EJNMMI Reports

ORIGINAL ARTICLE Open Access

Comparison of inter- and intraobserver agreement between [^{18}F]AlF-NOTA-octreotide and [^{68}Ga]Ga-DOTA-SSA PET/CT

Hannes Leupe[1], Niloefar Ahmadi Bidakhvidi[1], Karolien Goffin[1], Bliede Van den Broeck[2], Sander Jentjens[1], Annouschka Laenen[3], Elin Pauwels[1], Willem Lybaert[4], Eric Van Cutsem[5], Guy Bormans[6], Timon Vandamme[4,7], Frederik Cleeren[6], Jeroen Dekervel[5], Karen Geboes[8], Sigrid Stroobants[9], Chris Verslype[5] and Christophe M. Deroose[1*]

*Correspondence: christophe.deroose@uzleuven.be

[1] Nuclear Medicine, University Hospitals Leuven and Nuclear Medicine and Molecular Imaging, Department of Imaging and Pathology, KU Leuven, Herestraat 49, 3000 Leuven, Belgium
Full list of author information is available at the end of the article

Abstract

Purpose: This study compares inter- and intraobserver agreement between [^{18}F]AlF-NOTA-octreotide ([^{18}F]AlF-OC) and [^{68}Ga]Ga-DOTA-somatostatin analogues (SSAs) in PET/CT imaging for neuroendocrine neoplasm (NEN) patients.

Materials and methods: This is a secondary endpoint analysis from our multicenter trial (clin trial.gov identifier: NCT04552847) including 75 NEN patients who received both [^{68}Ga]Ga-DOTATATE (n = 56) or [^{68}Ga]Ga-DOTA-NOC (n = 19) and [^{18}F]AlF-OC PET imaging. Five readers assessed lesion detection and characterization across multiple organs, scoring lesions by number and conspicuity using a 5-point Likert scale. Agreement was measured using Gwet's agreement coefficient.

Results: Results demonstrated nearly perfect interobserver agreement for lesion characterization across all organs for both tracers (0.921 for [^{18}F]AlF-OC; 0.934 for [^{68}Ga]Ga-DOTA-SSA). Similar agreement was observed for the number of lesions across organs (0.736 for [^{18}F]AlF-OC and 0.749 for [^{68}Ga]Ga-DOTA-SSAs). Organ-specific analysis revealed strong agreement for bone and liver lesions, with slightly lower agreement for lymph nodes. Both tracers also showed excellent agreement in determining Krenning scores (0.925 for [^{18}F]AlF-OC and 0.927 for [^{68}Ga]Ga-DOTA-SSAs). While mean lesion conspicuity was similar between tracers, [^{18}F]AlF-OC had a higher global image quality score (4.22 vs. 3.86, $p < 0.0001$). Intraobserver agreement was consistent between tracers for lesion characterization (> 0.95 for both readers) and lesion count (> 0.80 for both readers).

Conclusion: [^{18}F]AlF-OC and [^{68}Ga]Ga-DOTA-SSAs demonstrate comparable and excellent inter- and intraobserver agreement, reinforcing the clinical interchangeability of [^{18}F]AlF-OC PET/CT with [^{68}Ga]Ga-DOTA-SSAs in routine practice.

Keywords: [^{18}F]AlF-NOTA-octreotide, [^{68}Ga]Ga-DOTATATE, Neuroendocrine tumor, Somatostatin receptor, Interobserver agreement

Introduction

Neuroendocrine neoplasms (NENs) are a heterogeneous group of tumors characterized by an overexpression of the somatostatin receptor (SSTR) (Bozkurt et al. 2017). Positron emission tomography/computed tomography (PET/CT) imaging with somatostatin analogues (SSAs) targeting the somatostatin receptor SSTR is a cornerstone in the clinical management of NEN patients. While [^{68}Ga]Ga-DOTA-SSA PET/CT imaging, the current gold standard in routine clinical practice, has revolutionized the detection and management of NENs, it does have several limitations. These include high cost and limited availability of germanium-68/gallium-68 generators, a short tracer half-life (67.6 min), limited production output (averaging 2–4 patients per batch), as well as a high positron energy resulting in a relatively long average positron range, which could impact spatial resolution. In recent years, fluorine-18-labeled PET tracers targeting the SSTR have garnered increasing attention for their potential advantages compared with gallium-68 (Pauwels et al. 2018). Fluorine-18 can be produced with a cyclotron and has a higher production yield (up to 15 patients per batch), a more favorable half-life (109.8 min) and lower average positron range (Leupe et al. 2023). These properties allow centralized production and distribution to distant PET centers without an on-site cyclotron or germanium-68/gallium-68 generator.

[^{18}F]AlF-NOTA-octreotide (or [^{18}F]AlF-OC) in particular has shown promising results in clinical trials. This [^{18}F]AlF-labeled SSTR agonist offers the advantages of fluorine-18 over gallium-68 and can be synthesized using a good-manufacturing-practice (GMP)–compliant, chelator-based, automated radiolabeling method (Tshibangu et al. 2020). Initial clinical trials comparing [^{18}F]AlF-OC PET/CT with [^{68}Ga]Ga-DOTA-SSA PET/CT demonstrated similar safety profiles, biodistribution and dosimetry (Pauwels et al. 2019, 2020; Hou et al. 2021; Haeger et al. 2023). To validate these results, Pauwels et al. conducted a prospective multicenter analysis of the diagnostic efficacy of [^{18}F]AlF-OC versus [^{68}Ga]Ga-DOTATATE or [^{68}Ga]Ga-DOTANOC in 75 patients with histologically confirmed NENs (Pauwels et al. 2023). They found a significantly higher detection ratio for [^{18}F]AlF-OC (91.1%) compared to [^{68}Ga]Ga-DOTA-SSAs (75.3%), illustrating diagnostic superiority, especially for the detection of liver metastases (Pauwels et al. 2023). Moreover, a secondary endpoint analysis on the collected data from this multicenter trial by our group showed that the use of [^{18}F]AlF-OC did not impact TNM staging or clinical management in the large majority of these NEN patients (86.7%), further validating the potential for routine clinical use of [^{18}F]AlF-OC PET/CT (Leupe et al. 2024; Hou et al. 2024).

This current study aims to validate the clinical interchangeability of [^{18}F]AlF-OC and [^{68}Ga]Ga-DOTA-SSA by comparing their inter- and intraobserver agreement in PET/CT imaging in NEN patients based on a read-out by five readers. This analysis, conducted as prespecified part of the larger multicenter trial (Pauwels et al. 2023), assesses lesion characterization, lesion number, lesion conspicuity, overall image quality and Krenning score, providing valuable insights for clinical adoption and read out of these PET tracers. Acquisition protocols were designed based on the clinical guidelines for [^{68}Ga]Ga-DOTA-SSA tracers and aimed to maximize the benefits of PET/CT imaging with fluorine-18-labeled SSA tracers, which offer higher production yield and a longer half-life.

Materials and methods

Study population and image acquisition

This study is a post-hoc analysis of the PET/CT images acquired in our previously published prospective multicenter trial (Pauwels et al. 2023). In this cohort, 75 NEN patients underwent both a [^{18}F]AlF-OC PET/CT and a routine clinical [^{68}Ga]Ga-DOTA-SSA PET/CT, either [^{68}Ga]Ga-DOTATATE (n = 56) or [^{68}Ga]Ga-DOTA-NOC (n = 19), within a three-month window of each other (median interval of 7 days and a maximum of 32 days). A detailed description of acquisition and reconstruction parameters (Pauwels et al. 2023) can be found in our previously published multicenter trial (Pauwels et al. 2023). Briefly, [^{18}F]AlF-OC was synthesized following an automated, GMP-compliant radiolabeling method outlined by Tshibangu et al. (2020), using an AllInOne® synthesis module (Trasis, Belgium). Similar to our current standardized operating procedure (SOP) for SSTR PET/CT imaging with [^{18}F]AlF-OC in UZ Leuven, patients received an intravenous injection of 4 MBq/kg of [^{18}F]AlF-OC, followed by a whole-body PET/CT scan, conducted at 2 h post-administration, using either a GE Discovery MI 4-ring PET/CT system (GE Healthcare, WI, USA) at 3 min. PET acquisition per bed position or a Siemens Biograph 40 Truepoint system (Siemens Medical, Germany) at 4 min. PET acquisition per bed position. The mean administered activity (± standard deviation) of [^{18}F]AlF-OC was 295 (± 60) MBq. Following intravenous injection of either [^{68}Ga]Ga-DOTATATE (UZ Leuven; n = 56; 45–60 min post-injection) or [^{68}Ga]Ga-DOTANOC (UZ Antwerp; n = 19; 60–90 min post-injection), a whole-body PET/CT scan was performed in accordance with the guidelines outlined by EANM (Bozkurt et al. 2017). Patients from UZ Antwerp underwent scanning using a GE Discovery MI 4-ring or MI 3-ring PET/CT system at 2.5 min. PET acquisition per bed position. The mean administered activity for [^{68}Ga]Ga-DOTATATE was 139 ± 27 MBq, while for [^{68}Ga]Ga-DOTANOC, it was 114 ± 20 MBq. All patients were advised to refrain from long-acting SSAs for a period of four to six weeks prior to the scan, irrespective of the tracer administered. The choice of a higher administered activity for [^{18}F]AlF-OC was based on the advantages offered by ^{18}F-labeling via a cyclotron, which allows for higher production yield of the tracer. Furthermore, the imaging time point of 2 h post-injection was selected based on findings from Pauwels et al. (2020), which demonstrated improved lesion detection at 3 h compared to 2 and 1 h post-injection. As a compromise between clinical feasibility and diagnostic performance, a 2-h post-injection time was chosen. Additionally, the longer half-life of fluorine-18 and the higher available activity make extended scanning durations without significant signal loss possible, allowing for prolonged acquisition times to optimize image quality.

Visual PET/CT image analysis

[^{18}F]AlF-OC PET/CT and [^{68}Ga]Ga-DOTA-SSA PET/CT images were analyzed using a dedicated software platform for multimodal image analysis (MIM version 7.3; MIM Software Inc., Cleveland, OH, USA). The data included standard Digital Imaging and Communications in Medicine (DICOM) files of computed tomography (CT) and attenuation-corrected positron emission tomography (PET) images. Clinical information related to the patient and the used radiopharmaceutical was eliminated from the

DICOM headers. Since [^{68}Ga]Ga-DOTA-SSAs have higher physiological salivary gland uptake compared to [^{18}F]AlF-OC (Pauwels et al. 2020; Hou et al. 2021), an independent operator masked the head region of all PET datasets. For each patient, a total of 19 organs (liver, bone, lymph nodes, small intestine, pancreas, peritoneum, heart, lung, pleura, oesophagus, stomach, colon, anorectum, kidney, spleen, adrenal gland, bone, muscle and breast) were evaluated for the presence of SSTR-positive lesions. Lymph node assessment was categorized as supradiaphragmatic, infradiaphragmatic and all lymph nodes. SSTR-positive lesions were considered as any lesion with [^{18}F]AlF-OC PET/CT or [^{68}Ga]Ga-DOTA-SSA uptake above the respective physiological uptake in the organ.

A randomized read-out of the masked images from both tracers was performed by 5 readers: 4 board-certified nuclear medicine physicians (KG, SJ, BVdB and NAB) and one nuclear medicine physician in training (HL). Readers were divided into two groups based on their experience: 2 highly experienced readers ('experienced', KG and SJ with 17 and 13 years of experience in interpreting SSTR PET, respectively) and 3 readers with relatively low experience ('less experienced', BvdB, NAB and HL with 8, 7 and 1 year(s) of experience in interpreting SSTR PET, respectively). Each year of experience correlates with 50–100 SSTR PET read-outs. The read-outs were performed independently from each other and blinded to clinical information and the tracer used.

The PET/CT scans were evaluated using a methodology similar to the one used in an interobserver study by Lens et al. (2023). All SSTR PET/CT scans were evaluated and rated for lesion characterization per organ using a 5-point Likert scale (1 = benign, 2 = probably benign, 3 = equivocal/indeterminate, 4 = probably malignant, 5 = malignant). If lesions within an organ had varying characterization scores, the highest score was recorded. The number of lesions corresponding to the assigned characterization score was determined for each organ. Regarding number of lesions per organ, lesions with characterization scores 1, 2 and 3 were grouped, as well as lesions with scores 4 and 5. Six bins were used to categorize the number of lesions (0, 1, 2–5, 6–10, 11–20 and > 20/diffuse/uncountable). Initially, > 20 and diffuse/uncountable were scored as separate bins, but were afterwards combined into a single category based on reader feedback regarding the ambiguous distinction between these groups. Furthermore, lesion conspicuity or visibility in contrast to surrounding tissue uptake (background uptake) was scored for malignant lesions (characterization 4 or 5) per organ using a 5-point Likert scale (1 = no increased uptake, 2 = barely visible, 3 = uptake higher than background, 4 = certainly detectable/uptake visually 2 × higher than background, 5 = clearly visible/uptake higher than all organs including the spleen). If lesions within an organ had varying conspicuity scores, the highest score was recorded. Additionally, each PET/CT scan received a visual global image quality score (5-point Likert scale: 1 = non-diagnostic, 2 = bad, 3 = average, 4 = good, 5 = excellent) and a global modified Krenning score based on the lesion with the highest uptake (1 = uptake much lower than liver, 2 = uptake slightly less than or equal to liver, 3 = uptake greater than liver, 4 = uptake greater than spleen) (Hope et al. 2019). All readers were trained on the read-out methodology prior to starting image analysis.

To evaluate intraobserver agreement, two readers (NAB and HL) reevaluated the PET/CT images from both tracers after a minimum interval of one month to minimize recall

bias. This reevaluation was conducted on a randomly selected subset of 20 patients, categorized by the number of lesions observed in the initial [^{68}Ga]Ga-DOTA-SSA PET/CT read-out: 5 patients with 0–25 lesions, 5 with 25–50 lesions, 5 with 50–75 lesions, and 5 with more than 75 lesions.

Statistical analysis

The Gwet's agreement coefficient (AC) 2 statistic was determined as a measure of agreement on ordinal scores. The Gwet's AC1 statistic was applied in case of binary ratings as a measure of agreement on categorical scores. Gwet's approach is applicable to two or more raters, and additionally handles missing data (Gwet 2001). Similar to the Kappa statistic, it is a chance-corrected measure of agreement, taking values from − 1 up to 1 with higher values indicating better agreement, though it usually falls between 0 and 1. The analyses were performed using the SAS MAGREE macro version 3.91 of SAS software (version 9.4 for Windows). Gwet's AC2 and Gwet's AC1 statistics are reported with their 95% confidence intervals and differences are interpreted upon clinical relevance. The Gwet's AC statistic is most commonly analyzed in a similar way to Fleiss' Kappa using the interpretation guidelines proposed by Landis and Koch and is summarized as follows (Landis and Koch 1977): < 0.0 = poor agreement; 0.0–0.20 = slight agreement; 0.21–0.40 = fair agreement; 0.41–0.60 = moderate agreement; 0.61–0.80 = substantial agreement; 0.81–1.0 = almost perfect agreement. Descriptive statistics were used to compare lesion conspicuity, Krenning score and global image quality.

For the statistical analysis of lesion characterization, results were dichotomized in benign (lesion characterization score 1, 2 or 3) versus malignant (lesion characterization score 4 or 5) to perform an additional sub-analysis for binary lesion characterization. For the statistical analysis of lesion count, only malignant lesions (characterization score 4 or 5) were considered. Additionally, a sub-analysis was conducted to compare lesion characterization and lesion count between experienced and less experienced readers.

Results

Patient characteristics

Patient characteristics are summarized in Table 1. The median age was 65 years (range 37–84 years). The primary tumors were mainly from intestinal (45/75 patients) or pancreatic (18/75 patients) origin. Organ involvement was determined based on the evaluation of all PET/CT images, and an organ was considered involved if at least one reader scored the presence of at least one lesion in that organ as either 4 (probably malignant) or 5 (definitely malignant). In total, 830 PET/CT observations were acquired: 750 (75 patients × 2 tracers × 5 readers) from the interobserver agreement and an additional 80 (20 patients × 2 tracers × 2 readers) from intraobserver agreement.

Interobserver agreement

Substantial to almost perfect interobserver agreement was observed for lesion characterization within individual organs using both ordinal (characterization scores 1 vs. 2 vs. 3 vs. 4 vs. 5) and binary analyses (characterization scores 1 + 2 + 3 vs. 4 + 5) (Table 2, Fig. 1a). When considering all organs collectively, the agreement was almost perfect and highly comparable between the two analysis methods (Fig. 1b). For the ordinal analysis

Table 1 Patient and clinical characteristics (n = 75)

Characteristic	Number (%) of patients or median (range)	
Age (y)	65 (37–84)	
Sex		
Male	46 (61.3%)	
Female	29 (38.7%)	
Primary tumor		
Intestine	45 (60.0%)	
Pancreas	18 (24.0%)	
Lung	7 (9.3%)	
CUP	4 (5.3%)	
Paraganglioma	1 (1.3%)	
Tumor grade		
G1*	37 (49.3%)	
G2	34 (45.3%)	
G3	2 (2.7%)	
NA	2 (2.7%)	
Ki-67 (%)	2.5 (0.4–29)	
Organ involvement (≥ 1 reader)	[^{68}Ga]Ga-DOTA-SSA	[^{18}F]AlF-OC
Lymph nodes	73 (97.3%)	75 (100.0%)
Liver	60 (80.0%)	58 (77.3%)
Bone	56 (74.7%)	56 (74.7%)
Pancreas	44 (58.7%)	48 (64.0%)
Peritoneum/omentum	35 (46.7%)	38 (50.7%)
Lung	27 (36.0%)	38 (50.7%)
Small intestine	18 (24.0%)	18 (24.0%)
Pleura	16 (21.3%)	31 (41.3%)
Prostate	17 (22.7%)	18 (24.0%)
Muscle	8 (10.7%)	14 (18.7%)
Total number of lesions (highest number of both tracers)		
1–25	30 (40.0%)	
25–50	12 (16.0%)	
50–70	10 (13.3%)	
> 75	23 (30.7%)	

*In the G1 group, 2 patients with Ki67% index < 5% were included

over all organs, the agreement coefficients were 0.918 (95% CI 0.899–0.937) for [^{68}Ga]Ga-DOTA-SSAs and 0.908 (95% CI 0.892–0.927) for [^{18}F]AlF-OC. Similarly, for the binary analysis, the agreement coefficients were 0.934 (95% CI 0.923–0.946) for [^{68}Ga]Ga-DOTA-SSAs and 0.921 (95% CI 0.904–0.937) for [^{18}F]AlF-OC.

Furthermore, interobserver agreement on the number of malignant lesions (characterization score 4 or 5) categorized in bins was substantial to almost perfect within individual organs (Supplemental Table 1, Fig. 2a). A slightly lower agreement was found for lymph nodes (0.537 and 0.502) and peritoneal/omental lesions (0.530 and 0.642) for [^{68}Ga]Ga-DOTA-SSAs and [^{18}F]AlF-OC, respectively. Agreement for all organs showed substantial agreement, with Gwet's AC2 agreement coefficients of 0.749 (95% CI 0.656–0.843) for [^{68}Ga]Ga-DOTA-SSAs and 0.736 (95% CI 0.654–0.818) for [^{18}F]AlF-OC (Fig. 2b).

Table 2 Comparison of ordinal and binary interobserver agreement in lesion characterization of [^{68}Ga]Ga-DOTA-SSA and [^{18}F]AlF-OC PET/CT across individual and all organs

Lesion characterization				
	Ordinal (1,2,3,4,5)		Binary (1 + 2 + 3 vs. 4 + 5)	
	[^{68}Ga]Ga-DOTA-SSA	[^{18}F]AlF-OC	[^{68}Ga]Ga-DOTA-SSA	[^{18}F]AlF-OC
Location	AC2 (95% CI)	AC2 (95% CI)	AC1 (95% CI)	AC1 (95% CI)
Per organ				
Bone	0.770 (0.674–0.866)	0.788 (0.654–0.921)	0.778 (0.658–0.899)	0.799 (0.665–0.932)
Heart	0.966 (0.933–0.999)	0.960 (0.925–0.995)	0.960 (0.919–1.000)	0.952 (0.904–1.000)
Liver	0.818 (0.714–0.923)	0.902 (0.831–0.974)	0.815 (0.697–0.932)	0.929 (0.852–1.000)
Lung	0.881 (0.806–0.955)	0.809 (0.715–0.903)	0.892 (0.816–0.969)	0.828 (0.735–0.922)
LN (all)	0.601 (0.452–0.750)	0.717 (0.566–0.868)	0.626 (0.444–0.808)	0.754 (0.598–0.910)
LN (infradiaphragmatic)	0.686 (0.550–0.823)	0.700 (0.575–0.825)	0.701 (0.533–0.869)	0.712 (0.579–0.846)
LN (supradiaphragmatic)	0.646 (0.517–0.775)	0.605 (0.429–0.781)	0.708 (0.579–0.836)	0.650 (0.452–0.847)
Muscle	0.962 (0.928–0.996)	0.932 (0.885–0.978)	0.959 (0.920–0.998)	0.924 (0.863–0.984)
Pancreas	0.718 (0.583–0.853)	0.639 (0.489–0.789)	0.751 (0.597–0.906)	0.637 (0.491–0.782)
Peritoneum/omentum	0.744 (0.560–0.929)	0.661 (0.462–0.861)	0.702 (0.467–0.938)	0.614 (0.378–0.850)
Pleura	0.919 (0.844–0.995)	0.842 (0.711–0.972)	0.909 (0.814–1.000)	0.823 (0.681–0.964)
Prostate	0.926 (0.853–0.999)	0.905 (0.814–0.997)	0.995 (0.980–1.000)	0.986 (0.963–1.000)
Small intestine	0.898 (0.844–0.953)	0.894 (0.820–0.968)	0.898 (0.831–0.965)	0.887 (0.805–0.968)
All organs	0.918 (0.899–0.937)	0.908 (0.890–0.927)	0.934 (0.923–0.946)	0.921 (0.904–0.937)

Gwet's AC2 = agreement coefficient for ordinal ratings by multiple raters. Gwet's AC1 = agreement coefficient for categorical rating by multiple raters; LN = lymph nodes

We found no clinically meaningful differences between the mean lesion conspicuity per organ for [^{68}Ga]Ga-DOTA-SSAs and [^{18}F]AlF-OC (Fig. 3a, Supplemental Table 4). Global image quality was significantly higher with [^{18}F]AlF-OC than with [^{68}Ga]Ga-DOTA-SSAs (mean score 4.22 (95% CI 3.85–4.58) vs 3.86 (95% CI 3.50–4.23), $p < 0.0001$; Figs. 3b, 4). Agreement on Krenning score was almost perfect for both tracers, with Gwet's AC2 agreement coefficients of 0.927 (95% CI 0.886–0.969) for [^{68}Ga]Ga-DOTA-SSAs and 0.925 (0.886–0.964) for [^{18}F]AlF-OC (Fig. 3c).

Intraobserver agreement

Analysis of intraobserver agreement in two readers showed high reproducibility in terms of lesion characterization per organ and across all organs (Fig. 5a, b, Supplemental Table 5), with Gwet's AC2 values of 0.910 or higher across all organs for both readers and both tracers. Similarly, no clinically relevant differences were found between both tracers in terms of lesion count, both at an organ level and across all organs (Fig. 5c, d, Supplemental Table 6). Gwet's AC2 values for intraobserver lesion number scoring were 0.899 and 0.909 for reader 1, and 0.791 and 0.866 for reader 2 for [^{68}Ga]Ga-DOTA-SSAs and [^{18}F]AlF-OC, respectively.

Experience level

The sub-analysis based on reader experience showed no clinically meaningful differences between experienced readers and less experienced readers (Fig. 1c–e). Similar to lesion characterization, the sub-analysis based on reader experience showed no

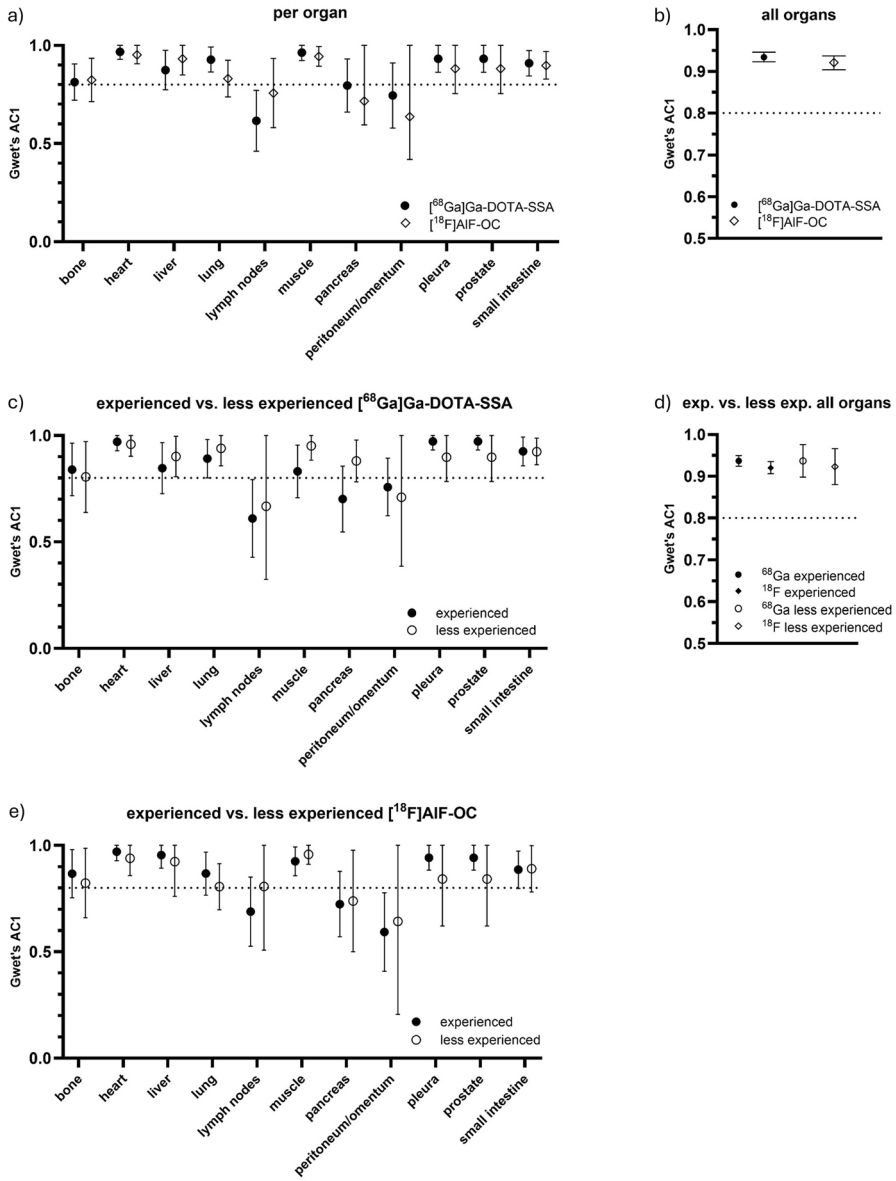

Fig. 1 Comparison of interobserver agreement (Gwet's AC1) in binary lesion characterization (score 1 + 2 + 3 vs. 4 + 5) for [^{68}Ga]Ga-DOTA-SSA PET/CT and [^{18}F]AlF-OC PET/CT: analysis across individual organs (**a**) and all organs (**b**) and comparison between experienced versus less experienced readers across individual (**c**, **e**) and all organs (**d**). The error bars in the graphs represent the lower and upper limit of the 95% confidence intervals, whereas the square/circle represent the Gwet's AC value. The horizontal dotted line represents the cut-off between substantial and almost perfect agreement

clinically meaningful differences between experienced readers and less experienced readers (Fig. 2c–e). With both tracers, agreement on lesion count was somewhat lower for lymph nodes and peritoneal/omental lesions in the experienced group compared to the less experienced group.

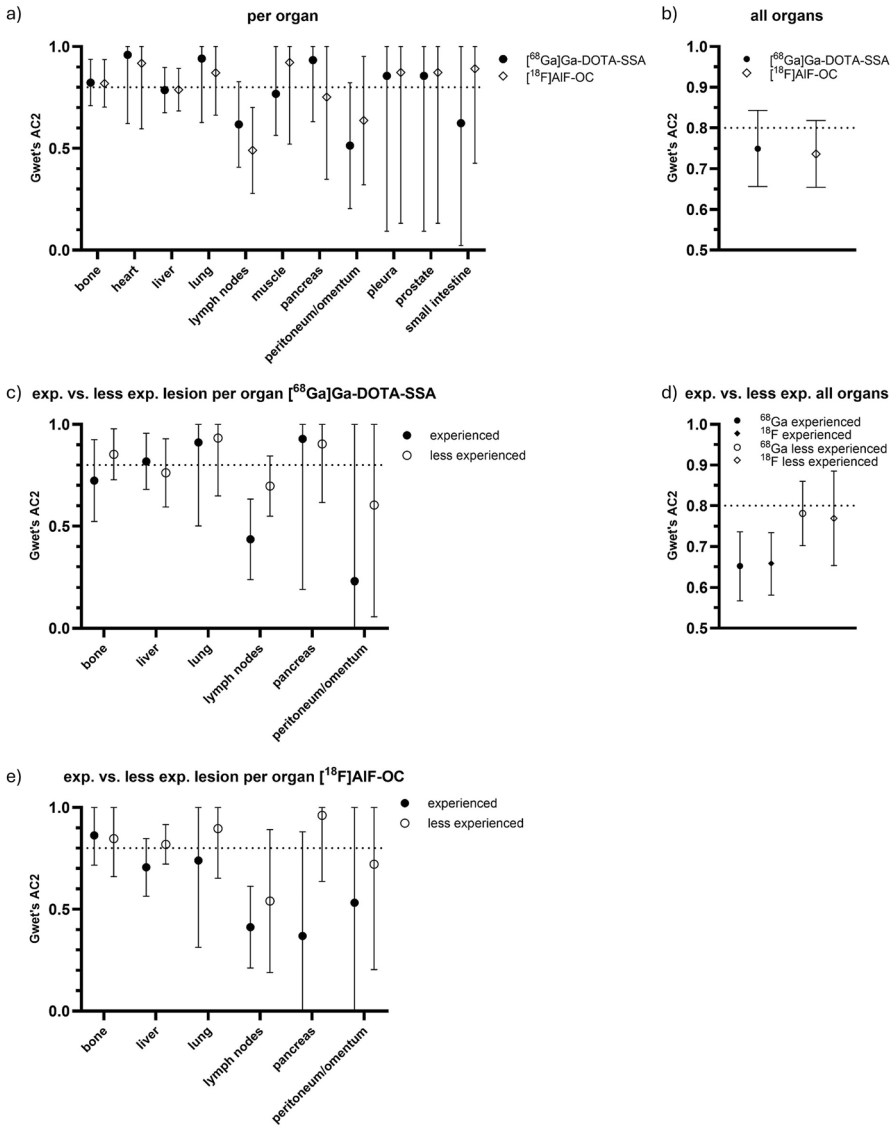

Fig. 2 Comparison of interobserver agreement (Gwet's AC2) in malignant lesion count (characterization score 4 or 5) for [^{68}Ga]Ga-DOTA-SSA PET/CT and [^{18}F]AlF-OC PET/CT: analysis across individual organs (**a**) and all organs (**b**) and comparison between experienced versus less experienced readers across individual (**c**, **e**) and all organs (**d**). The error bars in the graphs represent the lower and upper limit of the 95% confidence intervals, whereas the square/circle represent the Gwet's AC value. The horizontal dotted line represents the cut-off between substantial and almost perfect agreement

Discussion

In this study (Leupe et al. 2023, 2024; Pauwels et al. 2023), we conducted a comprehensive evaluation of inter- and intraobserver agreement for the use of [^{18}F]AlF-OC and [^{68}Ga]Ga-DOTA-SSAs in PET/CT imaging of NEN patients. Interobserver agreement analysis is crucial for assessing a PET/CT tracer's reliability, consistency, and accuracy across different observers. We chose the Gwet's AC statistic to evaluate observer agreement because it is applicable to two or more raters and it handles missing data better than the Kappa statistic. To the best of our knowledge, this is the first

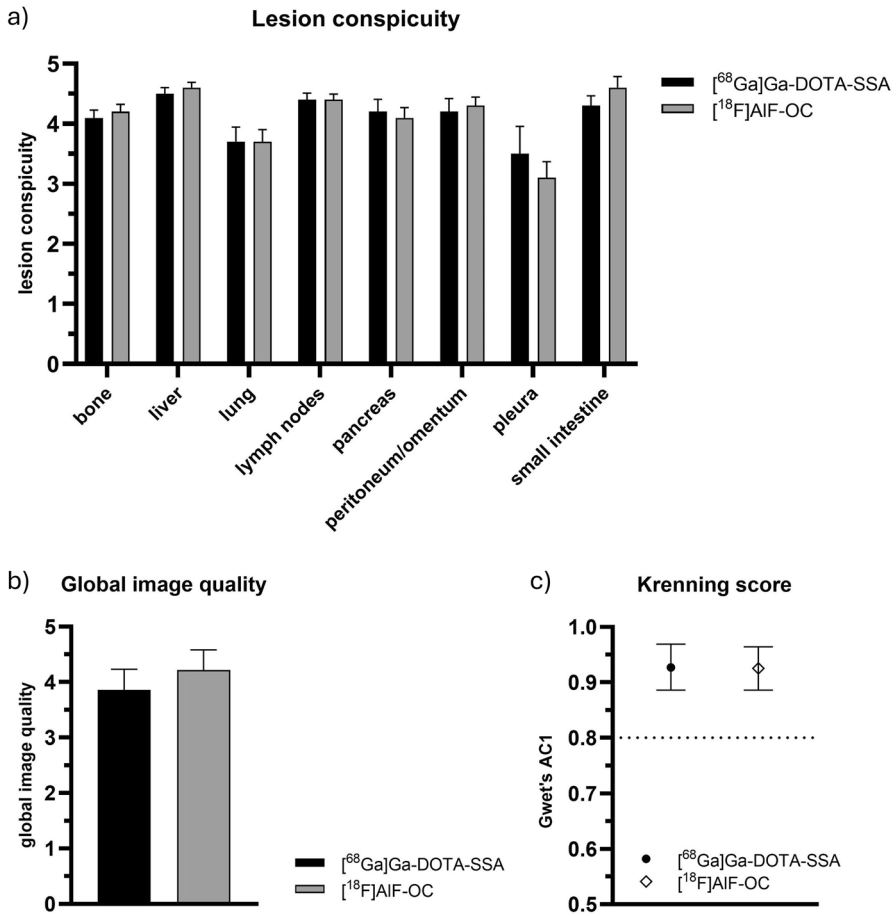

Fig. 3 Comparison of mean lesion conspicuity per organ for the most commonly affected organs (30 observations or more for all readers combined based on [^{68}Ga]Ga-DOTA-SSA PET/CT) (**a**), mean global image quality (*p < 0.0001) (**b**) and Gwet's AC1 for Krenning score (**c**) for [^{68}Ga]Ga-DOTA-SSA PET/CT and [^{18}F]AlF-OC PET/CT. The error bars in the graphs represent the lower and/or upper limit of the 95% confidence intervals, whereas the square/circle represent the Gwet's AC value

study to directly compare inter- and intraobserver agreement between [^{18}F]AlF-OC and [^{68}Ga]Ga-DOTA-SSAs, based on image evaluations by five readers.

Our results demonstrated an almost perfect interobserver agreement for lesion characterization across all organs for both tracers, with Gwet's AC1 agreement coefficients of 0.934 for [^{68}Ga]Ga-DOTA-SSAs and 0.921 for [^{18}F]AlF-OC for the binary analysis (benign vs. malignant). This high level of agreement underscores the reliability of both [^{18}F]AlF-OC and [^{68}Ga]Ga-DOTA-SSAs in the assessment of NEN lesions. Furthermore, no clinically meaningful differences were found when comparing the experienced with the less experienced readers. This consistency underscores the robustness of lesion characterization across different levels of reader expertise for both tracers.

These results are in line with previous studies. A prospective study by Fendler et al. involving 50 [^{68}Ga]Ga-DOTATATE PET/CT scans in NEN patients demonstrated high interobserver reliability between seven readers (350 observations) in terms of organ involvement ($\kappa = 0.70$) and number of organs affected (ICC = 0.84), even when images were interpreted by less experienced observers (Fendler et al. 2017). Additionally,

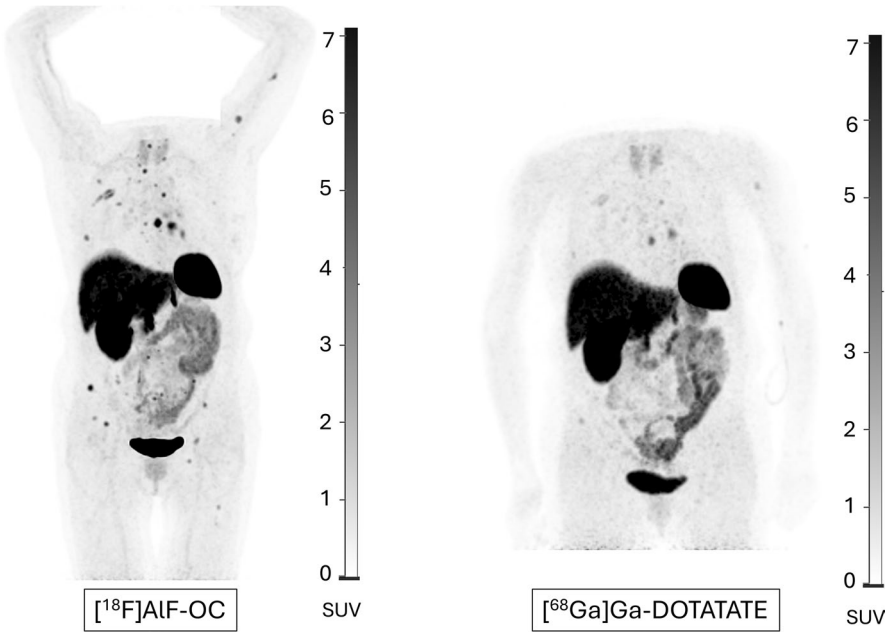

Fig. 4 Comparison of image quality between [^{18}F]AlF-OC and [^{68}Ga]Ga-DOTATATE PET/CT maximal intensity projection (MIP) images from the same patient (65-year-old female with a NET of unknown primary origin, Ki67 index 12%, and metastases in the bone, liver, and lymph nodes). The overall image quality of [^{18}F]AlF-OC PET/CT was rated higher, with an average score of 4.6 on a 5-point Likert scale by five readers, compared to an average score of 3.4 for [^{68}Ga]Ga-DOTATATE PET/CT. Note: the head region was removed from the PET/CT images to account for differences in tracer uptake in the salivary glands (Pauwels et al. 2020; Hou et al. 2021), ensuring a blinded PET/CT readout

Deppen et al. reported almost perfect reproducibility (k = 0.82) between two blinded observers and one non-blinded observer in interpreting [^{68}Ga]Ga-DOTATATE PET/CT images of 78 NEN patients (234 observations) (Deppen et al. 2016). Ruf et al. also found substantial agreement for PET (κ = 0.77) but only fair to moderate agreement for CT (κ = 0.48) in the assessment of [^{68}Ga]Ga-DOTATOC PET and CT images of 51 NEN patients by two readers (102 observations) (Ruf et al. 2011). This emphasizes the superior reliability of PET imaging over traditional CT scans for NEN evaluation, consistent with our high interobserver agreement results. Our study, with 830 observations, is the largest study on inter- and intraobserver variability for SSTR PET and the first one assessing two different PET tracers in the same study in NEN patients.

Organ-specific analysis revealed almost perfect agreement for most organs, including bone and liver lesions in both tracers, which are critical sites for NEN metastases. However, there was a notable discrepancy in the agreement for lymph node and peritoneum/omentum assessment, where both tracers showed slightly lower agreement in the number of lesions compared to other organs. The lower agreement for lymph nodes lesions can be attributed to a discrepancy in the lesion number scoring of malignant lymph nodes by two readers in patients with diffuse lymph node involvement (Supplemental Fig. 1) of both tracers. Additionally, in patients with diffuse peritoneal/omental involvement, counting separate lesions can be challenging due to the confluence of different lesions. Furthermore, the differentiation between lymph nodes and peritoneal/omental lesions can be challenging with SSTR PET imaging without contrast-enhanced CT, as

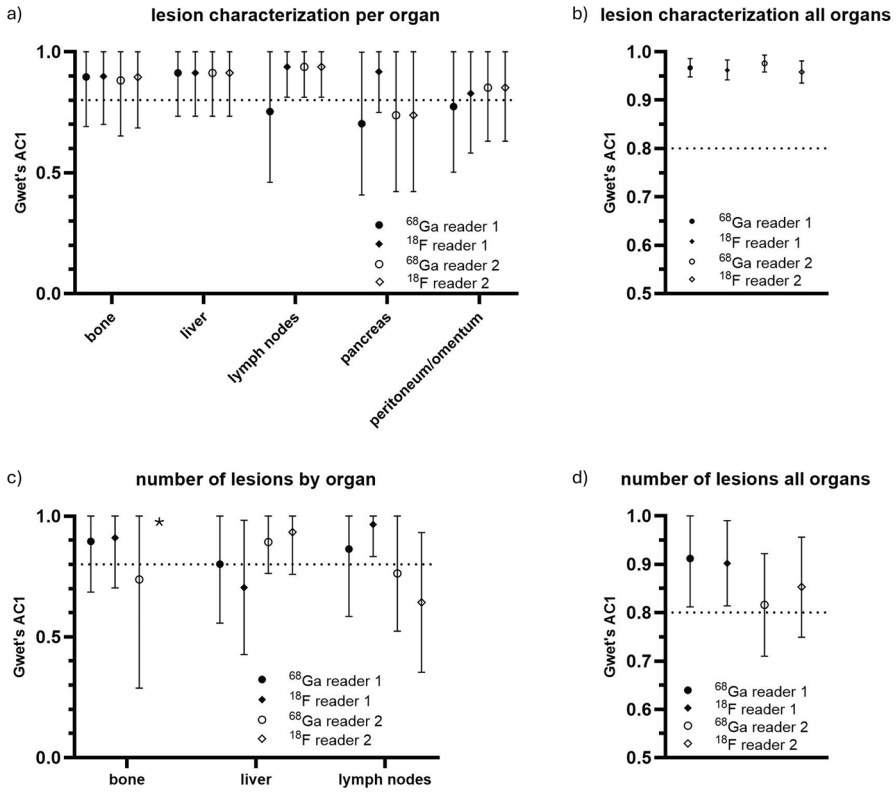

Fig. 5 Comparison of intraobserver agreement on lesion characterization, both per organ (**a**) and across all organs (**b**) and malignant lesion count (characterization 4 or 5) per organ, both per organ (**c**) and across all organs (**d**). *For [^{18}F]AlF-OC, all ratings by reader 2 were identical in bone, so no statistics on variability could be computed. Only organs for which an intraobserver agreement coefficient could be calculated are shown. The error bars in the graphs represent the lower and upper limit of the 95% confidence intervals, whereas the square/circle represent the Gwet's AC value. The horizontal dotted line represents the cut-off between substantial and almost perfect agreement

was the case in this study. Differences in the number of patients with pleural or lung involvement between the two tracers were influenced by several cases where at least one reader detected a single lesion or oligometastatic disease on [^{18}F]AlF-OC PET/CT, whereas [^{68}Ga]Ga-DOTA-SSA PET/CT did not. Lung lesions were confirmed by reviewing the corresponding CT-images. The superiority of [^{18}F]AlF-OC PET/CT in detecting lung lesions was also demonstrated in our multicenter trial (Pauwels et al. 2023).

The intraobserver agreement analysis, based on a reevaluation of 20 patients by two readers also highlighted the reproducibility of lesion characterization and counting using both tracers. The consistency in intraobserver agreement supports the robustness of [^{18}F]AlF-OC in routine clinical use, as it indicates that individual readers can reliably interpret scans from this tracer over multiple assessments.

Both tracers showed similar mean lesion conspicuity across all organs, suggesting that the visibility and contrast of lesions are highly comparable between [^{18}F]AlF-OC and [^{68}Ga]Ga-DOTA-SSAs. However, [^{18}F]AlF-OC demonstrated a higher mean global image quality score, which can be at least partially explained by higher injected activity, longer scan time and longer injection-scan interval. The reproducibility of Krenning score was also almost perfect for both tracers, which is important in the

clinical decision making and more specifically in selecting patients for peptide receptor radionuclide therapy (PRRT) (Bidakhvidi et al. 2022).

Our study's findings have important clinical implications. Interobserver agreement plays a crucial role in the evaluation of new PET tracers, ensuring that the results are reliable and reproducible across different clinicians (Cummings and Kinney 2022; Food and Drug Administration 2018). High interobserver agreement indicates that a tracer provides consistent diagnostic information, regardless of the observer's experience level, which is essential for its clinical adoption. For [^{18}F]AlF-OC, achieving strong agreement across multiple readers reinforces its reliability in detecting lesions, matching or exceeding the performance of established tracers like [^{68}Ga]Ga-DOTA-SSAs. This consistency across readers is key to validating [^{18}F]AlF-OC as a PET tracer for accurate imaging in NEN patients. Moreover, the higher global image quality associated with [^{18}F]AlF-OC may improve diagnostic confidence and accuracy, potentially leading to better patient outcomes. The findings further support the interchangeability of [^{18}F]AlF-OC and [^{68}Ga]Ga-DOTA-SSAs, establishing [^{18}F]AlF-OC as a validated SSTR PET tracer for routine clinical use.

The use of long axial field-of-view (LAFOV) scanners would likely yield results in line with the current study: we expect excellent lesion detection and high inter and intra-observer reproducibility. Due to their higher sensitivity and spatial resolution, LAFOV scanners could provide more comparable results between the tracers used. LAFOV scanners with > 1m axial FOV warrant an camera-specific optimized imaging protocol, where lower injected activities (e.g. 2 or even 1 MBq/kg can be considered). Additionally, the shorter positron range of fluorine-18 used in [^{18}F]AlF-OC might offer an advantage in detecting small lesions, similar as observed from [^{64}Cu]Cu-DOTATATE (Johnbeck et al. 2017). On these cameras, the theoretical advantages of fluorine-18 might emerge more clearly than on classical PET cameras.

While the study provides robust evidence showing reliable read-out, supporting the use of [^{18}F]AlF-OC, some limitations should be acknowledged. Firstly, evaluation of intraobserver agreement was performed based on a reevaluation of only two readers, who were both categorized as less experienced readers (HL and NAB). Given that there is no significant difference in terms of intraobserver agreement for both readers, we hypothesize similar or even better results for experienced readers. Secondly, our analysis was not powered to show statistically significant differences in agreement coefficients between both tracers. This was not the primary endpoint of our trial, which was powered to demonstrate non-inferiority in lesion detection. Additionally, differences in injected activity and post-injection acquisition time between the two tracers can influence our study outcomes. However, they reflect distinct protocols for SSTR PET/CT imaging that were developed to optimize imaging based on the production routes, yields and physical characteristics of the radionuclides used. These protocols were designed based on the clinical guidelines for [^{68}Ga]Ga-DOTA-SSA tracers applicable at the time of study conduct and aimed to maximize the benefits of PET/CT imaging with fluorine-18-labeled SSA tracers, which offer higher production yield and a longer half-life. Our study aimed to compare imaging protocols rather than the tracers themselves, with differences in protocols driven by the differences in the radionuclides and where each one is used in its optimal setting. We acknowledge

that this was not a comparison in terms of absolute tracer performance but rather an evaluation of distinct imaging protocols.

Conclusion

In conclusion, [^{18}F]AlF-OC demonstrates comparable inter- and intraobserver agreement to [^{68}Ga]Ga-DOTA-SSAs in PET/CT imaging of NEN patients, with the added benefits of slightly higher image quality and practical advantages in production and distribution. These findings contribute to the growing body of evidence supporting the use of [^{18}F]AlF-OC alongside [^{68}Ga]Ga-DOTA-SSAs in the imaging and management of NEN patients, establishing [^{18}F]AlF-OC as a validated SSTR PET tracer for routine clinical use.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s41824-025-00250-y.

Additional file 1.

Acknowledgements
Christophe M. Deroose and Timon Vandamme are Senior Clinical Investigators and Niloefar Ahmadi Bidakhvidi a PhD fellow at the Research Foundation—Flanders (FWO). Frederik Cleeren has been awarded an ERF Nuclear Medicine pilot research grant award by the Neuroendocrine Tumor Research Foundation (NETRF).

Author contributions
We declare that this manuscript is our original, unpublished work and that there are no related manuscripts currently submitted or in press elsewhere. We declare that all authors have contributed, read and approved the final manuscript. The specific contributions are as follows: Design of study: HL, EP, AL, FC, GB, SS, TV, BVdB, CMD; Data collection (PET readout): HL, NAB, KG, BVdB, SJ; Data analysis: HL, AL, CMD; Writing manuscript, including final approval: all.

Funding
This research was funded by the Project from Kom op tegen Kanker: "PET/MRI of the Norepinephrine Transporter and Somatostatin Receptor in Neural Crest and Neuroendocrine Tumors for Better Radionuclide Therapy Selection."

Availability of data and materials
The datasets generated during and/or analysed during the current study are not publicly available, but are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
All procedures performed in the study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee (Ethische Commissie Onderzoek UZ/KU Leuven S63678) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Written informed consent was obtained from all individual participants included in the study (NCT03883776).

Consent to publish
Informed written consent was obtained from all participants, with clear communication that the study's findings and data would be published in a scientific journal. All participants agreed to this condition.

Competing interests
No conflicts of interest (CoIs) relevant to this work exist. Nonrelated potential CoIs include: CMD is/has been a consultant for (fees received by institution): Sirtex, Advanced Accelerator Applications, Novartis, Ipsen, Terumo, PSI CRO, Immedica Pharma, Molecular Partners, Nuclear Research and consultancy. His institution has received travel fees from: GE Healthcare, Sirtex. Timon Vandamme received consultancy fees from AstraZeneca, Astellas, Bayer, Bristol Myers Squibb, Eisai, Elmedix, Ipsen, MyNeoTx, Nordic Pharma, Novartis, Roche, Sirtex, Servier, Takeda; research funding (institutional) from Ipsen and Novartis; support for travel/accomodation from Ipsen, AstraZeneca and Servier. KG received consultancy fees from Bayer, Novartis, Telix Pharmaceuticals, GE Healthcare, Lightpoint, Blue Earth Diagnostics, Merck Sharp & Dome, PSI CRO.

Author details
[1]Nuclear Medicine, University Hospitals Leuven and Nuclear Medicine and Molecular Imaging, Department of Imaging and Pathology, KU Leuven, Herestraat 49, 3000 Leuven, Belgium. [2]Nuclear Medicine, Department of Medical Imaging, Ghent University Hospital, Ghent, Belgium. [3]Leuven Biostatistics and Statistical Bioinformatics Center, KU Leuven, Leuven, Belgium. [4]Department of Oncology, Antwerp University Hospital and NETwerk Antwerpen-Waasland CoE, Edegem, Belgium. [5]Digestive Oncology, University Hospitals Leuven, Leuven, Belgium. [6]Radiopharmaceutical Research, Department

of Pharmacy and Pharmacology, KU Leuven, Leuven, Belgium. [7]Center for Oncological Research (CORE), Integrated Personalized and Precision Oncology Network (IPPON), University of Antwerp, Antwerp, Belgium. [8]Digestive Oncology, Department of Gastroenterology, Ghent University Hospital, Ghent, Belgium. [9]Nuclear Medicine, Antwerp University Hospital and Molecular Imaging and Radiology, Faculty of Medicine and Health Sciences, University of Antwerp, Wilrijk, Belgium.

Received: 28 October 2024 Accepted: 29 March 2025
Published online: 15 May 2025

References

Bidakhvidi NA, Goffin K, Dekervel J, Baete K, Nackaerts K, Clement P et al (2022) Peptide receptor radionuclide therapy targeting the somatostatin receptor: basic principles, clinical applications and optimization strategies. Cancers (Basel) 14:129

Bozkurt MF, Virgolini I, Balogova S, Beheshti M, Rubello D, Decristoforo C et al (2017) Guideline for PET/CT imaging of neuroendocrine neoplasms with 68Ga-DOTA-conjugated somatostatin receptor targeting peptides and 18F–DOPA. Eur J Nucl Med Mol Imaging 44:1588–1601

Cummings J, Kinney J (2022) Biomarkers for Alzheimer's disease: context of use, qualification, and roadmap for clinical implementation. Medicina 58:952

Deppen SA, Blume J, Bobbey AJ, Shah C, Graham MM, Lee P et al (2016) 68Ga-DOTATATE compared with 111In-DTPA-octreotide and conventional imaging for pulmonary and gastroenteropancreatic neuroendocrine tumors: a systematic review and meta-analysis. J Nucl Med 57:872–878

Fendler WP, Barrio M, Spick C, Allen-Auerbach M, Ambrosini V, Benz M et al (2017) 68Ga-DOTATATE PET/CT interobserver agreement for neuroendocrine tumor assessment: results of a prospective study on 50 patients. J Nucl Med 58:307–311

Food and Drug Administration. U.S. Department of Health and Human Services Food and Drug Administration. Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER). Biomarker Qualification: Evidentiary Framework Guidance for Industry and FDA Staff DRAFT GUIDANCE. 2018. http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm

Gwet KL (2001) Handbook of inter-rater reliability: how to measure the level of agreement between two or multiple raters. Stataxis Publishing Company, Gaithersburg

Haeger A, Soza-Ried C, Kramer V, Hurtado de Mendoza A, Eppard E, Emmanuel N et al (2023) Al[18F]F-NOTA-octreotide is comparable to [68Ga]Ga-DOTA-TATE for PET/CT imaging of neuroendocrine tumours in the Latin-American population. Cancers (Basel) 15:1–11

Hope TA, Calais J, Zhang L, Dieckmann W, Millo C (2019) 111In-pentetreotide scintigraphy versus 68Ga-DOTATATE PET: impact on krenning scores and effect of tumor burden. J Nucl Med 60:1266–1269

Hou G, Cheng X, Yang Y, Zhao D, Wang G, Zhao H et al (2024) Diagnostic performance and clinical impact of 18F-AlF-NOTA-octreotide in a large cohort of patients with neuroendocrine neoplasms: a prospective single-center study. Theranostics 14:3213–3220

Hou J, Long T, He Z, Zhou M, Yang N, Chen D et al (2021) Evaluation of 18F AlF NOTA octreotide for imaging neuroendocrine neoplasms: comparison with 68Ga-DOTATATE PET/CT. EJNMMI Res 11:55

Johnbeck CB, Knigge U, Loft A, Berthelsen AK, Mortensen J, Oturai P et al (2017) Head-to-head comparison of 64Cu-DOTATATE and 68Ga-DOTATOC PET/CT: a prospective study of 59 patients with neuroendocrine tumors. J Nucl Med 58:451–457

Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33:159–174

Lens G, AhmadiBidakhvidi N, Vandecaveye V, Grauwels S, Laenen A, Deckers W et al (2023) Intra-individual qualitative and quantitative comparison of [68Ga]Ga-DOTATATE PET/CT and PET/MRI. Ther Adv Med Oncol 15:17588359231189132

Leupe H, Ahenkorah S, Dekervel J, Unterrainer M, Van Cutsem E, Verslype C et al (2023) 18F-labeled somatostatin analogs as PET tracers for the somatostatin receptor: ready for clinical use. J Nucl Med 64:835–841

Leupe H, Pauwels E, Vandamme T, Van den Broeck B, Lybaert W, Dekervel J et al (2024) Clinical impact of using [18F]AlF-NOTA-octreotide PET/CT instead of [68Ga]Ga-DOTA-SSA PET/CT: secondary endpoint analysis of a multicenter, prospective trial. J Neuroendocrinol 36:e13420. https://doi.org/10.1111/jne.13420

Pauwels E, Cleeren F, Bormans G, Deroose CM (2018) Somatostatin receptor PET ligands—the next generation for clinical practice. Am J Nucl Med Mol Imaging 8:311–331

Pauwels E, Cleeren F, Tshibangu T, Koole M, Serdons K, Dekervel J et al (2019) Al18F-NOTA-octreotide: first comparison with 68Ga-DOTATATE in a neuroendocrine tumour patient. Eur J Nucl Med Mol Imaging 46:2398–2399

Pauwels E, Cleeren F, Tshibangu T, Koole M, Serdons K, Dekervel J et al (2020) [18F]AlF-NOTA-octreotide PET imaging: biodistribution, dosimetry and first comparison with [68Ga]Ga-DOTATATE in neuroendocrine tumour patients. Eur J Nucl Med Mol Imaging 47:3033–3046

Pauwels E, Cleeren F, Tshibangu T, Koole M, Serdons K, Boeckxstaens L et al (2023) 18F-AlF-NOTA-octreotide outperforms 68Ga-DOTA-TATE/-NOC PET in neuroendocrine tumor patients: results from a prospective, multicenter study. J Nucl Med 64:632–638

Ruf J, Schiefer J, Furth C, Kosiek O, Kropf S, Heuck F et al (2011) 68Ga-DOTATOC PET/CT of neuroendocrine tumors: spotlight on the CT phases of a triple-phase protocol. J Nucl Med 52:697–704

Tshibangu T, Cawthorne C, Serdons K, Pauwels E, Gsell W, Bormans G et al (2020) Automated GMP compliant production of [18F]AlF-NOTA-octreotide. EJNMMI Radiopharm Chem 5:1–23

Publisher's Note