–0.44)
Any	0.62* (0.59–0.64)	0.74* (0.71–0.76)	0.88* (0.86–0.91)	0.46 (0.44–0.49)
BCR and BCP ($n = 30$)				
High	0.73* (0.64–0.82)	0.81* (0.72–0.91)	0.84* (0.75–0.93)	0.65* (0.56–0.74)
Intermediate	0.55 (0.43–0.66)	0.75* (0.64–0.86)	0.76* (0.65–0.87)	0.49 (0.37–0.60)
Low	0.35 (0.24–0.46)	0.59 (0.47–0.70)	0.92* (0.81–1.00)	0.45 (0.34–0.57)
Any	0.51 (0.48–0.54)	0.72* (0.69–0.76)	0.84* (0.80–0.87)	0.48 (0.44–0.51)

*Substantial to almost-perfect reproducibility.
BCR = biochemical recurrence; BCP = biochemical persistence.
Data are mean Fleiss' κ, with 95% confidence intervals in parentheses.

TABLE 4
Sensitivity and Specificity for Observer with High, Intermediate, or Low Experience and for All Observers (Any)

	T		N		Mb		Mc	
Experience	SE	SP	SE	SP	SE	SP	SE	SP
High	100 (100–100)	93 (78–98)	95 (93–97)	85 (75–100)	100 (93–100)	93 (32–100)	58 (33–67)	98 (95–100)
Intermediate	100 (100–100)	88 (78–98)	93 (93–100)	75 (75–95)	100 (100–100)	91 (86–94)	50 (50–67)	95 (91–100)
Low	100 (89–100)	73 (59–100)	93 (93–100)	90 (35–95)	100 (93–100)	94 (89–97)	50 (33–67)	95 (77–100)
Any	100 (89–100)	87 (59–100)	93 (93–100)	85 (35–100)	100 (93–100)	91 (32–100)	50 (33–67)	97 (77–100)

SE = sensitivity; SP = specificity.
Median and range are given in percentage, separately for T, N, Mb, or Mc staging.

DISCUSSION

This prospective study on 50 ^{68}Ga-PSMA-11 PET/CT scans demonstrated that readings are highly reproducible for high- and intermediate-experienced observers. Observers in the low-experience group provided highly reproducible reads for bone metastases but achieved lower agreement for local tumor, lymph node, and organ metastases assessments.

Semiquantitative analyses of tumor lesions and background activity was highly reproducible for all levels of observer experience. On the basis of our predefined criteria, we recommend initial training on at least 30 representative patient cases to reach acceptable diagnostic performance for clinical and research interpretations of ^{68}Ga-PSMA-11 PET/CT scans. Training cases should include routine findings (ranging from unremarkable to extensive disease) and typical pitfalls, such as PET-positive ganglia or degenerative/posttraumatic bone lesions.

Interobserver agreement is an important aspect of clinical applicability. ^{68}Ga-PSMA-11 PET/CT scan interpretation is not without pitfalls: PSMA expression has been observed in tissues other than prostate cancer. Common examples are ganglia, hemangioma, Paget's bone disease, and other benign and malignant tumors (*26–32*). Sources of misinterpretations include normal and variable PSMA ligand uptake due to background activity in salivary glands, liver, spleen, small intestine, colon, and kidney or in the urinary system. In general, the list of false-positive pitfalls is still evolving, prompting any clinician to stay vigilant with the current literature. Visceral metastases to the liver can occasionally exhibit no to low uptake, which cannot always be differentiated reliably from background activity. Approximately 5%–10% of all primary prostate cancers as well as their metastases do not exhibit significant PSMA expression (*11,33*), stressing the importance of reader experience for interpretation of the PET/CT study.

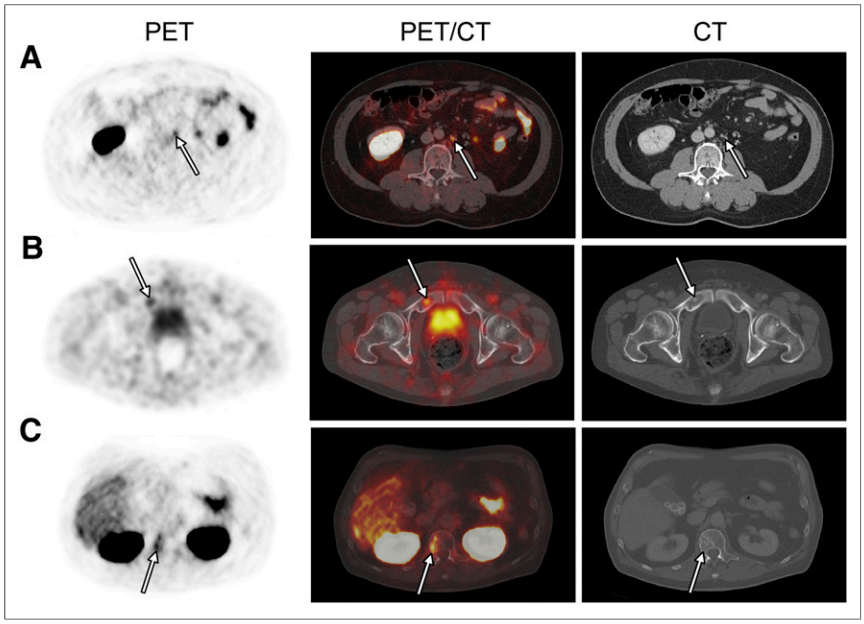

FIGURE 2. ^{68}Ga-PSMA-11 PET/CT lymph node and bone findings with low degree of interobserver agreement. Axial fused ^{68}Ga-PSMA-11 PET/CT (middle), PET (left), and CT (right) are shown for 3 patients (1 row per patient). (A) Retroperitoneal lymph node metastasis with low tracer uptake (SUV_{max}, 4.6; PET arrow) was confirmed by histopathology, however judged negative by 4 of 16 (25%) observers (false negative). (B) Pelvic bone metastasis with small sclerotic lesion (CT arrow) but faint ^{68}Ga-PSMA-11 uptake (SUV_{max}, 3.7; PET arrow) was confirmed by adequate PSA drop after external-beam radiation therapy, however judged negative by 6 of 16 (38%) observers (false negative). (C) Right L1 hemangioma with moderate uptake (SUV_{max}, 5.6; PET arrow) was judged positive by 8 of 16 (50%) observers (false positive). Bone metastasis was ruled out by follow-up imaging.

To reduce error rates, reported studies used consensus readings by multiple physicians (*7–9,11,34,35*). However, this does not solve the issue of observer variability in the clinical setting. The current cases, selected from 2 databases, contained a considerable proportion of pitfalls (Table 2). This approach was chosen to also challenge readers with difficult cases. Despite this additional level of difficulty, readers with intermediate and high experience levels achieved substantial to almost-perfect agreement for all clinically relevant categories.

Intermediate- and low-experienced observers demonstrated substantial or almost-perfect agreement for the N and Mb categories. This may be a result of high tumor-to-background uptake for ^{68}Ga-PSMA-11 and basic understanding of common metastatic pathways. False-positive findings for local involvement with potential implication on management, such as substantial changes of a salvage

TABLE 5
Interobserver Agreement for SUVs

Tissue	ICC	MeanΔ ± SD		
		High	Intermediate	Low
Tumor SUV_{max}				
T	0.99 (0.98–1.00)	2.2 ± 3.5	2.1 ± 3.7	2.6 ± 4.0
N	0.99 (0.97–1.00)	2.2 ± 4.8	1.2 ± 3.4	0.9 ± 2.8
Mb	0.96 (0.92–0.99)	4.3 ± 8.7	4.9 ± 8.9	2.7 ± 3.7
Blood pool				
SUV_{mean}	0.97 (0.96–0.98)	0.2 ± 0.2	0.1 ± 0.2	0.2 ± 0.2
SUV_{max}	0.95 (0.93–0.97)	0.3 ± 0.3	0.3 ± 0.3	0.3 ± 0.3
Muscle				
SUV_{mean}	0.94 (0.91–0.96)	0.1 ± 0.1	0.1 ± 0.1	0.1 ± 0.1
SUV_{max}	0.83 (0.76–0.89)	0.3 ± 0.3	0.2 ± 0.2	0.2 ± 0.2

ICC and corresponding 95% CI are given for categories with more than 10 observations. Mean absolute difference (Δ) ± SD when compared with findings of reference standard was calculated separately for reader groups.
Data in parentheses are 95% CIs.

radiation therapy plan, occurred more often in the low-experience group. Thus observers with low experience (<30 previous ^{68}Ga-PSMA-11 PET/CT readings) showed only moderate interobserver agreement for T staging with somewhat reduced specificity. Indeed, the judgment of local tumor can be challenging because small recurrence frequently occurs near the base of the bladder, causing problems with signal overlay by excreted tracer; and background uptake in normal prostate especially in benign hypertrophy as well as after local radiation therapy decreases signal-to-noise ratio (*13,35,36*).

Agreement for Mc staging was lower than for T, N, and Mb for all observer groups. In particular, intermediate- and low-experienced observers exhibited only fair to moderate agreement. This is likely due to a low number of observations (6 patients were true positive) combined with the relatively high portion of pitfalls: metastasis in the thyroid cartilage was missed by more than half of observers, especially those with intermediate and low experience. In general, false-negative visceral findings were triggered by reader bias due to low incidence (e.g., 5% in patients with biochemical recurrence (*8*)) and absent or low PSMA expression (*37–39*) impeding ^{68}Ga-PSMA-11 PET interpretation.

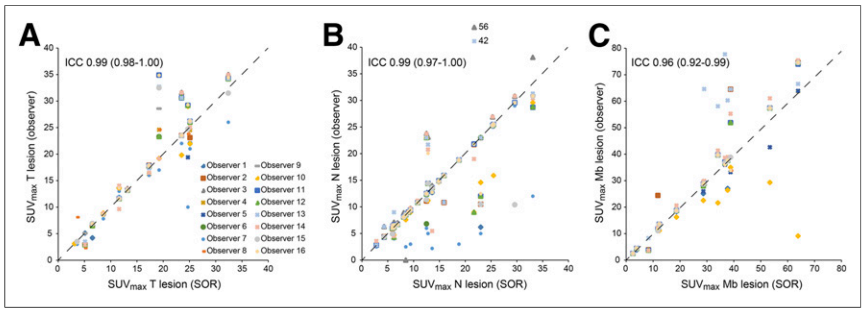

FIGURE 3. Interobserver agreement for tumor SUV_{max} shown separately for T (A, n = 184), N (B, n = 401), and Mb (C, n = 203) lesions. SUVs were sorted by SUV obtained from SOR. Dashed diagonal lines indicate perfect agreement. ICC and corresponding 95% CI are given. Two y-axis outliers were drawn in relative position outside scale, and absolute Y value is given. Observers 3 and 7 systematically measured false target bone region and were excluded from SUV_{max} Mb analysis.

Observer agreement levels of the current study are in line with PET procedures using high-affinity radioligands. Two recent studies reported almost-perfect reproducibility for ^{68}Ga-DOTATATE PET/CT interpretations (κ = 0.82 and 0.80) in patients with neuroendocrine tumors (*40,41*). Thus, interpretations of radioligand PET/CT studies in patients with neuroendocrine and prostate cancer, respectively, are equally robust. ^{68}Ga-DOTATATE and ^{68}Ga-PSMA-11 PET/CT are characterized by specific and high tumor signal. These hallmarks contribute to a high level of reader agreement even after short training period.

The present study has several limitations. First, observers were grouped on the basis of experience with ^{68}Ga-PSMA-11 PET/CT interpretation. However, the skill of a reader is determined by multiple factors, including clinical knowledge and general experience in imaging of prostate cancer. This may have led to a relatively broad variance in overall agreement, for example, observed for the low-experienced observers in our study (Fig. 1B). Second, the sensitivities reported might be overestimated because it is difficult to identify false-negative lesions especially in the setting of recurrence when histologic validation is image driven. Third, lymph node metastases within versus outside the pelvis were not separated in our staging system, which was organ-focused to analyze findings based on their PET/CT appearance. American Joint Committee on Cancer staging focuses on patient prognosis and thus discriminates intra- from extrapelvic lymph node metastases. Fourth, intraobserver agreement was not assessed, which might have given insight into reliability and confidence for individual judgments. However, applicability of our findings is supported by selection of representative patients and pitfalls as well as inclusion of a high number of observers from Europe, the United States, Asia, and Australia.

CONCLUSION

Both visual and semiquantitative ^{68}Ga-PSMA-11 PET/CT interpretations in

prostate cancer patients are highly reproducible among observers with intermediate and high experience. Our findings indicate acceptable reader performance after initial training on at least 30 representative patient cases.

DISCLOSURE

This study was partially funded by the U.S. Department of Energy, Office of Science Award DE-SC0012353. Wolfgang Peter Fendler received a scholarship from the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG, grant 807122). Jeremie Calais received a grant from the Fondation ARC pour la recherche sur le cancer (grant no. SAE20160604150). No other potential conflict of interest relevant to this article was reported.

REFERENCES

1. Eder M, Schafer M, Bauder-Wust U, et al. ^{68}Ga-complex lipophilicity and the targeting property of a urea-based PSMA inhibitor for PET imaging. *Bioconjug Chem.* 2012;23:688–697.
2. Bostwick DG, Pacelli A, Blute M, Roche P, Murphy GP. Prostate specific membrane antigen expression in prostatic intraepithelial neoplasia and adenocarcinoma: a study of 184 cases. *Cancer.* 1998;82:2256–2261.
3. Mannweiler S, Amersdorfer P, Trajanoski S, Terrett JA, King D, Mehes G. Heterogeneity of prostate-specific membrane antigen (PSMA) expression in prostate carcinoma with distant metastasis. *Pathol Oncol Res.* 2009;15:167–172.
4. Perera M, Papa N, Christidis D, et al. Sensitivity, specificity, and predictors of positive ^{68}Ga-prostate-specific membrane antigen positron emission tomography in advanced prostate cancer: a systematic review and meta-analysis. *Eur Urol.* 2016;70:926–937.
5. Afshar-Oromieh A, Malcher A, Eder M, et al. PET imaging with a [^{68}Ga]gallium-labelled PSMA ligand for the diagnosis of prostate cancer: biodistribution in humans and first evaluation of tumour lesions. *Eur J Nucl Med Mol Imaging.* 2013;40:486–495.
6. Afshar-Oromieh A, Haberkorn U, Eder M, Eisenhut M, Zechmann CM. [^{68}Ga]gallium-labelled PSMA ligand as superior PET tracer for the diagnosis of prostate cancer: comparison with ^{18}F-FECH. *Eur J Nucl Med Mol Imaging.* 2012;39:1085–1086.
7. Afshar-Oromieh A, Avtzi E, Giesel FL, et al. The diagnostic value of PET/CT imaging with the ^{68}Ga-labelled PSMA ligand HBED-CC in the diagnosis of recurrent prostate cancer. *Eur J Nucl Med Mol Imaging.* 2015;42:197–209.
8. Eiber M, Maurer T, Souvatzoglou M, et al. Evaluation of hybrid ^{68}Ga-PSMA ligand PET/CT in 248 patients with biochemical recurrence after radical prostatectomy. *J Nucl Med.* 2015;56:668–674.
9. Morigi JJ, Stricker PD, van Leeuwen PJ, et al. Prospective comparison of ^{18}F-fluoromethylcholine versus ^{68}Ga-PSMA PET/CT in Prostate cancer patients who have rising PSA after curative treatment and are being considered for targeted therapy. *J Nucl Med.* 2015;56:1185–1190.
10. Evangelista L, Briganti A, Fanti S, et al. New clinical indications for ^{18}F/^{11}C-choline, new tracers for positron emission tomography and a promising hybrid device for prostate cancer staging: a systematic review of the literature. *Eur Urol.* 2016;70:161–175.
11. 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.
12. van Leeuwen PJ, Emmett L, Ho B, et al. Prospective evaluation of 68gallium-PSMA positron emission tomography/computerized tomography for preoperative lymph node staging in prostate cancer. *BJU Int.* 2017;119:209–215.
13. Fendler WP, Schmidt DF, Wenter V, et al. ^{68}Ga-PSMA PET/CT Detects the location and extent of primary prostate cancer. *J Nucl Med.* 2016;57:1720–1725.
14. Kratochwil C, Giesel FL, Eder M, et al. [^{177}Lu]lutetium-labelled PSMA ligand-induced remission in a patient with metastatic prostate cancer. *Eur J Nucl Med Mol Imaging.* 2015;42:987–988.
15. Rahbar K, Ahmadzadehfar H, Kratochwil C, et al. German multicenter study investigating ^{177}Lu-PSMA-617 radioligand therapy in advanced prostate cancer patients. *J Nucl Med.* 2017;58:85–90.
16. Fendler WP, Reinhardt S, Ilhan H, et al. Preliminary experience with dosimetry, response and patient reported outcome after ^{177}Lu-PSMA-617 therapy for metastatic castration-resistant prostate cancer. *Oncotarget.* 2017;8:3581–3590.
17. Heck MM, Retz M, D'Alessandria C, et al. Systemic radioligand therapy with ^{177}Lu labeled prostate specific membrane antigen ligand for imaging and therapy in patients with metastatic castration resistant prostate cancer. *J Urol.* 2016;196:382–391.
18. Pyka T, Okamoto S, Dahlbender M, et al. Comparison of bone scintigraphy and ^{68}Ga-PSMA PET for skeletal staging in prostate cancer. *Eur J Nucl Med Mol Imaging.* 2016;43:2114–2121.
19. Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. Getting better but still not good. *JAMA.* 1995;274:645–651.
20. Bankier AA, Levine D, Halpern EF, Kressel HY. Consensus interpretation in imaging research: is there a better way? *Radiology.* 2010;257:14–17.
21. Rauscher I, Maurer T, Fendler WP, Sommer WH, Schwaiger M, Eiber M. ^{68}Ga-PSMA ligand PET/CT in patients with prostate cancer: how we review and report. *Cancer Imaging.* 2016;16:14.
22. Taylor A, Rockall AG, Powell ME. An atlas of the pelvic lymph node regions to aid radiotherapy target volume definition. *Clin Oncol (R Coll Radiol).* 2007;19:542–550.
23. Hale CA, Fleiss JL. Interval estimation under two study designs for kappa with binary classifications. *Biometrics.* 1993;49:523–534.
24. Scheffe H. *The Analysis of Variance.* New York, NY: Wiley; 1959.
25. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159–174.
26. Krohn T, Verburg FA, Pufe T, et al. [^{68}Ga]PSMA-HBED uptake mimicking lymph node metastasis in coeliac ganglia: an important pitfall in clinical practice. *Eur J Nucl Med Mol Imaging.* 2015;42:210–214.
27. Sawicki LM, Buchbender C, Boos J, et al. Diagnostic potential of PET/CT using a ^{68}Ga-labelled prostate-specific membrane antigen ligand in whole-body staging of renal cell carcinoma: initial experience. *Eur J Nucl Med Mol Imaging.* 2017;44:102–107.
28. Verburg FA, Krohn T, Heinzel A, Mottaghy FM, Behrendt FF. First evidence of PSMA expression in differentiated thyroid cancer using [^{68}Ga]PSMA-HBED-CC PET/CT. *Eur J Nucl Med Mol Imaging.* 2015;42:1622–1623.
29. Demirci E, Ocak M, Kabasakal L, et al. ^{68}Ga-PSMA PET/CT imaging of metastatic clear cell renal cell carcinoma. *Eur J Nucl Med Mol Imaging.* 2014;41:1461–1462.
30. Silver DA, Pellicer I, Fair WR, Heston WD, Cordon-Cardo C. Prostate-specific membrane antigen expression in normal and malignant human tissues. *Clin Cancer Res.* 1997;3:81–85.
31. Chang SS, O'Keefe DS, Bacich DJ, Reuter VE, Heston WD, Gaudin PB. Prostate-specific membrane antigen is produced in tumor-associated neovasculature. *Clin Cancer Res.* 1999;5:2674–2681.
32. Chang SS, Reuter VE, Heston WD, Bander NH, Grauer LS, Gaudin PB. Five different anti-prostate-specific membrane antigen (PSMA) antibodies confirm PSMA expression in tumor-associated neovasculature. *Cancer Res.* 1999;59:3192–3198.
33. Budäus L, Leyh-Bannurah SR, Salomon G, et al. Initial Experience of ^{68}Ga-PSMA PET/CT imaging in high-risk prostate cancer patients prior to radical prostatectomy. *Eur Urol.* 2016;69:393–396.
34. Verburg FA, Pfister D, Heidenreich A, et al. Extent of disease in recurrent prostate cancer determined by [^{68}Ga]PSMA-HBED-CC PET/CT in relation to PSA levels, PSA doubling time and Gleason score. *Eur J Nucl Med Mol Imaging.* 2016;43:397–403.
35. Meredith G, Wong D, Yaxley J, et al. The use of 68 Ga-PSMA PET CT in men with biochemical recurrence after definitive treatment of acinar prostate cancer. *BJU Int.* 2016;118:49–55.
36. Eiber M, Weirich G, Holzapfel K, et al. Simultaneous ^{68}Ga-PSMA HBED-CC PET/MRI improves the localization of primary prostate cancer. *Eur Urol.* 2016;70:829–836.
37. Laidler P, Dulinska J, Lekka M, Lekki J. Expression of prostate specific membrane antigen in androgen-independent prostate cancer cell line PC-3. *Arch Biochem Biophys.* 2005;435:1–14.
38. Parimi V, Goyal R, Poropatich K, Yang XJ. Neuroendocrine differentiation of prostate cancer: a review. *Am J Clin Exp Urol.* 2014;2:273–285.
39. Yuan TC, Veeramani S, Lin MF. Neuroendocrine-like prostate cancer cells: neuroendocrine transdifferentiation of prostate adenocarcinoma cells. *Endocr Relat Cancer.* 2007;14:531–547.
40. Deppen SA, Liu E, Blume JD, et al. Safety and efficacy of ^{68}Ga-DOTATATE PET/CT for diagnosis, staging and treatment management of neuroendocrine tumors. *J Nucl Med.* 2016;57:708–714.
41. Fendler WP, Barrio M, Spick C, et al. ^{68}Ga-DOTATATE PET/CT interobserver agreement for neuroendocrine tumor assessments: results from a prospective study on 50 patients. *J Nucl Med.* 2017;58:307–311.