
High Interobserver Agreement for the Standardized Reporting System SSTR-RADS 1.0 on Somatostatin Receptor PET/CT

Rudolf A. Werner^{*1}, Thorsten Derlin^{*1}, Steven P. Rowe², Lena Bundschuh³, Gabriel T. Sheikh⁴, Martin G. Pomper², Sebastian Schulz¹, Takahiro Higuchi^{5,6}, Andreas K. Buck⁵, Frank M. Bengel¹, Ralph A. Bundschuh^{*3}, and Constantin Lapa^{*4}

¹Department of Nuclear Medicine, Hannover Medical School, Hannover, Germany; ²Russell H. Morgan Department of Radiology and Radiological Science, Division of Nuclear Medicine, Johns Hopkins School of Medicine, Baltimore, Maryland; ³Department of Nuclear Medicine, University Medical Hospital Bonn, Medical Faculty, Bonn, Germany; ⁴Nuclear Medicine, Medical Faculty, University of Augsburg, Augsburg, Germany; ⁵Department of Nuclear Medicine, University Hospital Würzburg, Würzburg, Germany; and ⁶Okayama University Graduate School of Medicine, Dentistry, and Pharmaceutical Sciences, Okayama, Japan

Recently, a standardized framework system for interpreting somatostatin receptor (SSTR)-targeted PET/CT, termed the SSTR reporting and data system (RADS) 1.0, was introduced, providing reliable standards and criteria for SSTR-targeted imaging. We determined the interobserver reliability of SSTR-RADS for interpretation of ⁶⁸Ga-DOTATOC PET/CT scans in a multicentric, randomized setting. **Methods:** A set of 51 randomized ⁶⁸Ga-DOTATOC PET/CT scans was independently assessed by 4 masked readers with different levels of experience (2 experienced readers and 2 inexperienced readers) trained on the SSTR-RADS 1.0 criteria (based on a 5-point scale from 1 [definitively benign] to 5 [high certainty that neuroendocrine neoplasia is present]). For each scan, SSTR-RADS scores were assigned to a maximum of 5 target lesions (TLs). An overall scan impression based on SSTR-RADS was indicated, and interobserver agreement rates on a TL-based, on an organ-based, and on an overall SSTR-RADS score-based level were computed. The readers were also asked to decide whether peptide receptor radionuclide therapy (PRRT) should be considered on the basis of the assigned RADS scores. **Results:** Among the selected TLs, 153 were chosen by at least 2 readers (all 4 readers selected the same TLs in 58 of 153 [37.9%] instances). The interobserver agreement for SSTR-RADS scoring among identical TLs was good (intraclass correlation coefficient [ICC] ≥ 0.73 for 4, 3, and 2 identical TLs). For lymph node and liver lesions, excellent interobserver agreement rates were derived (ICC, 0.91 and 0.77, respectively). Moreover, the interobserver agreement for an overall scan impression based on SSTR-RADS was excellent (ICC, 0.88). The SSTR-RADS-based decision to use PRRT also demonstrated excellent agreement, with an ICC of 0.80. No significant differences between experienced and inexperienced readers for an overall scan impression and TL-based SSTR-RADS scoring were observed ($P \geq 0.18$), thereby suggesting that SSTR-RADS seems to be readily applicable even for less experienced readers. **Conclusion:** SSTR-RADS-guided assessment demonstrated a high concordance rate, even among readers with

different levels of experience, supporting the adoption of SSTR-RADS for trials, clinical routine, or outcome studies.

Key Words: SSTR-RADS; reporting and data system; RADS; neuroendocrine tumor; somatostatin receptor; peptide receptor radionuclide therapy

J Nucl Med 2021; 62:514–520

DOI: 10.2967/jnumed.120.245464

The theranostic concept in patients with neuroendocrine neoplasms (NENs) is based on somatostatin receptor (SSTR)-directed imaging followed by systemic radiation with β -emitting radionuclides linked to the identical amino acid peptide used for imaging (1–5). Of note, the recent Food and Drug Administration approval of the most commonly used theranostic twin, ⁶⁸Ga-/¹⁷⁷Lu-DOTA-TATE, will most likely lead to an increased initiation of SSTR-targeting peptide receptor radionuclide therapies (PRRTs) all over the United States (6). As such, it is anticipated that endoradiotherapy for NENs will experience increasing use (7,8). Accurate scan interpretation, however, is of the utmost importance in triaging patients for theranostic interventions (9), in guiding the referring oncologist when handling challenging cases, or in recommending an appropriate work-up (e.g., biopsy of non-NEN malignant lesions) (10). With the rapid adoption and growth of SSTR-directed PET, numerous pitfalls have recently been described, such as radiotracer-avid degenerative structures in bone, accumulation in inflammatory diseases, or intense SSTR expression in a large variety of non-NEN tumors exhibiting discernible radiotracer uptake, such as meningioma, breast cancer, or hemangioendothelioma (10–12). As such, a standardized framework system that increases the reader's confidence in separating pathologic from physiologic findings and identifying putative sites of disease that may require further evaluation would be of significant value. Thus, in analogy to other reporting and data systems (RADS) for specific organs such as liver, thyroid, and breast (13–15), SSTR-RADS as a novel RADS classification for SSTR-targeted PET/CT has recently been introduced (16). As a generalizable harmonization system, SSTR-RADS should tackle the aforementioned issues regarding a

Received Mar. 18, 2020; revision accepted Jul. 20, 2020.

For correspondence or reprints contact: Rudolf A. Werner, Department of Nuclear Medicine, Hannover Medical School, Carl-Neuberg-Strasse 1, 30625 Hannover, Germany.

E-mail: werner.rudolf@mh-hannover.de

*Contributed equally to this work.

Published online Aug. 28, 2020.

COPYRIGHT © 2021 by the Society of Nuclear Medicine and Molecular Imaging.

thorough scan interpretation, thereby increasing the reader's level of confidence and easing communication between nuclear medicine specialists and referring oncologists so that they can initiate an appropriate work-up (16). Most importantly, such harmonization systems should help the interpreting physician identify the right patient for the right theranostic intervention at the right time (10). However, before a more widespread adoption in clinical routine or implementation in larger clinical trials, the reproducibility of such standardized reporting systems must be proven (9,17), as has recently been demonstrated for other theranostic radiotracers (18–20). Thus, in the present study, we aimed to determine the interobserver reliability of SSTR-RADS for interpretation of SSTR-targeted PET/CT scans in a real-world scenario mimicking the clinical workflow in a busy molecular imaging center.

MATERIALS AND METHODS

Patient Population

In total, 51 patients (Table 1) who had undergone SSTR-directed PET/CT were recruited over 6 mo and evaluated in this retrospective, multicentric study. Parts of this cohort have been described previously (without assessment of interrater reliability) (21). At the time of imaging, all patients had given written informed consent to the medical examination and to the retrospective use of the anonymized data. Any

requirement for additional approval was waived by the local ethics committee because of the retrospective character of the study.

Imaging Procedure

All subjects underwent ⁶⁸Ga-DOTATOC PET/CT at a single center as previously described (21). Imaging was performed in accordance with current guidelines (12,22). Integrated PET/CT using a Biograph 64 (Siemens Medical Solutions) operating in 3-dimensional emission mode with CT attenuation correction was performed on all patients. ⁶⁸Ga-DOTATOC (median, 119 MBq [3.21 mCi]) was injected intravenously. PET emission data were acquired in 3-dimensional mode with a 200 × 200 matrix and 2-min emission time per bed position from the vertex of the skull to the proximal thighs. Consecutively, transmission data were acquired using contrast-enhanced spiral CT (dose modulation with a quality reference of 180 mAs, 120 kV, 512 × 512 matrix, 5-mm slice thickness, increment of 30 mm/s, rotation time of 0.5 s, and pitch index of 1.4). PET data were reconstructed iteratively (3 iterations, 24 subsets, gaussian filtering of 2.0 mm in full width at half maximum) with attenuation correction using dedicated standard software (HD-PET e.soft, Siemens Healthineers).

Imaging Interpretation

PET images were analyzed using XD3 Software (Mirada Medical), InterviewFusion (Mediso Medical Imaging), or syngo.via (version

TABLE 1
Patient Characteristics

Parameter	Characteristic	Data
Age (y)		64 ± 12
Sex	Female	26/51 (51%)
	Male	25/51 (49%)
Indication for SSTR PET/CT	Staging	15/51 (29.4%)
	Restaging	36/51 (70.6%)
Ki-67 (%)		4 ± 13
Grade*	G1	23/48 (47.9%)
	G2	22/48 (45.8%)
	G3	3/48 (6.3%)
Primary	Gastroenteropancreatic NET, in total	41/51 (80.4%)
	Ileum/jejunum/mesenterial	25/41 (61%)
	Pancreas	11/41 (26.8%)
	Rectum	3/41 (7.3%)
	Colon	2/41 (4.9%)
Non-gastroenteropancreatic NET, in total		10/51 (19.6%)
	Lung	5/10 (50%)
	Cancer of unknown primary	5/10 (50%)
Prior therapies	Prior therapies, in total	43/51 (84.3%)
	Surgery	30/43 (69.8%)
	Cold somatostatin analogs	23/43 (53.5%)
	PRRT	15/43 (34.9%)
	Chemotherapy	10/43 (23.3%)
	Locoregional procedures	3/43 (7%)
	Tyrosine kinase inhibitor	1/43 (2.3%)

*Available in 48/51 (94.1%).

Qualitative data are numbers followed by percentages in parentheses; continuous data are mean ± SD.

10B; Siemens Healthcare). Images were interpreted as previously described (18). In brief, PET, CT, and PET/CT images were assessed for all individuals. Two board-certified nuclear medicine physicians with more than 10 y of experience in reading PET/CT (experienced readers 1 and 2) and 2 residents with less than 1 y of experience in reading PET/CT (inexperienced readers 1 and 2), masked to the clinical status of the patients, evaluated all scans independently. Before beginning the independent interpretations, the readers underwent a training session with 5 additional cases to gain familiarity with the workstations. SSTR-RADS-1A lesions are benign and have no abnormal radiotracer uptake, whereas SSTR-RADS-1B lesions involve previous conventional imaging or a histologic diagnosis. As such, SSTR-RADS-1A and -1B were subsumed under SSTR-RADS-1 in the present masked analysis, as described previously (18). No other changes to the SSTR-RADS system were implemented in this study. A complete summary of the SSTR-RADS scoring system (from SSTR-RADS-1 to -5) can be found in a previous publication (16), which also explains the 3-point qualitative assessment scoring for defining the uptake level in an SSTR-avid lesion (level 1: lesion uptake \leq blood pool uptake; level 2: lesion uptake $>$ blood pool uptake but \leq physiologic liver uptake; level 3: lesion uptake $>$ physiologic liver uptake). In addition, the following rules were set to comply with SSTR-RADS. First, a maximum of 5 target lesions (TLs) was selected by the readers. SSTR-RADS suggests that TLs be those that are largest or have the most intense radiotracer uptake, and a maximum of 3 TLs per organ can be included. The following organ compartments were defined: lymph nodes (LNs), skeleton, soft tissue (other than LNs), liver, and lung (23). An SSTR-RADS score had to be assigned to every TL (16). Additionally, all involved organ compartments were identified by the readers, and an overall scan score was assigned (the highest SSTR-RADS score of any of the individual TLs) (16). Various general parameters were assessed by each observer in a binary fashion (18): overall scan result (positive in cases of suggestive radiotracer uptake above the background level), organ involvement, and LN involvement (18). Additionally, the number of organs affected, the number of organ metastases, the number of LN regions, and the number of LNs had to be indicated on a 5-point scale (from 1 to \geq 5 organ metastases, LNs, or number of organs or LN areas affected). The following LN areas were defined: cervical, thoracic or axillary, retroperitoneal, sacral or presacral, and pelvic (23,24). In addition, SSTR density on SSTR PET/CT was assessed on a 3-point scale (none, 0; low, 1; intermediate, 2; or high, 3) as previously described (24). Moreover, all readers had to decide in a binary fashion whether PRRT based on SSTR-RADS should be considered (with SSTR-RADS-4 or -5 TLs guiding the interpreting specialist toward endoradiotherapy) (16).

Statistical Analysis

Continuous data are presented as mean \pm SD. The categorical variables are presented as frequency and percentage. The degrees of agreement were assessed using intraclass correlation coefficients (ICCs) and their 95% CIs based on a mean-rating, single-measure, consistency model. According to Cicchetti, an ICC of less than 0.4 indicates poor interobserver agreement, 0.4–0.59 indicates fair agreement, 0.6–0.74 indicates good agreement, and 0.75–1 indicates excellent agreement (25). A detailed description of the performed statistics can also be found in a previous publication (18). Statistical analysis was performed using MedCalc statistical software (version 18.2.1). The statistical significance level was set at a *P* value of less than 0.05.

RESULTS

General Parameters

Considering the main findings, the ICC was excellent for the overall scan impression (0.76 [95% CI, 0.66–0.84]) and LN involvement (0.82 [95% CI, 0.74–0.88]), whereas for organ involvement,

good interreader agreement was found (0.62 [95% CI, 0.50–0.74]). Moreover, the number of LN metastases, number of LN areas affected, and number of organ metastases showed an excellent concordance rate between readers (ICC, 0.77 [95% CI, 0.68–0.85], 0.78 [95% CI, 0.69–0.86], and 0.85 [95% CI, 0.78–0.90], respectively). For the number of affected organs, a lower interobserver agreement with a good concordance was achieved (ICC, 0.62 [95% CI, 0.50–0.74]).

TL- and Compartment-Based Interobserver Agreement

Among the selected TLs, 153 were chosen by at least 2 individual observers and were assigned to the following compartments: soft tissues (49/153 [32.0%]), LNs (44/153 [28.8%]), liver (42/153 [27.5%]), skeleton (12/153 [7.8%]), and lung (6/153 [3.9%]).

Identical TLs Included by 4 Readers. The identical TLs were included in 58 of 153 (37.9%) instances, with a balanced distribution among the investigated compartments (soft tissue, 19/58 [32.8%]; LNs, 17/58 [29.3%]; liver, 14/58 [24.1%]; skeleton, 6/58 [10.3%]; and lung, 2/58 [3.5%]). The interobserver agreement for SSTR-RADS scoring among identical TLs was excellent, with an ICC of 0.83 (95% CI, 0.76–0.89). On an organ-based compartment level for all 4 readers selecting the same LN lesion, soft-tissue lesion, or liver lesion, the interobserver agreement was also excellent (ICC, 0.91 [95% CI, 0.82–0.96], 0.81 [95% CI, 0.67–0.91], and 0.77 [95% CI, 0.58–0.91], respectively). Table 2 and Figure 1 provide an overview of the distribution of the SSTR-RADS scores for 4 identical TLs.

Identical TLs Included by 3 Readers. In 50 of 153 instances (32.7%), 3 readers identified the identical TLs. Again, the distribution among the indicated compartments was balanced (soft tissue, 18/50 [36%]; liver, 16/50 [32%]; LN, 13/50 [26%]; and lung, 3/50 [6%]). Similar to when the identical TLs were included by 4 readers, an excellent ICC of 0.77 (95% CI, 0.66–0.85) was achieved. However, this result was driven mostly by a high agreement on soft-tissue lesions (ICC, 0.74 [95% CI, 0.53–0.88]), whereas for LN and liver lesions, the agreement rate was poor (ICC, 0.36 [95% CI, 0.02–0.70] and 0.08 [95% CI, -0.18 to 0.44], respectively).

Identical TLs Included by 2 Readers. In 45 of 153 instances (29.4%), 2 readers identified the identical TLs (LN, 14/45 [31.1%]; soft tissue and liver, 12/45 [26.7%]; skeleton, 6/45 [13.3%]; and lung, 1/45 [2.2%]). The ICC was 0.73 (95% CI, 0.56–0.85).

Taken together, good interrater agreement was achieved on a TL-based level, with an ICC of 0.73–0.83, independently of whether identical TLs had been chosen by 2, 3, or 4 observers.

Overall SSTR-RADS Score

An excellent agreement rate of 0.88 (95% CI, 0.82–0.92) was derived, with most scans being rated with an overall SSTR-RADS score of either 4 or 5. Table 2 and Figure 2 give an overview of the distribution of the overall SSTR-RADS score among all readers.

Decision for PRRT

The interobserver agreement rate for considering PRRT on the basis of the assigned SSTR-RADS scores was excellent (ICC, 0.80 [95% CI, 0.72–0.87]). Experienced reader 1 recommended PRRT in 37 of 51 instances (72.5%), experienced reader 2 in 35 of 51 instances (68.7%), inexperienced reader 1 in 34 of 51 instances (66.7%), and inexperienced reader 2 in 40 of 51 instances (78.4%). These findings were further confirmed by a high interobserver rate on SSTR density, with an ICC of 0.80 (95% CI, 0.71–0.87). Figure 3 shows the distribution on recommending PRRT among all 4 readers.

TABLE 2

Distribution of SSTR-RADS Score for 4 Identical TLs ($n = 58$) and for Overall SSTR-RADS Score (Scans, $n = 51$) Among All 4 Readers

Parameter	Reader	SSTR-RADS score							
		1*	2	3A	3B	3C	3D	4	5
4 identical TLs	ER1	3/58 (5.2)	1/58 (1.7)	4/58 (6.9)	1/58 (1.7)	2/58 (3.4)	3/58 (5.2)	13/58 (22.4)	31/58 (53.5)
	ER2	2/58 (3.4)	1/58 (1.7)	5/58 (8.6)		4/58 (6.9)	2/58 (3.4)	2/58 (3.4)	42/58 (72.4)
	IR1	2/58 (3.4)		3/58 (5.2)		4/58 (6.9)	3/58 (5.2)	13/58 (22.4)	33/58 (56.9)
	IR2	2/58 (3.4)	2/58 (3.4)	2/58 (3.4)	1/58 (1.7)	2/58 (3.4)	3/58 (5.2)	4/58 (6.9)	42/58 (72.4)
Overall SSTR-RADS score	ER1	3/51 (5.9)	4/51 (7.8)	1/51 (2)	3/51 (5.9)		2/51 (3.9)	1/51 (2)	37/51 (72.5)
	ER2	3/51 (5.9)	2/51 (3.9)		7/51 (13.7)	1/51 (2)	1/51 (2)	2/51 (3.9)	35/51 (68.6)
	IR1	7/51 (13.7)	1/51 (2)			1/51 (2)	2/51 (3.9)	11/51 (21.5)	29/51 (56.9)
	IR2	5/51 (9.8)	3/51 (5.9)		1/51 (2)		2/51 (3.9)	2/51 (3.9)	38/51 (74.5)

*SSTR-RADS-1A and -1B were subsumed under SSTR-RADS-1, as previously described (18).
 ER = experienced reader; IR = inexperienced reader.
 Values in parentheses are percentages.

Experienced Versus Inexperienced Readers

For overall SSTR-RADS score, the ICC among both experienced readers (0.85 [95% CI, 0.75–0.91]) and inexperienced readers (0.89 [95% CI, 0.82–0.94]) was excellent, with no significant differences between the 2 groups ($P = 0.18$). Similar findings were derived for a comparison of experienced versus inexperienced readers on a TL-based level (identical TLs identified by all 4 readers; ICC of 0.83 [0.73–0.90] for experienced readers vs. 0.89 [95% CI, 0.82–0.94] for inexperienced readers; $P = 0.29$). When comparing the SSTR-RADS–based decision to use PRRT, significance was reached between the 2 groups (ICC of 0.91 [0.85–0.95] for experienced readers vs. 0.73 [95% CI, 0.58–0.84] for inexperienced readers; $P = 0.02$). Figure 4 and Supplemental Figure 1 (supplemental materials are available at <http://jnm.snmjournals.org>) show 2 patients for whom the different levels of experience may have had an impact on SSTR-RADS scoring.

DISCUSSION

The randomized NETTER-1 trial has reported a markedly improved outcome in patients with midgut neuroendocrine tumors treated with PRRT (2), even across a spectrum of different tumor sizes (26). As such, the Food and Drug Administration recently granted approval for the diagnostic agent ⁶⁸Ga-DOTATOC and its therapeutic counterpart ¹⁷⁷Lu-DOTATATE (6,27). Thus, the theranostic concept for NENs is expected to evolve from an orphan treatment restricted to centers specialized in molecular endoradiotherapies mainly outside the United States to a nationwide standardized diagnostic and therapeutic procedure (28). Accurate interpretation of all available scan components (PET, CT, and PET/CT) is of the utmost importance to triage patients for such a theranostic intervention (1). However, given the increasing number of patients screened to determine their eligibility for PRRT, a large variety of pitfalls and normal variants in interpreting SSTR

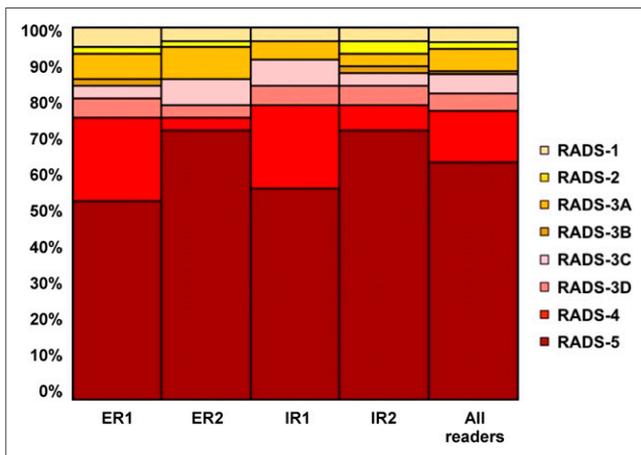


FIGURE 1. TL assessment (identical TL included by all 4 readers). SSTR-RADS-1A and -1B were subsumed under SSTR-RADS-1, as previously described (18). ER = experienced reader; IR = inexperienced reader.

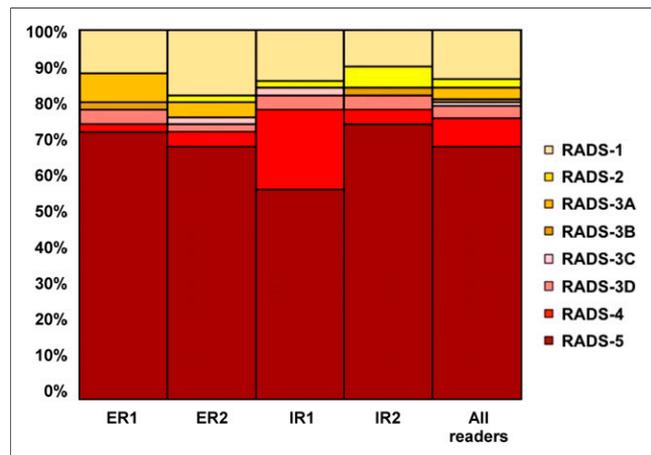


FIGURE 2. Overall SSTR-RADS scoring assessment (for all 4 readers). SSTR-RADS-1A and -1B were subsumed under SSTR-RADS-1, as previously described (18). ER = experienced reader; IR = inexperienced reader.

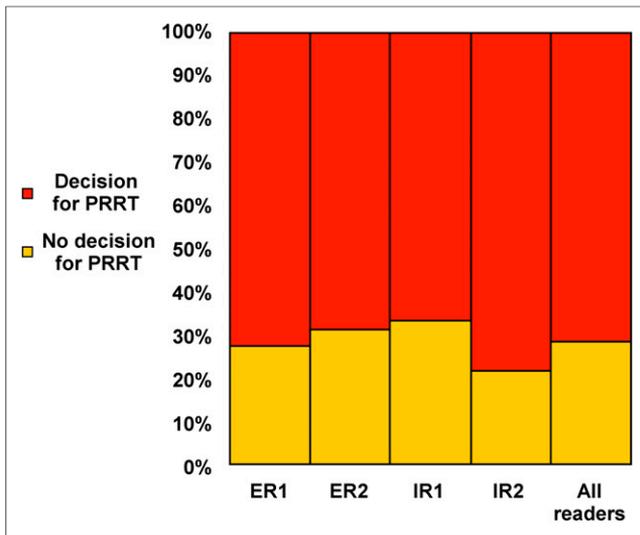


FIGURE 3. Decision for PRRT based on SSTR-RADS among all 4 readers (16). ER = experienced reader; IR = inexperienced reader.

PET/CT have been reported in recent years (10,11,17,29). Such misinterpretations, which may also be caused by a large variability of terminology in written reports, can lead to false-positive recommendations for PRRT or trigger inappropriate work-up (17,24). For instance, in conventional radiology, the level of clinically significant errors has been tabulated to be from 2% to 20% (30), and thus, the American College of Radiology has established numerous RADS system to enable standardized reporting on imaging findings in a large variety of diagnostic settings (13–15). A

RADS reporting system has also been recently introduced for SSTR-directed imaging, and this standardized framework may help to navigate certain pitfalls in scan interpretation; provide the foundation to initiate molecular-imaging-based treatment strategies on a lesion-by-lesion level; and, ultimately, tailor therapeutic approaches to individual patient needs (16). However, before implementation in clinical routine and larger trials, further data confirming high interobserver reproducibility for SSTR-RADS, preferably in a real-world scenario mimicking the workflow of a busy molecular imaging center, are needed.

In the present analysis, the ICC for the overall SSTR-RADS score was excellent (0.88), with a high interobserver agreement rate for both inexperienced and experienced readers (ICC \geq 0.85), thereby suggesting that SSTR-RADS seems to be readily applicable even for less experienced readers. This suggestion is in line with a report of Fendler et al. describing an excellent agreement rate when evaluating the overall scan result of SSTR PET/CT in a binary fashion by less trained observers ($\kappa = 0.80$) (24). In neuroendocrine liver lesions, contrast-enhanced CT revealed a substantial concordance rate on a visual assessment among junior versus senior abdominal radiologists only when using a nonstandardized approach ($\kappa = 0.62$) (31). Of note, SSTR PET/CT enables a noninvasive whole-body readout and, thus, may be associated with a higher degree of complexity as it is not restricted to a single organ but allows for investigation of every putative site of disease (32). Given the markedly higher agreement rates achieved in the present analysis, such differences in interobserver concordance rates may emphasize the need for standardization in interpreting imaging findings in complex tumor entities, such as those of neuroendocrine origin (33).

On an overall-scan-impression level, most PET studies were assigned a SSTR-RADS-4 or -5 score by all readers (Table 2), and thus, in a manner similar to PSMA-RADS for prostate-specific membrane antigen PET/CT (18), one may speculate that this observation derives from the high accuracy of SSTR-directed radiotracers (29,34). Apart from overall scan results assessed with RADS, the concordance rate was also good on a TL-based level (ICC \geq 0.73), independently of whether identical TLs had been chosen by 2, 3, or 4 observers. Of note, when applying SSTR-RADS to liver lesions, the interobserver rate exhibited a large ICC range from 0.08 (in the case of 3 readers) to 0.77 (in the case of 4 readers) for identical liver lesions identified. In comparison to MRI, the diagnostic superiority of SSTR PET/CT for assessing carcinoid liver lesions is still a matter of debate, with studies showing a lower sensitivity for SSTR-directed imaging (74% vs. MRI, 88%) (35). Thus, SSTR PET/MRI appears to be more sensitive than PET/CT to assess hepatic metastases in NEN patients, thereby suggesting that future studies should also validate SSTR-RADS on such novel hybrid devices.

Optimized risk stratification to select individuals who will most likely benefit from treatment is of the utmost importance

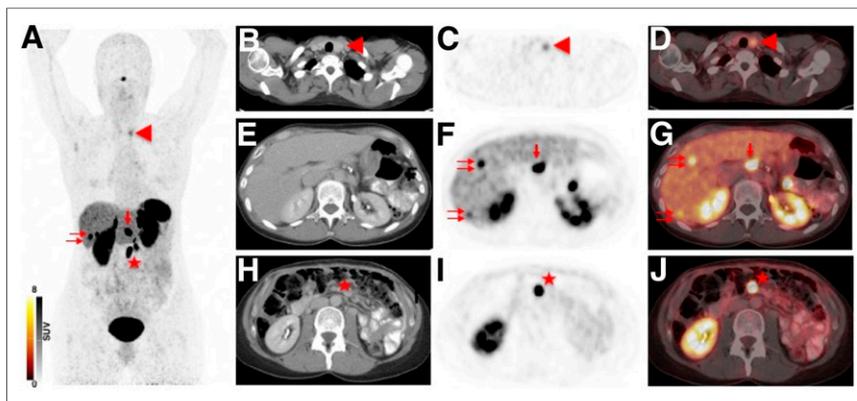


FIGURE 4. Example of different SSTR-RADS classifications of experienced readers vs. inexperienced readers. Images are from 49-y-old woman undergoing restaging for pancreatic neuroendocrine neoplasia. (A) ^{68}Ga -DOTATOC maximum-intensity projection demonstrated intense radiotracer uptake in thyroid (arrowhead), liver (double arrow), and celiac (single arrow) and mesenteric (star) LNs. (B–D) Axial CT (B), ^{68}Ga -DOTATOC PET (C), and ^{68}Ga -DOTATOC PET/CT (D) showed intense SSTR expression in left lobe of thyroid (arrowhead), which was rated as SSTR-RADS-3C by experienced readers and as SSTR-RADS-1 by inexperienced readers. SSTR-RADS-3C describes lesions with intense radiotracer accumulation in sites highly atypical of NENs, and thus, these lesions are suggestive of SSTR-expressing, non-NEN benign tumor. Fine-needle aspiration biopsy of nodule revealed no abnormality. (E–G) Axial CT (E), ^{68}Ga -DOTATOC PET (F), and ^{68}Ga -DOTATOC PET/CT (G) demonstrate intense uptake in liver (double arrows) and celiac LN (single arrow), with no corresponding finding on CT. These lesions were rated as SSTR-RADS-4 by all readers. (H–J) Axial CT (H), ^{68}Ga -DOTATOC PET (I), and ^{68}Ga -DOTATOC PET/CT (J) show intense uptake in mesenteric LN (star). Given the corresponding finding on CT, this lesion was classified SSTR-RADS-5 by all observers. All readers recommended PRRT. ER = experienced readers; IR = inexperienced readers.

(1). To date, current established strategies for selecting appropriate candidates are assessing the SSTR density on pretherapeutic SSTR-directed SPECT or PET, mainly using the Krenning score. Although a reliable imaging metric, its predictive efficacy is rather limited, as only 60% of the subjects with substantial radiotracer accumulation (Krenning score of 4) will show a minor or complete remission (1,36). As a possible explanation, the Krenning score, which had initially been developed for a ^{111}In -pentetreotide scan, considers only liver or spleen uptake versus tumor uptake on functional imaging, unlike a standardized framework system such as SSTR-RADS, which takes into account all available imaging components (PET, CT, and hybrid imaging) and, thus, may provide a more elaborate assessment of the entire tumor burden (16). For instance, the Krenning score may miss certain lesions, such as SSTR-RADS-negative lesions, which are visible exclusively on the CT component (SSTR-RADS-3D). However, if substantial SSTR expression can still be visualized in all other lesions, the patient may benefit more from a combined treatment approach of PRRT together with a locoregional procedure, such as selective internal radiation therapy (37). Second, if there would be an increasing number of dedifferentiated lesions with most of the remaining metastases still being SSTR-positive, a combination of chemotherapy and PRRT would also be feasible (38). Nonetheless, outcome studies as conducted with the Krenning score are still lacking for SSTR-RADS, and future studies must show the efficacy of structured reporting in response prediction. However, as a preliminary step, readers with varying levels of experience should agree on the appropriateness of choosing PRRT based on imaging. Recent reports showed that such recommendations for or against a theranostic intervention significantly vary among multiple observers ($\kappa = 0.64$) when no standardized framework for reporting is applied, whereas in the present study an excellent interobserver rate was seen (ICC, 0.80) (24). Although a significant difference between inexperienced and experienced readers for deciding on PRRT was observed, less experienced readers still achieve a higher ICC (0.58–0.84) when using SSTR-RADS than when using a nonstandardized approach ($\kappa = 0.41$ – 0.68), thereby suggesting that SSTR-RADS may be a useful tool in recommending theranostic interventions. SSTR intensity on SSTR-directed diagnostic procedures is still considered the gold standard in selecting PRRT candidates, but the intrinsic heterogeneity of radiotracer accumulation in putative sites of disease has led to the development of additional molecular tools for outcome prediction (1,39). For instance, patient-specific multigene genomic signature testing is currently penetrating the clinical arena and has already demonstrated a high accuracy in predicting PRRT response (40). However, combining such innovative biomarkers with SSTR-RADS applied to routinely conducted PET/CT may further reduce the number of patients with a suboptimal outcome or radioresistance.

This study had several limitations. First, histopathologic comparisons validating each TL were not feasible. Second, none of the readers knew the clinical status; unrestricted access to clinical data may further influence the herein observed ICCs and impact the concordance rate for recommending PRRT. Nonetheless, we aimed to investigate the robustness of SSTR-RADS as an imaging-finding-driven construct (18), and future studies should definitely assess the impact on agreement rates when the readers are provided with clinical information such as Ki-67, grade, or previous therapies. A substantial proportion of scans rated as SSTR-RADS-4 or -5 were not considered eligible for PRRT, and thus, a more sophisticated approach of applying SSTR-RADS for treatment recommendation should be implemented in future versions, such as by taking the entire tumor burden into account

(particularly in cases with heterogeneous receptor expression). Nonetheless, current inclusion and exclusion criteria should still apply if patients are considered for PRRT (41,42). Moreover, interobserver reliability could have been further increased by providing stricter advice on selecting TLs instead of randomly selecting lesions, such as by providing a detailed list of every organ or LN of interest (written guide to imaging interpretation) (24). In this regard, SSTR-RADS 1.1 should also provide a more suitable characterization of TLs, which in turn will ensure that readers from different study sites will choose the same lesions. Moreover, SSTR-RADS is based on imaging findings, and thus, its informative value is also limited by technical aspects, such as system resolutions or partial-volume effects (21). Further studies could also evaluate the performance of inexperienced readers versus a reference standard established by a consensus interpretation by several experienced readers and should preferably include a higher number of readers.

CONCLUSION

In the present analysis validating the structured reporting system SSTR-RADS for SSTR PET/CT, a high concordance rate, even among readers with different levels of experience, was demonstrated. As such, SSTR-RADS is nearing readiness to be implemented in larger trials or clinical routine.

DISCLOSURE

This work was funded by the German Research Foundation (DFG), through the PRACTIS – Clinician Scientist Program of Hannover Medical School (ME 3696/3-1, RAW). No other potential conflict of interest relevant to this article was reported.

KEY POINTS

QUESTION: Will SSTR-RADS as a standardized reporting system for interpreting SSTR PET/CT achieve a high interobserver agreement rate, even for less experienced readers?

PERTINENT FINDINGS: In a cohort study evaluating 51 SSTR-targeted PET/CT scans, 4 readers, unaware of the clinical status, achieved a high interobserver agreement rate on a target-lesion and overall-scan-impression level, independently of level of previous experience in reading SSTR PET/CT. When applying SSTR-RADS, readers with varying levels of experience also showed high concordance on the appropriateness of choosing PRRT based on imaging results, thereby suggesting that standardized reporting may be a useful tool in recommending theranostic interventions.

IMPLICATIONS FOR PATIENT CARE: Given the high interobserver agreement rates, SSTR-RADS can be implemented in collecting data for large prospective trials or in clinical routine, which in turn may minimize the risk of communication errors between molecular imaging experts and referring oncologists.

REFERENCES

1. Bodei L, Herrmann K, Baum RP, Kidd M, Malczewska A, Modlin IM. Caveat emptor: let our acclaim of the apotheosis of PRRT not blind us to the error of Prometheus. *J Nucl Med.* 2019;60:7–8.
2. Strosberg J, El-Haddad G, Wolin E, et al. Phase 3 trial of ^{177}Lu -Dotatate for midgut neuroendocrine tumors. *N Engl J Med.* 2017;376:125–135.
3. Kunikowska J, Pawlak D, Bak MI, Kos-Kudla B, Mikolajczak R, Krollicki L. Long-term results and tolerability of tandem peptide receptor radionuclide therapy with $^{90}\text{Y}/^{177}\text{Lu}$ -DOTATATE in neuroendocrine tumors with respect to the primary location: a 10-year study. *Ann Nucl Med.* 2017;31:347–356.

4. Lütje S, Feldmann G, Essler M, Brossart P, Bundschuh RA. Immune checkpoint imaging in oncology: a game changer toward personalized immunotherapy? *J Nucl Med.* 2020;61:1137–1144.
5. Baum RP, Kluge AW, Kulkarni H, et al. [¹⁷⁷Lu-DOTA]⁰-D-Phe¹-Tyr³-octreotide (¹⁷⁷Lu-DOTATOC) for peptide receptor radiotherapy in patients with advanced neuroendocrine tumours: a phase-II study. *Theranostics.* 2016;6:501–510.
6. FDA approves lutetium Lu 177 dotatate for treatment of GEP-NETS. U.S. Food and Drug Administration website. <https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-lutetium-lu-177-dotatate-treatment-gep-nets>. Published January 26, 2018. Accessed November 23, 2020.
7. Sowa-Staszczak A, Pach D, Chrzan R, et al. Peptide receptor radionuclide therapy as a potential tool for neoadjuvant therapy in patients with inoperable neuroendocrine tumours (NETs). *Eur J Nucl Med Mol Imaging.* 2011;38:1669–1674.
8. Prasad V, Brenner W, Modlin IM. How smart is peptide receptor radionuclide therapy of neuroendocrine tumors especially in the salvage setting? The clinician's perspective. *Eur J Nucl Med Mol Imaging.* 2014;41:202–204.
9. Cho SY. Proposed criteria positions PSMA PET for the future. *J Nucl Med.* 2018;59:466–468.
10. Werner RA, Bundschuh RA, Bundschuh L, et al. Molecular imaging reporting and data systems (MI-RADS): a generalizable framework for targeted radiotracers with theranostic implications. *Ann Nucl Med.* 2018;32:512–522.
11. Hofman MS, Lau WF, Hicks RJ. Somatostatin receptor imaging with ⁶⁸Ga DOTATATE PET/CT: clinical utility, normal patterns, pearls, and pitfalls in interpretation. *Radiographics.* 2015;35:500–516.
12. Hope TA, Bergsland EK, Bozkurt MF, et al. Appropriate use criteria for somatostatin receptor PET imaging in neuroendocrine tumors. *J Nucl Med.* 2018;59:66–74.
13. Liver Reporting & Data System (LI-RADS). American College of Radiology website. <https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/LI-RADS>. Accessed November 23, 2020.
14. Thyroid Imaging Reporting & Data System (TI-RADS™). American College of Radiology website. <https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/TI-RADS>. Accessed November 23, 2020.
15. ACR BI-RADS Atlas® 5th Edition. American College of Radiology website. <https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Bi-Rads>. Accessed November 23, 2020.
16. Werner RA, Solnes LB, Javadi MS, 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.
17. Werner RA, Bundschuh RA, Bundschuh L, et al. Novel structured reporting systems for theranostic radiotracers. *J Nucl Med.* 2019;60:577–584.
18. Werner RA, Bundschuh RA, Bundschuh L, et al. Interobserver agreement for the standardized reporting system PSMA-RADS 1.0 on ¹⁸F-DCFPyL PET/CT imaging. *J Nucl Med.* 2018;59:1857–1864.
19. Derwael C, Lavergne O, Lovinfosse P, et al. Interobserver agreement of [⁶⁸Ga] Ga-PSMA-11 PET/CT images interpretation in men with newly diagnosed prostate cancer. *EJNMMI Res.* 2020;10:15.
20. Toriihara A, Nobashi T, Baratto L, et al. Comparison of three interpretation criteria of ⁶⁸Ga-PSMA11 PET based on inter- and intra-reader agreement. *J Nucl Med.* 2020;61:533–539.
21. Werner RA, Hanscheid H, Leal JP, et al. Impact of tumor burden on quantitative [⁶⁸Ga] DOTATOC biodistribution. *Mol Imaging Biol.* 2019;21:790–798.
22. Virgolini I, Ambrosini V, Bomanji JB, et al. Procedure guidelines for PET/CT tumour imaging with ⁶⁸Ga-DOTA-conjugated peptides: ⁶⁸Ga-DOTA-TOC, ⁶⁸Ga-DOTA-NOC, ⁶⁸Ga-DOTA-TATE. *Eur J Nucl Med Mol Imaging.* 2010;37:2004–2010.
23. Fendler WP, Calais J, Allen-Auerbach M, et al. ⁶⁸Ga-PSMA-11 PET/CT interobserver agreement for prostate cancer assessments: an international multicenter prospective study. *J Nucl Med.* 2017;58:1617–1623.
24. Fendler WP, Barrio M, Spick C, et al. ⁶⁸Ga-DOTATATE PET/CT interobserver agreement for neuroendocrine tumor assessment: results of a prospective study on 50 patients. *J Nucl Med.* 2017;58:307–311.
25. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess.* 1994;6:284–290.
26. Strosberg J, Kunz PL, Hendifar A, et al. Impact of liver tumour burden, alkaline phosphatase elevation, and target lesion size on treatment outcomes with ¹⁷⁷Lu-Dotatate: an analysis of the NETTER-1 study. *Eur J Nucl Med Mol Imaging.* 2020;47:2372–2382.
27. FDA approves new diagnostic imaging agent to detect rare neuroendocrine tumors. U.S. Food and Drug Administration website. <https://www.fda.gov/news-events/press-announcements/fda-approves-new-diagnostic-imaging-agent-detect-rare-neuroendocrine-tumors>. Published June 1, 2016. Accessed November 23, 2020.
28. Kendi AT, Mailman J, Naraev BG, Underwood JK, Halfdanarson TR. Patient travel concerns after treatment with ¹⁷⁷Lu-DOTATATE. *J Nucl Med.* 2020;61:496–497.
29. Wild D, Bomanji JB, Benkert P, et al. Comparison of ⁶⁸Ga-DOTANOC and ⁶⁸Ga-DOTATATE PET/CT within patients with gastroenteropancreatic neuroendocrine tumors. *J Nucl Med.* 2013;54:364–372.
30. Goddard P, Leslie A, Jones A, Wakeley C, Kabala J. Error in radiology. *Br J Radiol.* 2001;74:949–951.
31. Zappa M, Hentic O, Vullierme MP, et al. Is visual radiological evaluation of liver tumour burden in patients with neuroendocrine tumours reproducible? *Endocr Connect.* 2017;6:33–38.
32. Antwi K, Nicolas G, Wild D, Christ E. Molecular imaging for neuroendocrine tumours. *Swiss Med Wkly.* 2019;149:w20017.
33. Ezziddin S, Adler L, Sabet A, et al. Prognostic stratification of metastatic gastroenteropancreatic neuroendocrine neoplasms by ¹⁸F-FDG PET: feasibility of a metabolic grading system. *J Nucl Med.* 2014;55:1260–1266.
34. Ilhan H, Todica A, Lindner S, et al. First-in-human ¹⁸F-SiFAlin-TATE PET/CT for NET imaging and theranostics. *Eur J Nucl Med Mol Imaging.* 2019;46:2400–2401.
35. Schreiter NF, Nogami M, Steffen I, et al. Evaluation of the potential of PET-MRI fusion for detection of liver metastases in patients with neuroendocrine tumours. *Eur Radiol.* 2012;22:458–467.
36. Kwekkeboom DJ, Kam BL, van Essen M, et al. Somatostatin-receptor-based imaging and therapy of gastroenteropancreatic neuroendocrine tumors. *Endocr Relat Cancer.* 2010;17:R53–R73.
37. Maccauro M, Follacchio GA, Spreafico C, Coppa J, Seregni E. Safety and efficacy of combined peptide receptor radionuclide therapy and liver selective internal radiation therapy in a patient with metastatic neuroendocrine tumor. *Clin Nucl Med.* 2019;44:e286–e288.
38. Yordanova A, Ahrens H, Feldmann G, et al. Peptide receptor radionuclide therapy combined with chemotherapy in patients with neuroendocrine tumors. *Clin Nucl Med.* 2019;44:e329–e335.
39. Graf J, Pape UF, Jann H, et al. Prognostic significance of somatostatin receptor heterogeneity in progressive neuroendocrine tumor treated with Lu-177 DOTATOC or Lu-177 DOTATATE. *Eur J Nucl Med Mol Imaging.* 2020;47:881–894.
40. Bodei L, Kidd MS, Singh A, et al. PRRT genomic signature in blood for prediction of ¹⁷⁷Lu-octreotate efficacy. *Eur J Nucl Med Mol Imaging.* 2018;45:1155–1169.
41. Bodei L, Mueller-Brand J, Baum RP, et al. The joint IAEA, EANM, and SNMMI practical guidance on peptide receptor radionuclide therapy (PRRT) in neuroendocrine tumours. *Eur J Nucl Med Mol Imaging.* 2013;40:800–816.
42. Hicks RJ, Kwekkeboom DJ, Krenning E, et al. ENETS consensus guidelines for the standards of care in neuroendocrine neoplasia: peptide receptor radionuclide therapy with radiolabeled somatostatin analogues. *Neuroendocrinology.* 2017;105:295–309.