68Ga-DOTATATE PET/CT Interobserver Agreement for Neuroendocrine Tumor Assessment: Results of a Prospective Study on 50 Patients

Wolfgang Peter Fendler[1], Martin Barrio[1], Claudio Spick[1], Martin Allen-Auerbach[1], Valentina Ambrosini[2], Matthias Benz[3], Christina Bluemel[4], Ravinder Kaur Grewal[5], Constantin Lapa[4], Matthias Miederer[6], Guillaume Nicolas[3], Tibor Schuster[7], Johannes Czernin[1], and Ken Herrmann[1]

[1]Department of Molecular and Medical Pharmacology, David Geffen School of Medicine, UCLA, Los Angeles, California; [2]Nuclear Medicine, Department of Experimental Diagnostic and Specialized Medicine, University of Bologna and S. Orsola-Malpighi Hospital, Bologna, Italy; [3]Clinic of Radiology and Nuclear Medicine, University of Basel Hospital, Basel, Switzerland; [4]Department of Nuclear Medicine, Julius-Maximilians-University, Würzburg, Germany; [5]Molecular Imaging and Therapy Service, Department of Radiology, Memorial Sloan Kettering Cancer Center, New York, New York; [6]Department of Nuclear Medicine, University Medical Center Mainz, Mainz, Germany; and [7]Department of Family Medicine, McGill University, Montreal, Quebec, Canada

O verexpression of cell surface somatostatin receptors (SSRs) in well-differentiated neuroendocrine tumors (NETs) can be exploited for imaging and therapy with radiolabeled somatostatin analogs. SSR scintigraphy using 111In-octreotide (OctreoScan; Mallinckrodt Pharmaceuticals) has been available for more than 20 y. Over the past few decades, scintigraphy was gradually replaced by 68Ga-DOTATOC or 68Ga-DOTATATE PET/CT imaging because of its superior accuracy for NET staging (1–4). PET/CT using 68Ga-labeled somatostatin analogs is now considered the gold standard, and its superiority over scintigraphy is also emphasized in the European Neuroendocrine Tumor Society guidelines (5).

68Ga-DOTATATE PET/CT demonstrates high accuracy for NET staging and is an important companion diagnostic to the highly effective 177Lu-DOTATATE peptide receptor radionuclide therapy (PRRT) (6). High remission rates after PRRT correlated positively with intense 111In-octreotide uptake on pretherapy SSR imaging using a liver-based 4-point scale for tracer accumulation (7). 68Ga-DOTATATE uptake might thus, similarly to the Krenning scale, predict the likelihood of response to PRRT. 68Ga-DOTATATE has now received approval by the U.S. Food and Drug Administration for NET imaging. However, little is known about interobserver differences in 68Ga-DOTATATE PET/CT interpretation. The overall value of an imaging method is associated with the degree of observer agreement. Knowledge of interobserver variability and reproducibility is therefore essential for interpreting study results and designing future trials. The aim of this study was to determine interobserver agreement for interpretations of 68Ga-DOTATATE PET/CT images and to compare findings between observers with low and high levels of experience.

MATERIALS AND METHODS

Patients and Image Acquisition

From June 2013 until March 2014, 50 patients with known or suspected NET of the small bowel ($n = 19$), pancreas ($n = 14$), lung ($n = 4$), or other location ($n = 13$) were prospectively recruited under a Food and Drug Administration–approved Investigational New Drug application. The patients had been referred for NET staging ($n = 10$) or restaging ($n = 40$). All had been included in a previously published study on the impact of ^{68}Ga-DOTATATE PET/CT on patient management (8). The prospective study was approved by the University of California, Los Angeles, Institutional Review Board, and all subjects gave written informed consent. The patients were prepared and the images acquired as previously described (8). In brief, ^{68}Ga-DOTATATE was injected intravenously at a dose of 190 ± 17 MBq (5.1 ± 0.5 mCi; range, 130–211 MBq [3.5–5.7 mCi]). A tracer uptake period of 60 min (mean, 62 ± 7 min) was allowed before imaging, which was performed using a Biograph True Point 64 or Biograph mCT device (Siemens). For anatomic correlation and identification of organ lesions, intravenous contrast medium (90–115 mL of Omnipaque 350 [GE Healthcare] at a flow rate of 2 mL/s) was administered to 47 of the 50 patients (94%). Oral contrast medium (~600 mL of barium sulfate [Readi-Cat 2; Bracco]) was given to all patients within 1 h before the scan.

Observers

Anonymized PET/CT images (one per patient) were electronically submitted to 7 nuclear medicine physicians from 5 centers in Europe ($n = 4$) or North America ($n = 1$). The data included standard DICOM files of CT, attenuation-corrected PET, and uncorrected PET images. All physicians had at least 5 y of experience with interpreting oncologic ^{18}F-FDG PET/CT images. All centers had been performing ^{68}Ga-DOTATATE PET/CT on a regular basis for at least 2 y, and all physicians had experience with ^{68}Ga-DOTATATE PET/CT. The observers reported their number of years of experience and the number of ^{68}Ga-DOTATATE PET/CT scans that they had interpreted. On the basis of these reports, the observers were grouped into two experience categories: low level ($<$500 scans or $<$5 y of experience with ^{68}Ga-DOTATATE PET/CT; $n = 4$) and high level ($\geq$500 scans or $\geq$5 y of experience with ^{68}Ga-DOTATATE PET/CT; $n = 3$).

Visual Interpretation

Each observer received a written guide to image interpretation and two patient examples, which are shown in the supplemental materials (available at http://jnm.snmjournals.org), and reported the results in a table template. The following patient information was disclosed to each observer before image interpretation: indication (staging or restaging), sex (male or female), age (y), height (cm), weight (kg), injected dose (mCi), uptake time (min), CT protocol (contrast-enhanced or nonenhanced), prior therapy (yes or no), prior chemotherapy (yes or no), prior PRRT (yes or no), prior surgery (yes or no), and prior octreotide therapy (yes or no). The observers were masked to all other clinical data.

The images were visually interpreted for the following: overall scan result for presence or absence of disease, SSR density in NET tissue (none, 0; low, 1; intermediate, 2; or high, 3), indication for PRRT (yes or no), organs affected (yes or no), number of organs affected (0, 1, 2, 3, 4, or $\geq$5), number of organ metastases detected (0, 1, 2, 3, 4, or $\geq$5), lymph nodes affected (yes or no), number of lymph node regions affected (0, 1, 2, 3, 4, or $\geq$5), and number of lymph node metastases detected (0, 1, 2, 3, 4, or $\geq$5). The criteria for a PRRT indication were, among others, intense tracer uptake by tumor lesions and metastatic spread.

SUV Measurement

Each observer recorded series number, image number, location, and SUV_{max} for up to 3 target lesions. The lesions were chosen from each diseased organ system following predefined criteria as outlined in the written guide to image interpretation.

Each observer measured SUV_{max} and SUV_{mean} using a 5-cm-diameter circular region of interest in a lesion-free location in the right hepatic lobe and a 2-cm-diameter circular region of interest in lesion-free splenic parenchyma. To ascertain consistency, the observers were provided with a PET/CT dataset and SUV data for one test patient for comparison.

Reference Standard and Statistical Analyses

Data acquisition and analyses were performed prospectively. Image interpretations by a University of California, Los Angeles, physician who had access to all baseline and follow-up clinical information served as the reference standard. For binary data, agreement between each observer and the reference observer was evaluated using the Cohen κ statistic (9). Overall agreement using pooled observer data was evaluated using generalized estimation equations (10). For nonbinary data, agreement among observers was evaluated by the intraclass correlation coefficient (ICC) using a 2-way mixed model for absolute agreement (single measures) (11). To calculate ICC for tumor SUV_{max}, one target lesion, that is, the lesion reported by most observers, was chosen per patient. Ninety-five percent confidence intervals (95%CIs) are reported for κ and ICC values. Interpretation of κ and ICC was based on a classification provided by Landis and Koch (12): 0.0, poor; 0.0–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; and 0.81–1.00, almost perfect reproducibility.

Discrepancies in quantitative ratings among observers were expressed as mean difference (Δ) $\pm$ SD. Statistical analyses were performed using R software (R Core Team 2015; R Foundation for Statistical Computing) with the package "irr" (version 0.84) for generalized estimation equation modeling and SPSS (version 15.0; SPSS Inc.) for all other statistical analyses.

RESULTS

Patient Characteristics

Table 1 summarizes the patient characteristics. The reference observer rated the ^{68}Ga-DOTATATE PET/CT studies of 37 (74%) of the 50 patients as positive for NET: 21 (42%) as stage N1 and 34 (68%) as stage M1.

Visual Interpretation

Interobserver agreement on the visual interpretation is shown in Table 2. Reproducibility was substantial to almost perfect for the overall scan result, organ involvement, lymph node involvement, and the respective subitems (i.e., number of organs, organ metastases, lymph node areas, lymph node metastases [each ICC or κ $\geq$ 0.70]). The mean absolute difference from the reference observer was low for number of organs, lymph node areas, and metastases (each $\Delta < 0.5$), and there was no relevant difference between observers with a low level of experience and those with a high level. However, interobserver agreement on whether PRRT was indicated ranged from only moderate to substantial (κ = 0.64; 95%CI, 0.57–0.70).

False-positive and false-negative findings, as well as level of agreement between individual observers and the reference observer, are listed in Table 3. For the overall scan result, organ involvement, and lymph node involvement, observers with either level of experience demonstrated a low frequency of false-positive and false-negative findings (range, 0–6).

There was a false-positive overall scan result for 5 of the 7 observers; an example is shown in Figure 1. Three of the 4 observers with a low level of experience had fair agreement with the reference observer on whether PRRT was indicated. The number of erroneous recommendations for PRRT was higher for observers with low experience (range, 7–12) than for those with high experience (range, 4–8).

TABLE 1
Characteristics of the 50 Patients

Characteristic	Data
Age (y)	
Mean	61
Range	27–78
Sex (n)	
Female	25 (50%)
Male	25 (50%)
Location of primary (n)	
Small bowel	19 (38%)
Pancreas	14 (28%)
Other	13 (26%)
Lung	4 (8%)
Indication (n)	
Staging	10 (20%)
Restaging	40 (80%)
Determined by reference observer:	
Scan positive for NET (n)	37 (74%)
N1 stage (n)	21 (42%)
M1 stage (n)	34 (68%)
Prior therapy (n)	
Chemotherapy	11 (22%)
PRRT	3 (6%)
Surgery	30 (60%)
Octreotide	22 (44%)

Details on individual test performance, as well as sensitivity and specificity values, are available in Supplemental Table 1.

SUV Measurements

Interobserver agreement on SUV measurements is given in Table 4. Agreement was almost perfect for tumor SUV_{max} (ICC, 0.99). Liver SUV_{max} and spleen SUV_{mean} were highly reproducible (ICC, 0.79 and 0.81, respectively), with a low mean absolute difference ($\Delta < 1.2$) when compared with the SUV measurements of the reference observer. The mean absolute difference was comparable between observers with low experience and those with high experience. Figure 2 illustrates agreement on individual SUV measurements.

DISCUSSION

^{68}Ga-DOTATATE PET/CT image interpretation is not without pitfalls. Up to 70% of high-grade NET lesions are ^{68}Ga-DOTATATE PET–negative because of low, or even absence of, SSR expression (13). On the other hand, inflammation with recruitment of SSR-expressing macrophages may lead to false-positive findings (14). Physiologic uptake in the adrenal glands, the pituitary gland, and the uncinate process of the pancreas potentially mask NET lesions or can be misinterpreted as tumor tissue (15). Other pitfalls include uptake at sites of osteoblastic activity, splenules, splenosis, and SSR expression in hemangioma and other benign or malignant tumors of nonneuroendocrine origin (16,17). All these processes may result in interpretative errors that ultimately limit the accuracy of ^{68}Ga-DOTATATE PET/CT. To reduce the frequency of errors, most study protocols include a consensus image analysis by multiple observers. However, the level of interobserver agreement cannot be discerned from consensus readings (8,13,18–20). Here, we provide evidence of substantial to almost perfect agreement among 7 observers with varying levels of experience. Agreement was determined for both visual and semiquantitative analyses in multiple, predefined, categories. Previous reports included reproducibility as secondary or tertiary endpoints and compared findings from no more than 3 observers, who had comparable or unknown levels of experience. Deppen et al. reported almost perfect reproducibility ($\kappa = 0.82$) between two masked observers and one nonmasked observer for interpretation of the ^{68}Ga-DOTATATE PET/CT images of 78 patients with pulmonary or gastroenteropancreatic NET (3). Ruf et al. analyzed agreement between two observers after they separately analyzed the ^{68}Ga-DOTATOC PET images and triple-phase CT images of 51 NET patients. Agreement was substantial for PET ($\kappa = 0.77$) but only fair to moderate for triple-phase and

TABLE 2
Interobserver Agreement on Visual Image Interpretation

Parameter	Agreement	95%CI	Mean Δ ± SD	
			Low experience	High experience
Overall scan result (pos/neg)	κ: 0.80	0.74–0.86		
PRRT indication (pos/neg)	κ: 0.64	0.57–0.70		
SSR density (0–3)	ICC: 0.84	0.77–0.89	0.30 ± 0.64	0.31 ± 0.65
Organ involvement (pos/neg)	κ: 0.70	0.64–0.76		
Number of organs (0–4, ≥5)	ICC: 0.75	0.66–0.83	0.42 ± 0.67	0.39 ± 0.66
Number of organ met. (0–4, ≥5)	ICC: 0.84	0.77–0.89	0.42 ± 0.95	0.47 ± 1.09
LN involvement	κ: 0.71	0.65–0.78		
Number of LN areas (0–4, ≥5)	ICC: 0.79	0.71–0.86	0.23 ± 0.46	0.28 ± 0.63
Number of LN met. (0–4, ≥5)	ICC: 0.77	0.69–0.84	0.42 ± 0.90	0.47 ± 1.00

Pos = positive; neg = negative; met = metastases; LN = lymph node.

TABLE 3
False Findings for Each Observer and Agreement with Reference Observer

Observer	Experience	Scan result			PRRT indication			Organ involvement			Lymph node involvement		
		FP	FN	κ	FP	FN	κ	FP	FN	κ	FP	FN	κ
1	Low	1	2	0.85	7	1	0.68	2	0	0.91	6	2	0.68
2	Low	3	1	0.78	11	4	0.41	3	3	0.72	3	3	0.75
3	Low	2	1	0.84	12	1	0.50	4	1	0.76	1	2	0.88
4	Low	1	0	0.95	10	1	0.57	2	1	0.86	2	2	0.84
5	High	0	0	1.00	4	2	0.76	4	0	0.80	6	1	0.72
6	High	2	3	0.75	8	1	0.65	4	5	0.59	2	2	0.84
7	High	1	5	0.72	4	2	0.76	2	3	0.77	4	3	0.72

FP = false-positive; FN = false-negative.

single-phase CT (*21*). Two trials used separate analyses by independent observers but did not report reproducibility (*22,23*).

Interobserver agreement is an important aspect of clinical applicability. Complex protocols can be associated with reduced interobserver agreement on NET staging, as has been demonstrated for triple-phase CT (*21*) and single- or multisequence MRI (*24,25*). Here, we have shown that both visual and semiquantitative ^{68}Ga-DOTATATE PET/CT interpretation is highly reproducible among observers with both low and high levels of experience. Our findings indicate that ^{68}Ga-DOTATATE PET/CT has a high interobserver reliability for NET staging in a clinical or research setting, even if the images are interpreted by less experienced observers. Interestingly, observers with lower experience levels inappropriately recommended PRRT with a higher frequency than did highly experienced observers. ^{68}Ga-DOTATATE PET/CT images should thus be interpreted by experienced observers if PRRT is being considered in NET patients.

The study had several limitations. First, the observers had access to only the ^{68}Ga-DOTATATE PET/CT images and a limited set of patient information, whereas the reference observer had access to all available clinical and image data. Clinical data such as type and duration of prior therapy are important in determining the appropriateness for PRRT. Lack of clinical data created a disadvantage that might have led to the higher frequency with which the observers inappropriately found that PRRT was indicated. Second, the observers were grouped on the basis of their experience with ^{68}Ga-DOTATATE PET/CT. However, the skills of an observer are determined by multiple additional factors, including clinical knowledge and experience with other imaging modalities. Thus, unknown factors might have led to an underestimation of the true expertise, especially for observers in the low-experience group.
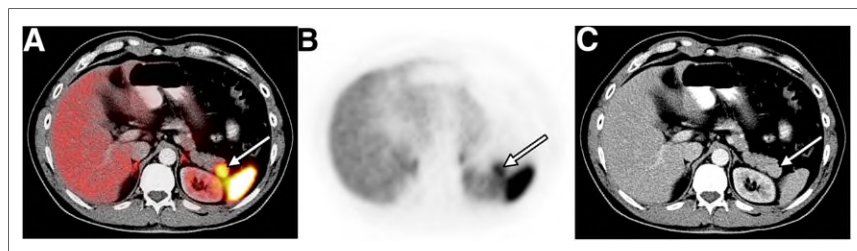

FIGURE 1. False-positive overall scan result in patient with repeated flushing. Axial ^{68}Ga-DOTATATE PET/CT (A), PET (B), and CT (C) images are shown. Focal ^{68}Ga-DOTATATE uptake (arrow) in region of pancreatic tail was judged to represent NET by 5 of 7 observers (false-positive). On the basis of clinical data and MRI follow-up, reference observer ruled out NET and confirmed presence of splenule abutting tail of pancreas.

TABLE 4
Interobserver Agreement on SUVs

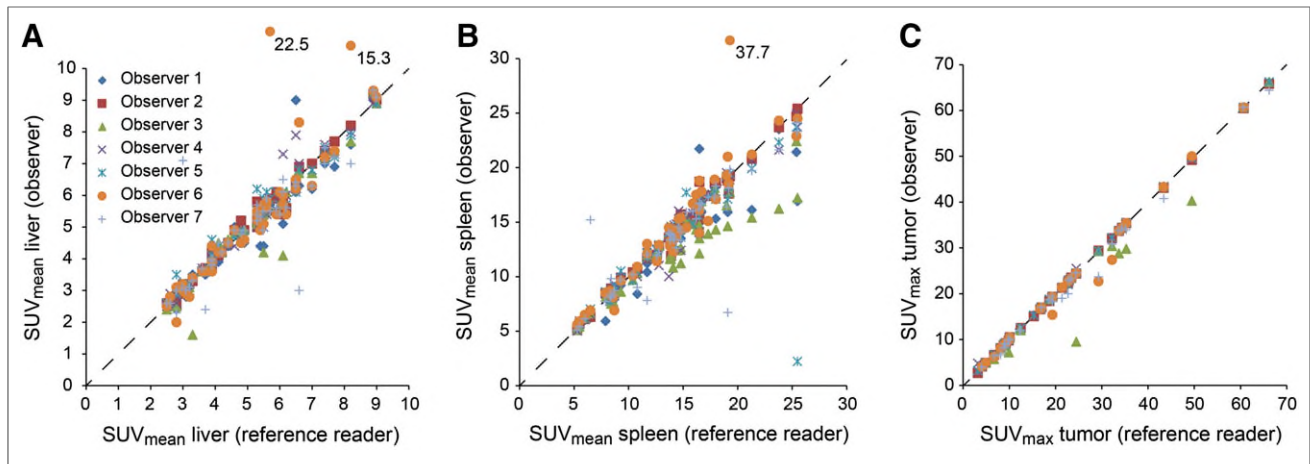| SUV type	ICC	95%CI	Mean Δ ± SD	
			Low experience	High experience
SUV_{mean}, liver	0.69	0.58–0.79	0.24 ± 0.34	0.47 ± 1.62
SUV_{max}, liver	0.79	0.70–0.86	0.46 ± 0.61	0.78 ± 1.20
SUV_{mean}, spleen	0.81	0.73–0.88	1.04 ± 1.47	1.17 ± 2.81
SUV_{max}, spleen	0.65	0.54–0.76	1.10 ± 4.70	1.07 ± 2.25
SUV_{max}, tumor	0.99	0.97–0.99	0.42 ± 1.73	0.46 ± 1.20

FIGURE 2. Interobserver agreement on SUV_{mean} for liver (A), SUV_{mean} for spleen (B), and SUV_{max} for tumor (C). SUVs were sorted by SUV obtained from reference observer. Dashed diagonal lines indicate perfect agreement. Three *y*-axis outliers were drawn in relative position outside scale, and absolute *y*-value is given.

CONCLUSION

Both visual and semiquantitative analysis of ^{68}Ga-DOTATATE PET/CT images is highly reproducible among observers with varying experience. Diagnostic information gained from ^{68}Ga-DOTATATE PET/CT in an appropriate clinical or research setting can thus be considered reliable. ^{68}Ga-DOTATATE PET/CT interpretation should be performed by experienced observers if PRRT is being considered.

DISCLOSURE

REFERENCES

1. Gabriel M, Decristoforo C, Kendler D, et al. ^{68}Ga-DOTA-Tyr3-octreotide PET in neuroendocrine tumors: comparison with somatostatin receptor scintigraphy and CT. *J Nucl Med.* 2007;48:508–518.
2. Srirajaskanthan R, Kayani I, Quigley AM, Soh J, Caplin ME, Bomanji J. The role of ^{68}Ga-DOTATATE PET in patients with neuroendocrine tumors and negative or equivocal findings on ^{111}In-DTPA-octreotide scintigraphy. *J Nucl Med.* 2010;51:875–882.
3. 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.
4. Deppen SA, Blume J, Bobbey AJ, et al. ^{68}Ga-DOTATATE compared with ^{111}In-DTPA-octreotide and conventional imaging for pulmonary and gastroenteropancreatic neuroendocrine tumors: a systematic review and meta-analysis. *J Nucl Med.* 2016;57:872–878.
5. Pape UF, Perren A, Niederle B, et al. ENETS consensus guidelines for the management of patients with neuroendocrine neoplasms from the jejuno-ileum and the appendix including goblet cell carcinomas. *Neuroendocrinology.* 2012;95:135–156.
6. Strosberg J, Wolin E, Chasen B, et al. ^{177}Lu-DOTATATE significantly improves progression-free survival in patients with midgut neuroendocrine tumors: results of the phase III NETTER-1 trial [abstract C39]. North American Neuroendocrine Tumor Society website. https://www.nanets.net/nanets_cd/2015/pdfs/C39.pdf. Published 2015. Accessed September 26, 2016.
7. Kwekkeboom DJ, Teunissen JJ, Bakker WH, et al. Radiolabeled somatostatin analog [^{177}Lu-DOTA0,Tyr^3]octreotate in patients with endocrine gastroenteropancreatic tumors. *J Clin Oncol.* 2005;23:2754–2762.
8. Herrmann K, Czernin J, Wolin EM, et al. Impact of ^{68}Ga-DOTATATE PET/CT on the management of neuroendocrine tumors: the referring physician's perspective. *J Nucl Med.* 2015;56:70–75.
9. Cohen JA. A coefficient of agreement for nominal scales. *Educ Psychol Meas.* 1960;20:37–46.
10. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics.* 1986;42:121–130.
11. Scheffe H. *The Analysis of Variance.* New York, NY: Wiley; 1959:221–260.
12. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159–174.
13. Kayani I, Bomanji JB, Groves A, et al. Functional imaging of neuroendocrine tumors with combined PET/CT using ^{68}Ga-DOTATATE (DOTA-DPhe1,Tyr^3-octreotate) and ^{18}F-FDG. *Cancer.* 2008;112:2447–2455.
14. Armani C, Catalani E, Balbarini A, Bagnoli P, Cervia D. Expression, pharmacology, and functional role of somatostatin receptor subtypes 1 and 2 in human macrophages. *J Leukoc Biol.* 2007;81:845–855.
15. Mapelli P, Tam HH, Sharma R, Aboagye EO, Al-Nahhas A. Frequency and significance of physiologic versus pathologic uptake of ^{68}Ga-DOTATATE in the pancreas: validation with morphologic imaging. *Nucl Med Commun.* 2014;35:613–619.
16. Hofman MS, Lau WF, Hicks RJ. Somatostatin receptor imaging with ^{68}Ga DOTATATE PET/CT: clinical utility, normal patterns, pearls, and pitfalls in interpretation. *Radiographics.* 2015;35:500–516.
17. Brogsitter C, Hofmockel T, Kotzerke J. ^{68}Ga DOTATATE uptake in vertebral hemangioma. *Clin Nucl Med.* 2014;39:462–463.
18. Naswa N, Sharma P, Kumar A, et al. Gallium-68-DOTA-NOC PET/CT of patients with gastroenteropancreatic neuroendocrine tumors: a prospective single-center study. *AJR.* 2011;197:1221–1228.
19. Haug AR, Cindea-Drimus R, Auernhammer CJ, et al. The role of ^{68}Ga-DOTATATE PET/CT in suspected neuroendocrine tumors. *J Nucl Med.* 2012;53:1686–1692.
20. Haug AR, Auernhammer CJ, Wangler B, et al. ^{68}Ga-DOTATATE PET/CT for the early prediction of response to somatostatin receptor-mediated radionuclide therapy in patients with well-differentiated neuroendocrine tumors. *J Nucl Med.* 2010;51:1349–1356.
21. Ruf J, Schiefer J, Furth C, et al. ^{68}Ga-DOTATOC PET/CT of neuroendocrine tumors: spotlight on the CT phases of a triple-phase protocol. *J Nucl Med.* 2011;52:697–704.
22. Sadowski SM, Neychev V, Millo C, et al. Prospective study of ^{68}Ga-DOTATATE positron emission tomography/computed tomography for detecting gastro-entero-pancreatic neuroendocrine tumors and unknown primary sites. *J Clin Oncol.* 2016;34:588–596.
23. Buchmann I, Henze M, Engelbrecht S, et al. Comparison of ^{68}Ga-DOTATOC PET and ^{111}In-DTPAOC (Octreoscan) SPECT in patients with neuroendocrine tumours. *Eur J Nucl Med Mol Imaging.* 2007;34:1617–1626.
24. Brenner R, Metens T, Bali M, Demetter P, Matos C. Pancreatic neuroendocrine tumor: added value of fusion of T2-weighted imaging and high b-value diffusion-weighted imaging for tumor detection. *Eur J Radiol.* 2012;81:e746–e749.
25. De Robertis R, D'Onofrio M, Zamboni G, et al. Pancreatic neuroendocrine neoplasms: clinical value of diffusion-weighted imaging. *Neuroendocrinology.* December 9, 2015 [Epub ahead of print].