

Asphericity of Tumor FDG Uptake in Non-Small Cell Lung Cancer: Reproducibility and Implications for Harmonization in Multicenter Studies

Julian MM Rogasch (✉ julian.rogasch@charite.de)

Charité-Universitätsmedizin Berlin <https://orcid.org/0000-0002-0817-6532>

Christian Furth

Charité-Universitätsmedizin Berlin

Stephanie Bluemel

Charité-Universitätsmedizin Berlin

Piotr Radojewski

Charité-Universitätsmedizin Berlin

Holger Amthauer

Charité-Universitätsmedizin Berlin

Frank Hofheinz

Helmholtz-Zentrum Dresden-Rossendorf

Original research

Keywords: FDG-PET, image reconstruction, spatial resolution, asphericity, non-small cell lung cancer, reproducibility, prognosis

Posted Date: June 19th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-35589/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at EJNMMI Research on November 2nd, 2020.

See the published version at <https://doi.org/10.1186/s13550-020-00725-y>.

Abstract

Background

Asphericity (ASP) of the primary tumor's metabolic tumor volume (MTV) in FDG-PET/CT is independently predictive for survival in patients with non-small cell lung cancer (NSCLC). However, comparability between PET systems may be limited. Therefore, reproducibility of ASP was evaluated at varying image reconstruction and acquisition times to assess feasibility of ASP assessment in multicenter studies.

Methods

Retrospective study of 50 patients with NSCLC (female, 20; median age, 69 years) undergoing pretherapeutic FDG-PET/CT (median, 3.7 MBq/kg; 180 s/bed position). Reconstruction used OSEM with TOF_{4/16} (iterations, 4; subsets, 16; in-plane filter, 2.0, 6.4 or 9.5 mm), TOF_{4/8} (4 it; 8 ss; filter, 2.0/6.0/9.5 mm), PSF + TOF_{2/17} (2 it; 17 ss; filter, 2.0/7.0/10.0 mm) or Bayesian penalized likelihood (Q.Clear; beta, 600/1750/4000). Resulting reconstructed spatial resolution (FWHM) was determined from hot sphere inserts of a NEMA IEC phantom. Data with approx. 5 mm FWHM were retrospectively smoothed to achieve 7 mm FWHM. List mode data were rebinned for acquisition times of 120/90/60 s. Threshold-based delineation of primary tumor MTV was followed by evaluation of relative ASP/SUVmax/MTV differences between datasets and resulting proportions of discordantly classified cases.

Results

Reconstructed resolution for narrow/medium/wide in-plane filter (or low/medium/high beta) was approx. 5/7/9 mm FWHM. Comparing different pairs of reconstructed resolution between TOF_{4/8}, PSF + TOF_{2/17}, Q.Clear and the reference algorithm TOF_{4/16}, ASP differences were lowest at FWHM of 7 vs. 7 mm. Proportions of discordant cases (ASP > 19.5% vs. ≤19.5%) were also lowest at 7 mm (TOF_{4/8}, 2%; PSF + TOF_{2/17}, 4%; Q.Clear, 10%). Smoothing of 5 mm data to 7 mm FWHM significantly reduced discordant cases (TOF_{4/8}, 38% reduced to 2%; PSF + TOF_{2/17}, 12–4%; Q.Clear, 10–6%) resulting in proportions comparable to original 7 mm data. Shorter acquisition time only increased proportions of discordant cases at < 90 s.

Conclusions

ASP differences were mainly determined by reconstructed spatial resolution, and multicenter studies should aim at comparable FWHM (e.g., 7 mm; determined by in-plane filter width). This reduces discordant cases (high vs. low ASP) to an acceptable proportion for TOF and PSF + TOF of < 5% (Q.Clear:

10%). Data with better resolution (i.e. lower FWHM) could be retrospectively smoothed to the desired FWHM resulting in a comparable number of discordant cases.

Background

Patients with early-stage or locally advanced non-small cell lung cancer (NSCLC) are potential candidates for curatively intended therapy; however, management decisions are primarily based on the clinical tumor stage as a single factor only (1). In the average of patients, adjuvant chemotherapy only showed modest survival benefits (2–4), and therefore, more effective methods of treatment selection are highly warranted.

Consequently, numerous additional prognostic or predictive factors (5–7), among them image-derived parameters (8–12), have been investigated aiming at more differentiated outcome prediction and more differentiated management decisions. Among parameters from positron emission tomography/computed tomography with [^{18}F]fluorodeoxyglucose (FDG-PET/CT), asphericity (ASP) is a parameter that reflects shape irregularity of the primary tumor's metabolic tumor volume (MTV), combining metric and metabolic features of the primary tumor. Three retrospective studies confirmed its independent prognostic value for progression-free (PFS) and overall survival (OS) in patients with NSCLC (13–15). The largest study (311 patients, UICC stage I-III) further showed that ASP, with an optimal cut-off of $> 19.5\%$, could identify patients with UICC stage II treated by surgery and adjuvant chemotherapy with high ASP and reduced PFS (median, 11 months vs. not reached) and OS (22 months vs. not reached) (15). ASP was superior for survival prediction compared to primary tumor's maximum standardized uptake value (SUVmax) and MTV, two other previously proposed and common PET parameters (8, 9, 16, 17).

Studies on quantitative PET parameters have mostly been monocentric, but the main limitation of any PET parameter is its dependence on numerous technical factors including image reconstruction algorithms. Therefore, results may fail to reproduce in a multicenter approach unless harmonization between centers is ensured (18–20). SUVmax and MTV may vary by $> 30\%$ if basic ordered subset expectation maximization (OSEM) reconstruction is combined with time-of-flight (TOF) information and/or scanner-specific compensation for the point spread function (PSF) (19–22).

Variability of ASP has not been investigated so far, but an impact of different reconstruction methods and resulting levels of image noise can be expected. The definition of ASP includes the MTV and its surface; therefore, a variability of MTV will cause variability of ASP. Since MTV also varies notably depending on the applied delineation algorithm (20, 23–25), there are two potential sources of variability of ASP: image generation and lesion delineation.

The goal of the current study was to investigate differences in ASP resulting from common reconstruction methods and acquisition times. The focus was on the assessment if the resulting variation is acceptable for application in multicenter studies and on defining the range of acceptable

variation of the influencing factors. To separate variability introduced by different reconstruction methods and variability caused by different delineation algorithms, all lesions were delineated using the same method as in the preceding studies on ASP in NSCLC (13–15). To facilitate interpretation, SUVmax and MTV were investigated analogously for comparison.

Methods

Phantom data

A NEMA IEC body phantom was examined using a GE Discovery MI PET scanner (GE Healthcare, General Electric, Boston, MA, USA) with a 3-ring detector with silicon photomultipliers (SiPM) and a reported sensitivity of 7.3 cps/kBq (26). Total activity in field of view was approximately 35 MBq. The absolute activities were measured in a certified dose calibrator (ISOMED 2010, MED Dresden GmbH, Germany). Sphere inserts (inner diameter, 10, 13, 17, 22, 28, and 37 mm) were filled with 24.4 kBq/ml F18-fluoride while the background was filled with 3.1 kBq/ml (sphere-to-background ratio, approx. 8:1). Acquisition time was 3 min per bed position (transaxial field of view, 70 cm; matrix size, 256×256 ; voxel size, $2.73 \times 2.73 \times 2.78 \text{ mm}^3$). CT data of the phantom were used for attenuation correction.

PET raw data were reconstructed using OSEM with time of flight (TOF; GE “VUE Point FX”) with 4 iterations and 16 subsets (i.e., $\text{TOF}_{4/16}$). This reconstruction was defined as the reference algorithm for subsequent analyses and used either a 2.0 mm, 6.4 mm or 9.5 mm in-plane Gaussian post-processing filter (i.e., $\text{TOF}_{4/16/2}$, $\text{TOF}_{4/16/6.4}$ or $\text{TOF}_{4/16/9.5}$). Further reconstruction was performed with OSEM and TOF with 4 iterations, 8 subsets and either 2.0 mm, 6.0 mm or 9.5 mm in-plane filter ($\text{TOF}_{4/8/2}$, $\text{TOF}_{4/8/6}$ or $\text{TOF}_{4/8/9.5}$).

Additionally, data were reconstructed using OSEM with TOF and point spread function (OSEM + PSF + TOF, hereafter referred to as PSF + TOF; GE “VUE Point FX” with “SharpIR”) with 2 iterations and 17 subsets and either 2.0 mm, 7.0 mm or 10.0 mm in-plane filter (PSF + $\text{TOF}_{2/17/2}$, PSF + $\text{TOF}_{2/17/7}$ or PSF + $\text{TOF}_{2/17/10}$), respectively. TOF and PSF + TOF reconstructions always included a “standard” z-axis filter.

All data were also reconstructed using Bayesian penalized likelihood reconstruction (GE “Q.Clear”) with a penalization factor β of 600, 1750 or 4000 (Q.Clear₆₀₀, Q.Clear₁₇₅₀ or Q.Clear₄₀₀₀), respectively.

Reconstructed spatial resolution was assessed as the full width at half maximum (FWHM) of the PSF in the reconstructed phantom images. PSF was modeled by a 3D Gaussian, and FWHM was determined by applying the method described in detail by Hofheinz *et al.* (27). This method is based on fitting the analytic solution for the radial activity profile of a homogeneous sphere convolved with a 3D Gaussian to the reconstructed data. In this process, the full 3D vicinity of each sphere is evaluated by transforming the data to spherical coordinates relative to the respective sphere's center. A summary of the used reconstructions, resulting spatial resolution and image noise (patient data) is given in Table 1.

Loading [MathJax]/jax/output/CommonHTML/jax.js [ig. 1.

Table 1
Reconstruction parameters and image noise

Spatial resolution	Parameters	Noise, % (median, IQR)
5 mm		
TOF _{4/8/2}	4 it, 8 ss, 2.0 mm	20.0 (17.5 to 22.8)
TOF _{4/16/2}	4 it, 16 ss, 2.0 mm	31.8 (28.5 to 37.2)
PSF + TOF _{2/17/2}	2 it, 17 ss, 2.0 mm	11.9 (10.2 to 13.2)
Q.Clear ₆₀₀	beta = 600	6.6 (5.9 to 7.6)
7 mm		
TOF _{4/8/6}	4 it, 8 ss, 6.0 mm	7.9 (6.8 to 9.1)
TOF _{4/16/6.4}	4 it, 16 ss, 6.4 mm	9.6 (8.7 to 11.4)
PSF + TOF _{2/17/7}	2 it, 17 ss, 7.0 mm	6.8 (5.7 to 7.6)
Q.Clear ₁₇₅₀	beta = 1750	4.0 (3.5 to 4.4)
9 mm		
TOF _{4/8/9.5}	4 it, 8 ss, 9.5 mm	5.2 (4.6 to 5.9)
TOF _{4/16/9.5}	4 it, 16 ss, 9.5 mm	6.5 (5.5 to 7.4)
PSF + TOF _{2/17/10}	2 it, 17 ss, 10.0 mm	5.1 (4.4 to 5.8)
Q.Clear ₄₀₀₀	beta = 4000	2.8 (2.4 to 3.4)
Reconstruction parameters for each reconstruction are displayed as well as resulting image noise in patient data (median and IQR of all 50 patients).		
it, iterations; ss, subsets		

To study effects of different acquisition time per bed position, PET list mode data were retrospectively rebinned to reconstruct further datasets representing an acquisition time of 120 s, 90 s or 60 s, respectively. Reconstruction was then performed with the algorithms that resulted in a reconstructed spatial resolution of 7 mm (i.e., TOF_{4/8/6}, TOF_{4/16/6.4}, PSF + TOF_{2/17/7} and Q.Clear₁₇₅₀).

Patients and scans

Fifty patients (female, 20; median age, 69 years; range, 46 to 83 years) with histologically proven NSCLC underwent pretherapeutic FDG-PET/CT between 07/2018 and 02/2019 using the same scanner. Imaging procedures were in accordance with the 2015 procedure guidelines for FDG-PET/CT tumor imaging by the

Loading [MathJax]/jax/output/CommonHTML/jax.js

European Association of Nuclear Medicine (EANM) (28). A median activity of 249 MBq (interquartile range [IQR], 238 to 257 MBq; range, 209 to 274 MBq) or 3.7 MBq/kg (IQR, 3.1 to 4.2 MBq/kg; range, 2.0 to 5.7 MBq/kg) was administered intravenously. PET data were acquired after a median uptake time of 65 minutes (IQR, 61 to 70 min; range, 55 to 96 min) from base of skull to the proximal femora in 3D acquisition mode (acquisition time, 180 s per bed position; bed overlap, approx. 25%). Attenuation correction was based on a non-enhanced low-dose CT (automated tube current modulation “Smart mA”; maximum tube current-time product, 100 mAs; tube voltage, 120 kV; gantry rotation time, 0.5 s) or non-enhanced diagnostic CT (maximum tube current-time product, 200 mAs).

PET raw data were reconstructed as described above (patient example in Fig. 2). Furthermore, data with 5 mm FWHM resolution were smoothed with a Gaussian filter (5 mm FWHM). According to

$$FWHM_{target}^2 = FWHM_{original}^2 + FWHM_{filter}^2$$

1

this results in a target spatial resolution of approximately 7 mm. Altogether, 25 image data per patient with different spatial resolution and noise (i.e., acquisition time) were generated.

Data evaluation

Evaluation of the data was performed with a dedicated software (ROVER, version 3.0.34, ABX advanced biochemical compounds GmbH, Radeberg, Germany) by an experienced physician in nuclear medicine. MTV of the primary tumor was delineated in each dataset using the same threshold-based, background-adapted algorithm (29). Delineation was visually inspected and manually corrected if deemed necessary. Tumoral FDG-avid tissue not related to the primary tumor and delineable from the latter (lymph nodes, metastases) was excluded. If the primary tumor was determined to be multifocal (i.e. separate ipsilateral tumor nodules) or the presence of lymphangitic carcinomatosis was diagnosed by interdisciplinary consensus, all tumor nodules and FDG-avid lymphangitic tissue were included in the MTV (see also (15)). SUVmax and ASP of the MTV were derived. Distribution of ASP values in all 50 patients is illustrated in Fig. 3.

In each dataset, a spherical volume of interest (VOI) of approx. 19 ml was placed in the unaffected right liver lobe to derive its SUVmean and SUV standard deviation and calculate image noise (SUV standard deviation / SUVmean).

Statistical analysis

Statistical analysis was performed using SPSS 22 (IBM Corporation, Armonk, NY, USA). Descriptive parameters were expressed as median and IQR. Relative differences between any dataset *a* and the reference dataset *b* were calculated as follows:

$$Relative\ difference(\%) = \frac{|a - b|}{b} \times 100\%$$

The significance of these differences was assessed with Wilcoxon signed-rank test for paired data. Proportions (%) of discordantly classified cases (high vs. low ASP/SUVmax/MTV) between algorithms were given with their 95% binomial proportion confidence intervals (95%-CI) which included the continuity correction of $\pm 0.5/n$ ($= \pm 0.5/50 = \pm 1\%$). Classification with ASP ($> 19.5\%$) was based on a previously identified cut-off in NSCLC patients (15) while cut-offs for SUVmax (> 10.5) and MTV (> 9.5 ml) were the respective median among the current 50 patients. Proportions between different pairs of algorithms were compared with two-sided McNemar's test. Statistical significance was generally assumed at $p < 0.05$.

Results

Relative differences

To identify the level of reconstructed spatial resolution that provides minimal relative ASP difference to the reference algorithm ($\text{TOF}_{4/16}$), different combinations of spatial resolution for candidate algorithms ($\text{TOF}_{4/8}$, PSF + $\text{TOF}_{2/17}$, Q.Clear) and the reference algorithm were compared pairwise (Table 2).

Table 2
Relative differences to the reference algorithm

Relative differences, % (median, IQR)					
	5 mm vs. 7 mm	7 mm vs. 7 mm	9 mm vs. 7 mm	5 mm vs. 5 mm	9 mm vs. 9 mm
ASP					
TOF _{4/8}	205 (108 to 314)	7.6 (3.1 to 18.0)	38.8 (25.1 to 55.7)	30.7 (22.7 to 42.7)	9.3 (3.7 to 29.7)
TOF _{4/16}	363 (187 to 603)	—	33.5 (23.2 to 58.9)	—	—
PSF + TOF _{2/17}	85.3 (35.8 to 162)	12.8 (6.3 to 26.9)	34.4 (21.8 to 56.9)	50.1 (36.4 to 73.5)	16.2 (4.5 to 34.1)
Q.Clear	24.7 (15.4 to 51.4)	31.3 (11.2 to 43.7)	47.8 (26.2 to 63.9)	73.1 (54.1 to 85.1)	29.1 (9.1 to 44.6)
SUVmax					
TOF _{4/8}	34.0 (25.7 to 41.3)	1.2 (0.6 to 2.0)	12.8 (10.2 to 17.2)	6.2 (3.8 to 10.4)	0.7 (0.4 to 2.0)
TOF _{4/16}	43.1 (35.4 to 53.9)	—	12.6 (9.9 to 16.4)	—	—
PSF + TOF _{2/17}	39.1 (28.6 to 46.2)	4.9 (1.9 to 7.2)	11.2 (7.5 to 13.9)	12.1 (3.3 to 20.7)	3.0 (1.6 to 5.1)
Q.Clear	17.2 (11.9 to 28.1)	5.1 (2.4 to 10.6)	11.7 (5.5 to 21.4)	14.6 (9.5 to 27.4)	7.1 (2.2 to 11.2)
MTV					
TOF _{4/8}	29.9 (21.4 to 39.8)	2.3 (1.4 to 4.4)	11.7 (6.6 to 22.0)	6.2 (3.9 to 15.9)	1.3 (0.6 to 2.8)
TOF _{4/16}	36.4 (24.9 to 50.7)	—	12.4 (6.4 to 22.4)	—	—
PSF + TOF _{2/17}	31.9 (24.2 to 39.4)	6.1 (3.1 to 12.7)	9.3 (5.3 to 17.3)	21.7 (8.4 to 41.3)	3.4 (1.5 to 7.3)
Q.Clear	14.6 (7.6 to 20.7)	6.3 (3.1 to 13.2)	10.3 (4.0 to 17.2)	30.0 (13.2 to 83.7)	10.8 (5.9 to 16.3)
Relative differences (interquartile range; IQR) are given for each algorithm relative to the reference algorithm TOF _{4/16} . Different pairs of reconstructed spatial resolution (FWHM) are compared. Missing values reflect pairs of identical datasets.					

Relative ASP differences with TOF_{4/8} and PSF + TOF_{2/17} compared to TOF_{4/16} were significantly lower at 7 vs. 7 mm than at 5 vs. 7 mm, 9 vs. 7 mm and 5 vs. 5 mm (each $p < 0.001$). In contrast, differences with Q.Clear vs. TOF_{4/16} at 7 vs. 7 mm (median, 31.3%; IQR, 11.2 to 43.7%) were similar to 9 vs. 7 mm (24.7%; 15.4 to 51.4%; $p = 0.25$). Relative ASP differences at 7 vs. 7 mm were similar to 9 vs. 9 mm with TOF_{4/8} (median, 7.6% vs. 9.3%; $p = 0.38$), PSF + TOF_{2/17} (12.8% vs. 16.2%; $p = 0.25$) and Q.Clear (31.3% vs. 29.1%; $p = 0.33$).

Relative SUVmax and MTV differences at 7 vs. 7 mm were significantly lower than corresponding ASP differences (each $p < 0.001$; Table 2).

Proportions of discordantly classified cases (original data)

The proportion of discordantly classified cases ($ASP > 19.5\%$ vs. $ASP \leq 19.5\%$) with TOF_{4/8} compared to the reference algorithm at 7 vs. 7 mm was 2% (95%-CI, 0 to 6.9%) and significantly lower than at 5 vs. 7 mm or 9 vs. 7 mm (38% and 16%, each $p < 0.05$; Table 3) but similar to 5 vs. 5 mm and 9 vs. 9 mm (6% and 2%, each $p > 0.5$).

Table 3
Discordant cases relative to the reference algorithm (ASP)

Discordant proportion, % (95%-CI)					
	5 mm vs. 7 mm	7 mm vs. 7 mm	9 mm vs. 7 mm	5 mm vs. 5 mm	9 mm vs. 9 mm
Strictly 19.5%					
TOF _{4/8}	38 (23.5 to 52.5)	2 (0 to 6.9)	16 (4.8 to 27.2)	6 (0 to 13.6)	2 (0 to 6.9)
TOF _{4/16}	44 (29.2 to 58.8)	—	14 (3.4 to 24.6)	—	—
PSF + TOF _{2/17}	12 (2.0 to 22.0)	4 (0 to 10.4)	12 (2.0 to 22.0)	32 (18.1 to 45.9)	6 (0 to 13.6)
Q.Clear	10 (0.7 to 19.3)	10 (0.7 to 19.3)	26 (12.8 to 39.2)	38 (23.5 to 52.5)	12 (2.0 to 22.0)
5% tolerance					
TOF _{4/8}	36 (21.7 to 50.3)	0 (0 to 1.0)	12 (2.0 to 22.0)	4 (0 to 10.4)	0 (0 to 1.0)
TOF _{4/16}	36 (21.7 to 50.3)	—	10 (0.7 to 19.3)	—	—
PSF + TOF _{2/17}	10 (0.7 to 19.3)	0 (0 to 1.0)	8 (0 to 16.5)	30 (16.3 to 43.7)	0 (0 to 1.0)
Q.Clear	10 (0.7 to 19.3)	6 (0 to 13.6)	22 (9.5 to 34.5)	36 (21.7 to 50.3)	10 (0.7 to 19.3)
Proportions of discordantly classified cases among all 50 patients are given in % (95%-confidence interval; 95%-CI) for each algorithm relative to the reference algorithm TOF _{4/16} . Different pairs of reconstructed spatial resolution (FWHM) are compared. Missing values reflect pairs of identical datasets. Proportions are provided either for a strict ASP cut-off (high, > 19.5%; low, ≤ 19.5%) or with 5% tolerance (i.e., ASP was also rated concordant if between 18.53% and 20.48%).					

Conversely, PSF + TOF_{2/17} showed significantly lower proportions at 7 vs. 7 mm (4%; 95%-CI, 0 to 10.4%) compared to 5 vs. 5 mm (32%, p = 0.001) while proportions were similar to 5 vs. 7 mm, 9 vs. 7 mm and 9 vs. 9 mm (12%, 12% and 6%, each p > 0.1).

Q.Clear resulted in significantly lower proportions of discordant cases at 7 vs. 7 mm (10%; 95%-CI, 0.7 to 19.3%) than at 9 vs. 7 mm and 5 vs. 5 mm (26% and 38%, each p < 0.01) while proportions were similar to 5 vs. 7 mm and 9 vs. 9 mm (10% and 12%, each p = 1.0).

Proportions at 7 vs. 7 mm were comparable between TOF_{4/8} and PSF + TOF_{2/17} (2% vs. 4%; p = 1.0) while both algorithms showed slightly less discordant cases than Q.Clear (10%; each p > 0.1).

Proportions of discordant cases at 7 vs. 7 mm were comparable between ASP, SUVmax and MTV with TOF_{4/8} (2% vs. 6% vs. 2%; each p > 0.5), PSF + TOF_{2/17} (4% vs. 0% vs. 4%; each p = 1.0) and Q.Clear (10% vs. 6% vs. 8%; each p = 1.0; supplementary table 1).

The number of discordantly classified cases tended to decrease when allowing a $\pm 5\%$ tolerance range around the ASP cut-off value (i.e., low ASP, < 20.48%; high ASP, > 18.53%; Table 3).

Relative differences and discordant cases (retrospectively smoothed data)

Comparing data that were retrospectively smoothed to achieve 7 mm reconstructed spatial resolution with the original 7 mm data, relative differences between TOF_{4/8} and the reference algorithm TOF_{4/16} were higher in retrospectively smoothed data for ASP but similar for SUVmax and MTV (details in Table 4). In contrast, relative differences with PSF + TOF_{2/17} were comparable for ASP and significantly higher in the smoothed data for SUVmax and MTV. With Q.Clear, relative differences for ASP, SUVmax and MTV were each significantly lower in the smoothed data compared to original 7 mm data.

Table 4
Relative differences to the reference algorithm: Smoothed data

	Relative differences, % (median, IQR)		P value
	Smoothed to 7 mm vs. 7 mm	Original 7 mm vs. 7 mm	
ASP			
TOF _{4/8}	13.9 (9.3 to 32.8)	7.6 (3.1 to 18.0)	0.001
TOF _{4/16}	9.1 (4.9 to 26.2)	—	—
PSF + TOF _{2/17}	17.7 (5.2 to 37.6)	12.8 (6.3 to 26.9)	0.9
Q.Clear	17.6 (5.9 to 37.2)	31.3 (11.2 to 43.7)	< 0.001
SUVmax			
TOF _{4/8}	1.6 (0.7 to 2.9)	1.2 (0.6 to 2.0)	0.08
TOF _{4/16}	1.9 (1.0 to 3.3)	—	—
PSF + TOF _{2/17}	8.9 (4.4 to 12.6)	4.9 (1.9 to 7.2)	< 0.001
Q.Clear	4.8 (1.8 to 6.6)	5.1 (2.4 to 10.6)	0.021
MTV			
TOF _{4/8}	3.2 (1.5 to 5.8)	2.3 (1.4 to 4.4)	0.18
TOF _{4/16}	3.1 (1.2 to 4.8)	—	—
PSF + TOF _{2/17}	10.6 (5.4 to 16.5)	6.1 (3.1 to 12.7)	< 0.001
Q.Clear	6.2 (3.4 to 9.8)	6.3 (3.1 to 13.2)	0.003
Relative differences (interquartile range; IQR) are given for each algorithm relative to the reference algorithm TOF _{4/16} at 7 mm FWHM (i.e., TOF _{4/16/6.4}). Differences are displayed separately for either the retrospectively smoothed data (5 mm smoothed to 7 mm FWHM) or original 7 mm data. Missing values reflect pairs of identical datasets. Significant p values are printed in bold.			

Proportions of discordantly classified cases at 7 vs. 7 mm were comparable between retrospectively smoothed data and original 7 mm data for TOF_{4/8} (smoothed vs. original, 2% vs. 2%; p = 1.0), for PSF + TOF_{2/17} (4% vs. 4%; p = 1.0) and Q.Clear (6% vs. 10%; p = 0.5). The rate of discordant cases between retrospectively smoothed data and original 7 mm data for the reference algorithm TOF_{4/16} itself was 2% (95%-CI, 0 to 6.9%).

Relative differences and discordant cases (reduced acquisition time)

Loading [MathJax]/jax/output/CommonHTML/jax.js

Relative differences in ASP, SUVmax and MTV at reconstructed spatial resolution of 7 mm (TOF_{4/8/6}, TOF_{4/16/6.4}, PSF + TOF_{2/17/7} and Q.Clear₁₇₅₀) and shorter acquisition times are displayed in supplementary tables 2 to 4. Independent from the acquisition time for the candidate algorithms, relative differences were always calculated with regard to the reference algorithm TOF_{4/16/6.4} at 180 s. Briefly, relative ASP, SUVmax and MTV differences with TOF_{4/8/6} and TOF_{4/16/6.4} were significantly higher at any shorter acquisition time (i.e., 120 s, 90 s and 60 s) than at 180 s. Relative differences with PSF_{2/17/7} tended to remain similar between 180 s and 90 s but increased significantly at 60 s. Q.Clear₁₇₅₀ mostly showed similar ASP, SUVmax and MTV differences between all acquisition times.

Proportions of discordantly classified cases of ASP, SUVmax and MTV with TOF_{4/8/6}, PSF + TOF_{2/17/7} and Q.Clear₁₇₅₀ did not increase significantly with shorter acquisition time (each compared to 180 s; supplementary tables 5 to 7). Discordant cases with TOF_{4/16/6.4} remained similar at 120 s and 90 s but increased with 60 s acquisition time (McNemar's test not applicable).

Discussion

This study found that ASP differences between reconstruction algorithms were significantly higher than corresponding SUVmax and MTV differences (Table 2). This may be explained by a combined effect of changes in SUVmax (suppression of local maxima and therefore a decreasing absolute threshold and increasing MTV size) and changes in MTV surface (smoothed, smaller MTV surface) on the ASP. Coarseness of the MTV surface is likely to differ with variation in reconstructed spatial resolution, which – in conventional iterative reconstruction algorithms – is mainly determined by the width of the in-plane post-processing filter. Therefore, if threshold-based MTV delineation is applied, wider post-processing filters can be expected to result in lower ASP. In Bayesian penalized likelihood reconstruction (e.g., GE's Q.Clear), post-processing is not applied, and smoother images are generated by increasing the penalization factor β .

However, since ASP is supposed to serve as part of prognostic/predictive models based on a predefined cut-off, even substantial inter-method differences may be clinically irrelevant if classification of individual patients into groups of high vs. low ASP remains concordant. Applying a strict cut-off for ASP of > 19.5% (15), discordantly classified cases compared to the reference algorithm accounted for 2% (TOF_{4/8}) or 4% (PSF + TOF_{2/17}) at spatial resolution of approx. 7 mm FWHM. This could be acknowledged as acceptably low for application of ASP in a multicenter study. If a less strict cut-off with $\pm 5\%$ tolerance (ASP between 18.53% and 20.48%) was applied, no discordant cases at 7 mm FWHM were observed for TOF_{4/8} and PSF + TOF_{2/17}. This underlines that inter-method ASP differences at comparable spatial resolution are clinically relevant only if ASP is close to the predefined cut-off. Furthermore, this range of tolerance is well covered by the range of possible ASP cut-offs (17–39%) within which ASP remained significantly prognostic for PFS in previously reported patients with UICC stage II NSCLC (15).

Relative differences and discordant proportions tended to be higher with Q.Clear. Notably, Q.Clear showed systematically lower image noise at any level of spatial resolution (Table 1 and Fig. 2). In contrast to conventional algorithms, relative ASP differences with Q.Clear compared to the reference algorithm were higher at 7 vs. 7 mm than at 5 vs. 7 mm (Table 2) or at 7 vs. 9 mm (*data not shown*). Simultaneously, noise levels at 5 vs. 7 mm and 7 vs. 9 mm were also more comparable to the reference algorithm than at 7 vs. 7 mm. However, the same observation was not true for SUVmax and MTV or with the conventional algorithms. Consequently, similar reconstructed spatial resolution rather than the noise level should guide the choice of reconstruction algorithms for harmonization for multicenter purposes. Furthermore, Q.Clear, or Bayesian penalized likelihood reconstruction in general, may not be optimal to achieve minimal ASP deviations if the reference is a conventional algorithm.

Variation of image noise between algorithms was especially prominent at spatial resolution of 5 mm FWHM (Table 1, Fig. 1) which partly explains high inter-method differences and discordant cases even if pairs of algorithms with 5 vs. 5 mm FWHM were compared. In addition to higher noise, Gibbs artifacts (edge elevations) caused by PSF + TOF and Q.Clear reconstruction increase with narrower in-plane filters or lower β (30). Consequently, SUVmax differences will be more prominent than at 7 mm or 9 mm FWHM. In contrast, in substantially smoothed data with 9 mm FWHM, PET parameters that are reflective of heterogeneity or irregularity of tracer accumulation, such as ASP, may lose discriminatory power to detect “real” and clinically relevant differences between tumors/patients. Therefore, 7 mm FWHM could be a feasible and reasonable target for harmonization in a multicenter approach.

If reconstructed spatial resolution is better than the target resolution (e.g., 5 mm instead of 7 mm FWHM), retrospective smoothing of data using formula (1) can be performed to achieve the anticipated resolution. This enabled inter-method differences and discordant proportions far closer to those observed with the original 7 mm data, irrespective of TOF, PSF + TOF or Q.Clear. Consequently, in a multicenter analysis, retrospective smoothing of data with better spatial resolution would be a valid option to ensure comparability. It is important to note that here the effective reconstructed spatial resolution is relevant (27) which can differ notably from the resolution determined via point sources.

A similar approach by the EANM Research Ltd. (EARL) harmonization project was reported by Kaalep *et al.* who analyzed SUV and MTV in FDG-PET data of NSCLC and lymphoma patients. Only after applying a Gaussian post-reconstruction filter of 6 to 7 mm FWHM to PET data reconstructed with PSF + TOF (compliant with the current EARL 2 standard) could SUV and MTV differences be reduced from approx. 30% to < 10% compared to reconstruction compliant with the former EARL 1 standard (31). In a different approach to harmonization, Tsutsui *et al.* examined OSEM + TOF data of a NEMA IEC phantom obtained with a Siemens Biograph mCT and showed that errors compared to a simulated reference phantom were lowest with an in-plane filter of approx. 7 to 8 mm FWHM (32). In a different study, the group achieved harmonization between 12 different PET scanners using contrast recovery (CR) of NEMA IEC phantom spheres by applying a scanner-specific Gaussian filter of up to 8 mm FWHM (33). The current results of low SUVmax differences < 5% and MTV differences \leq 6% at 7 vs. 7 mm FWHM imply that both CR and

Loading [MathJax]/jax/output/CommonHTML/jax.js able surrogates for harmonization.

Shorter acquisition times of 120 s, 90 s or 60 s increased inter-method differences compared to 180 s with TOF_{4/8/6} and TOF_{4/16/6.4} while the increase was insignificant or less prominent with PSF + TOF_{2/17/7} and Q.Clear₁₇₅₀. More importantly, proportions of discordantly classified cases by ASP, SUVmax or MTV remained similar or did not increase significantly – especially between 180 s and 90 s. Therefore, equal acquisition times between PET systems/centers may be of secondary importance to achieve comparability in the investigated parameters, and differences as high as 180 s vs. 90 s might be tolerable.

Voxel sizes may also vary between PET systems in a multicenter study. However, due to technical restrictions voxel size could not be freely varied during image reconstruction in this study. Therefore, the influence on ASP, SUVmax and MTV and the correcting effect of retrospective reslicing to the original voxel size could not be assessed. A further limitation of the current analysis is that the variation in reconstruction algorithms and acquisition time may not fully reflect differences between PET scanners beyond these factors. This would require comparative examinations with different scanners in each patient under identical conditions (20, 34). For methodological consistency with the previous studies (13–15), the same threshold-based algorithm (29) was used to delineate all lesions. Consequently, the presented results are not necessarily valid when lesions are delineated differently.

Conclusions

Differences in ASP, SUVmax and MTV resulting from TOF_{4/8}, PSF + TOF_{2/17} or Q.Clear compared to the reference algorithm TOF_{4/16} were mainly determined by differences in reconstructed spatial resolution. Therefore, harmonization for ASP in multicenter studies should aim at comparable reconstructed spatial resolution between PET systems, which is determined by either in-plane filter width or the penalization factor β . A resolution of 7 mm FWHM could ensure that discordantly classified cases of high vs. low ASP are at an acceptable proportion for TOF and PSF + TOF of < 5% (Q.Clear: 10%). Retrospectively smoothing data with better spatial resolution (i.e., lower FWHM) to the desired FWHM resulted in comparable results. These results require confirmation in a multicenter study.

List of abbreviations

95%-CI 95%-confidence interval

ASP asphericity

CR contrast recovery

EANM European Association of Nuclear Medicine

EARL EANM Research Ltd.

FDG fluorodeoxyglucose

Loading [MathJax]/jax/output/CommonHTML/jax.js

FWHM full width at half maximum

IEC International Electrotechnical Commission

it iterations

IQR interquartile range

MTV metabolic tumor volume

NEMA National Electrical Manufacturers Association

NSCLC non-small cell lung cancer

OS overall survival

OSEM ordered subset expectation maximization

PET/CT positron emission tomography / computed tomography

PFS progression-free survival

PSF point spread function

SiPM silicon photomultipliers

ss subsets

SUV standardized uptake value

TOF time of flight

VOI volume of interest

Declarations

Declarations

Ethics approval and consent to participate

All procedures were in accordance with the Charité ethics commission (vote, EA4/163/18), and informed consent was obtained from all individual participants included in the study.

Consent for publication

Loading [MathJax]/jax/output/CommonHTML/jax.js ented in Fig. 2 for publication was obtained.

Availability of data and material

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

No funding.

Authors' contributions

JMMR and FH participated in data reconstruction, analysis and interpretation as well as preparation of the manuscript. SB and PR contributed to obtaining data. HA and CF participated in data interpretation and review of the manuscript.

Acknowledgements

None.

References

1. Postmus PE, Kerr KM, Oudkerk M, Senan S, Waller DA, Vansteenkiste J, et al. Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol.* 2017;28(suppl_4):iv1–21.
2. Arriagada R, Bergman B, Dunant A, Le Chevalier T, Pignon JP, Vansteenkiste J, et al. Cisplatin-based adjuvant chemotherapy in patients with completely resected non-small-cell lung cancer. *N Engl J Med.* 2004;350(4):351–60.
3. Douillard JY, Rosell R, De Lena M, Carpagnano F, Ramlau R, Gonzales-Larriba JL, et al. Adjuvant vinorelbine plus cisplatin versus observation in patients with completely resected stage IB-IIIa non-small-cell lung cancer (Adjuvant Navelbine International Trialist Association [ANITA]): a randomised controlled trial. *Lancet Oncol.* 2006;7(9):719–27.
4. Artal Cortes A, Calera Urquizu L, Hernando Cubero J. Adjuvant chemotherapy in non-small cell lung cancer: state-of-the-art. *Transl Lung Cancer Res.* 2015;4(2):191–7.
5. Sharpnack MF, Ranbaduge N, Srivastava A, Cerciello F, Codreanu SG, Liebler DC, et al. Proteogenomic Analysis of Surgically Resected Lung Adenocarcinoma. *J Thorac Oncol.* 2018;13(10):1519–29.

6. Wang L, Dong T, Xin B, Xu C, Guo M, Zhang H, et al. Integrative nomogram of CT imaging, clinical, and hematological features for survival prediction of patients with locally advanced non-small cell lung cancer. *Eur Radiol.* 2019;29(6):2958–67.
7. Desseroit MC, Visvikis D, Tixier F, Majdoub M, Perdrisot R, Guillevin R, et al. Development of a nomogram combining clinical staging with (18)F-FDG PET/CT image features in non-small-cell lung cancer stage I-III. *Eur J Nucl Med Mol Imaging.* 2016;43(8):1477–85.
8. Liu J, Dong M, Sun X, Li W, Xing L, Yu J. Prognostic Value of 18F-FDG PET/CT in Surgical Non-Small Cell Lung Cancer: A Meta-Analysis. *PLoS One.* 2016;11(1):e0146195.
9. Paesmans M, Garcia C, Wong CY, Patz EF Jr, Komaki R, Eschmann S, et al. Primary tumour standardised uptake value is prognostic in nonsmall cell lung cancer: a multivariate pooled analysis of individual data. *Eur Respir J.* 2015;46(6):1751–61.
10. Vanhove K, Mesotten L, Heylen M, Derwael R, Louis E, Adriaensens P, et al. Prognostic value of total lesion glycolysis and metabolic active tumor volume in non-small cell lung cancer. *Cancer Treat Res Commun.* 2018;15:7–12.
11. Park S, Ha S, Lee SH, Paeng JC, Keam B, Kim TM, et al. Intratumoral heterogeneity characterized by pretreatment PET in non-small cell lung cancer patients predicts progression-free survival on EGFR tyrosine kinase inhibitor. *PLoS One.* 2018;13(1):e0189766.
12. Arshad MA, Thornton A, Lu H, Tam H, Wallitt K, Rodgers N, et al. Discovery of pre-therapy 2-deoxy-2-(18)F-fluoro-D-glucose positron emission tomography-based radiomics classifiers of survival outcome in non-small-cell lung cancer patients. *Eur J Nucl Med Mol Imaging.* 2019;46(2):455–66.
13. Apostolova I, Ego K, Steffen IG, Buchert R, Wertz H, Achenbach HJ, et al. The asphericity of the metabolic tumour volume in NSCLC: correlation with histopathology and molecular markers. *Eur J Nucl Med Mol Imaging.* 2016;43(13):2360–73.
14. Apostolova I, Rogasch J, Buchert R, Wertz H, Achenbach HJ, Schreiber J, et al. Quantitative assessment of the asphericity of pretherapeutic FDG uptake as an independent predictor of outcome in NSCLC. *BMC Cancer.* 2014;14:896.
15. Rogasch JMM, Furth C, Chibolela C, Hofheinz F, Ochsenreither S, Ruckert JC, et al. Validation of Independent Prognostic Value of Asphericity of (18)F-Fluorodeoxyglucose Uptake in Non-Small-Cell Lung Cancer Patients Undergoing Treatment With Curative Intent. *Clin Lung Cancer.* 2019.
16. Sharma A, Mohan A, Bhalla AS, Sharma MC, Vishnubhatla S, Das CJ, et al. Role of Various Metabolic Parameters Derived From Baseline 18F-FDG PET/CT as Prognostic Markers in Non-Small Cell Lung Cancer Patients Undergoing Platinum-Based Chemotherapy. *Clin Nucl Med.* 2018;43(1):e8–17.
17. Ma W, Wang M, Li X, Huang H, Zhu Y, Song X, et al. Quantitative (18)F-FDG PET analysis in survival rate prediction of patients with non-small cell lung cancer. *Oncol Lett.* 2018;16(4):4129–36.
18. Houdu B, Lasnon C, Licaj I, Thomas G, Do P, Guizard AV, et al. Why harmonization is needed when using FDG PET/CT as a prognosticator: demonstration with EARL-compliant SUV as an independent prognostic factor in lung cancer. *Eur J Nucl Med Mol Imaging.* 2019;46(2):421–8.

19. Lasnon C, Enilorac B, Popotte H, Aide N. Impact of the EARL harmonization program on automatic delineation of metabolic active tumour volumes (MATVs). *EJNMMI Res.* 2017;7(1):30.
20. Zhuang M, Garcia DV, Kramer GM, Frings V, Smit EF, Dierckx R, et al. Variability and Repeatability of Quantitative Uptake Metrics in (18)F-FDG PET/CT of Non-Small Cell Lung Cancer: Impact of Segmentation Method, Uptake Interval, and Reconstruction Protocol. *J Nucl Med.* 2019;60(5):600–7.
21. Akamatsu G, Mitsumoto K, Taniguchi T, Tsutsui Y, Baba S, Sasaki M. Influences of point-spread function and time-of-flight reconstructions on standardized uptake value of lymph node metastases in FDG-PET. *Eur J Radiol.* 2014;83(1):226–30.
22. Armstrong IS, Kelly MD, Williams HA, Matthews JC. Impact of point spread function modelling and time of flight on FDG uptake measurements in lung lesions using alternative filtering strategies. *EJNMMI Phys.* 2014;1(1):99.
23. Fleckenstein J, Hellwig D, Kremp S, Grgic A, Groschel A, Kirsch CM, et al. F-18-FDG-PET confined radiotherapy of locally advanced NSCLC with concomitant chemotherapy: results of the PET-PLAN pilot trial. *Int J Radiat Oncol Biol Phys.* 2011;81(4):e283-9.
24. Dewalle-Vignion AS, Yeni N, Petyt G, Verscheure L, Huglo D, Beron A, et al. Evaluation of PET volume segmentation methods: comparisons with expert manual delineations. *Nucl Med Commun.* 2012;33(1):34–42.
25. Nestle U, Kremp S, Schaefer-Schuler A, Sebastian-Welsch C, Hellwig D, Rube C, et al. Comparison of different methods for delineation of 18F-FDG PET-positive tissue for target volume definition in radiotherapy of patients with non-Small cell lung cancer. *J Nucl Med.* 2005;46(8):1342–8.
26. Vandendriessche D, Uribe J, Bertin H, De Geeter F. Performance characteristics of silicon photomultiplier based 15-cm AFOV TOF PET/CT. *EJNMMI Phys.* 2019;6(1):8.
27. Hofheinz F, Dittrich S, Potzsch C, Hoff J. Effects of cold sphere walls in PET phantom measurements on the volume reproducing threshold. *Phys Med Biol.* 2010;55(4):1099–113.
28. Boellaard R, Delgado-Bolton R, Oyen WJ, Giammarile F, Tatsch K, Eschner W, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging.* 2015;42(2):328–54.
29. Hofheinz F, Langner J, Petr J, Beuthien-Baumann B, Steinbach J, Kotzerke J, et al. An automatic method for accurate volume delineation of heterogeneous tumors in PET. *Med Phys.* 2013;40(8):082503.
30. Rogasch JM, Suleiman S, Hofheinz F, Bluemel S, Lukas M, Amthauer H, et al. Reconstructed spatial resolution and contrast recovery with Bayesian penalized likelihood reconstruction (Q.Clear) for FDG-PET compared to time-of-flight (TOF) with point spread function (PSF). *EJNMMI Phys.* 2020;7(1):2.
31. Kaalep A, Burggraaff CN, Pieplensbosch S, Verwer EE, Sera T, Zijlstra J, et al. Quantitative implications of the updated EARL 2019 PET-CT performance standards. *EJNMMI Phys.* 2019;6(1):28.
32. Tsutsui Y, Awamoto S, Himuro K, Umezumi Y, Baba S, Sasaki M. Characteristics of Smoothing Filters to Achieve the Guideline Recommended Positron Emission Tomography Image without Harmonization.

33. Tsutsui Y, Daisaki H, Akamatsu G, Umeda T, Ogawa M, Kajiwara H, et al. Multicentre analysis of PET SUV using vendor-neutral software: the Japanese Harmonization Technology (J-Hart) study. EJNMMI Res. 2018;8(1):83.
34. Kramer GM, Frings V, Hoetjes N, Hoekstra OS, Smit EF, de Langen AJ, et al. Repeatability of Quantitative Whole-Body 18F-FDG PET/CT Uptake Measures as Function of Uptake Interval and Lesion Selection in Non-Small Cell Lung Cancer Patients. J Nucl Med. 2016;57(9):1343–9.

Figures

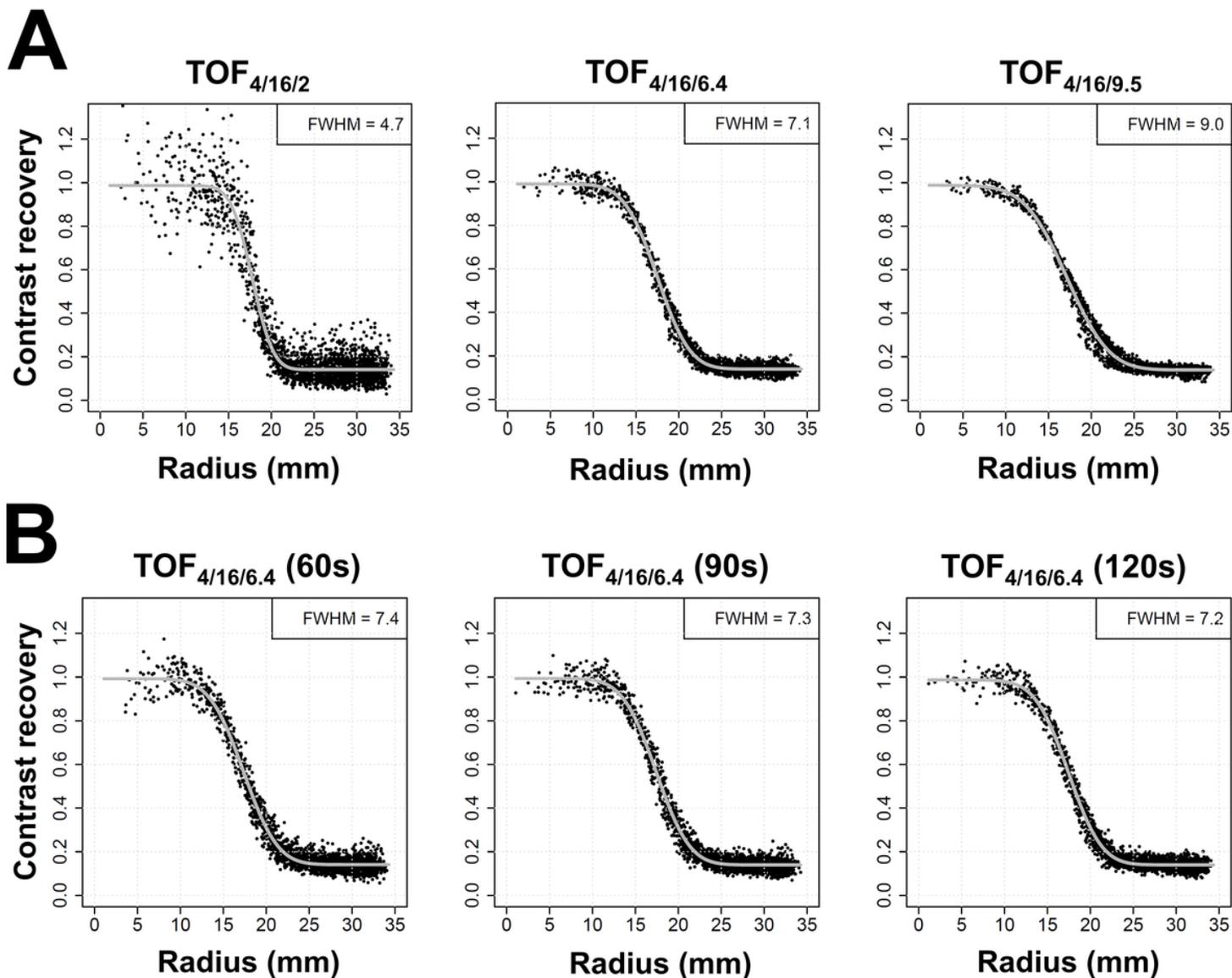


Figure 1

Sphere activity profiles A shows radial activity profiles of the 37 mm sphere for the reference algorithm with different in-plane filter width to achieve different levels of reconstructed spatial resolution (FWHM). Acquisition time was 180s. Substantial noise propagation can be observed at FWHM of approx. 5 mm. B

Loading [MathJax]/jax/output/CommonHTML/jax.js

gives corresponding profiles for 6.4 mm in-plane filter width at shorter acquisition times. Noise especially increases between 90s and 60s acquisition time while reconstructed spatial resolution remains similar.

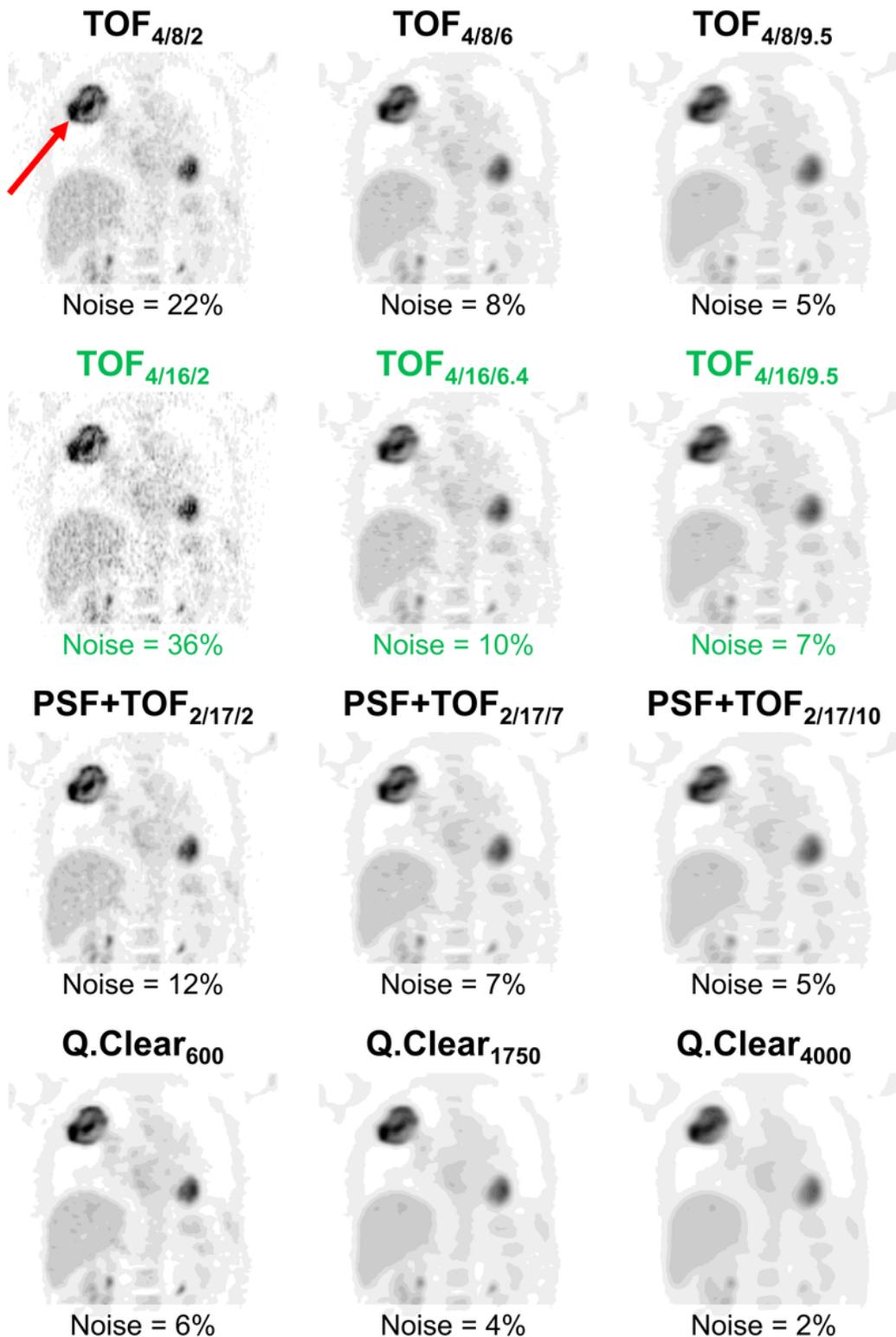


Figure 2

Patient example Coranar FDG-PET images of the thorax for a patient are displayed for all 12 reconstruction algorithms (body mass index, 22.5 kg/m²; injected activity, 3.5 MBq/kg; acquisition time, s) and 6.4 mm in-plane filter width. The image is the median of all 50 patients. Data are separated by

reconstructed spatial resolution of approx. 5 mm (left column), 7 mm (middle column) or 9 mm FWHM (right column), respectively. The reference algorithm is highlighted in green. At 7 mm FWHM spatial resolution, ASP of the primary tumor (red arrow) was concordantly high (>19.5%) with all algorithms except for Q.Clear.

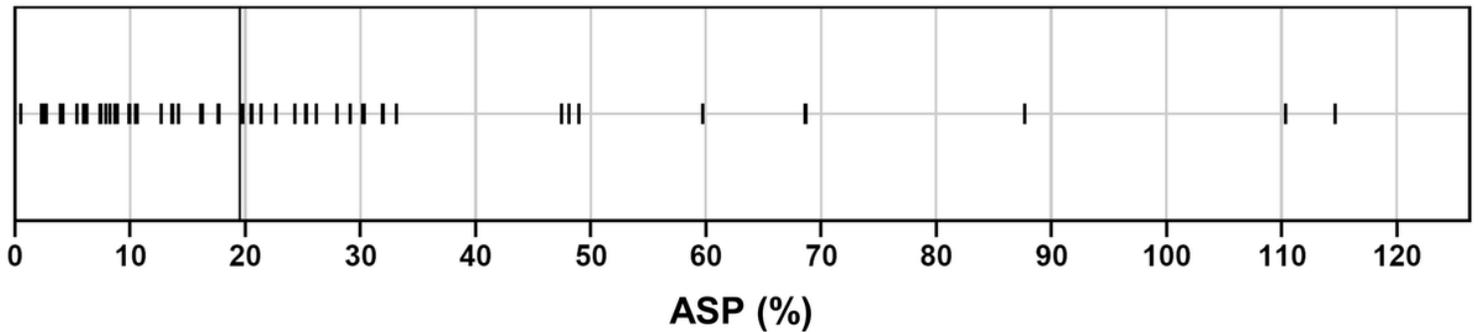


Figure 3

Distribution of ASP values with the reference algorithm The distribution of ASP values of all 50 patients for the reference algorithm at 7 mm reconstructed spatial resolution (i.e., TOF4/16/6.4) is displayed with each patient represented by a mark. The cut-off at 19.5% is highlighted. Several tumors feature ASP in proximity to this cut-off.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterial.docx](#)