The effect of reconstruction algorithms on semi-quantitative measurements in ^{18}F-FDG PET/CT imaging

Filiz Özülker MD,
Gündüzalp Buğrahan Babacan MD,
Safiya Cengiz MD,
Tamer Özülker MD

University of Health Sciences, Prof. Dr. CemilTaşcıoğlu Hospital, Department of Nuclear Medicine, İstanbul, Turkey

Keywords: Bayesian penalty likelihood - ^{18}F-FDG - Ordered subset expectation maximization - Reconstruction algorithm - Q.Clear.

Corresponding author:
Filiz Özülker MD
Çolak İsmail Sok No:25
Daire 7,
Suadiye/Kadıköy/İstanbul/Turkey
Tel: 90-506-3026157
fozulker@gmail.com

Received:
1 July 2024
Accepted revised:
20 July 2024

Abstract

Objective: This study was carried out to understand whether Q.Clear and ordered subset expectation maximization (OSEM), reconstruction algorithms used in fluorine-18-fluorodeoxyglucose (^{18}F-FDG) positron emission tomography/computed tomography (PET/CT) applications, and parameters such as time of flight (TOF) and point spread function (PSF) cause different results in semi-quantitative measurements. **Subjects and Methods:** Raw PET data of 264 patients who were referred to ^{18}F-FDG PET/CT imaging with the purpose of evaluation of known or suspicious malignant disease were reconstructed separately with Q. Clear (GE Healthcare), a BPL, an OSEM algorithm, PSF (SharpIR®) and TOF (VUE Point FX®) methods. Each patient's liver, mediastinal blood pool, metabolic tumor volume (MTV), total lesion glycolysis (TLG), and standardized uptake values (SUV) (SUVmax, SUVmean, and SUVpeak) of a total of 264 lesions selected from the patients were performed. **Results:** β350 + ToF yielded higher measurement results than all other variables for all of the lesion SUVmax, lesion SUVmean, L/AP SUVmax, and L/AP SUVmean parameters. OSEM+ToF and OSEM+TOF+PSF algorithms yielded higher mean and median SUVmax values for the reference structures (liver and mediastinum) and for lesions SUVmax and SUVmean values were statistically significantly lower than the β 350+ToF method. The method with the lowest mean value for the L/Liver SUVmax variable was OSEM+ToF 4iter16ss (mean=1.76), while the method with the highest mean value was β350+ToF (mean=2.26). β 350+ToF was the reconstruction method with the highest ratios for L/AP SUVmax and SUVmean for both lesions below and above 1cm. β 350 + ToF algorithm had also statistically significantly higher results for these variables compared to all other parameters in malignant lesions. **Conclusions:** When comparing ^{18}F-FDG PET/CT images, the use of different reconstruction algorithms may lead to misleading results, especially in the evaluation of response to treatment of malignancies.

Hell J Nucl Med 2024; 27(2): 85-92 *Epub ahead of print: 6 August 2024* *Published online: 28 August 2024*

Introduction

Fluorine-18-fluorodeoxyglucose (^{18}F-FDG) positron emission tomography/computed tomography (PET/CT) is a hybrid imaging system that enables more accurate staging of many malignant diseases and provides important contributions in the follow-up of these patients. Metabolic change in malignant tumors, rather than the size of the lesions, is widely being used as a measure of response to therapy. Qualitative evaluation of visual images, which is often performed in comparison with reference organs such as the liver and mediastinum, and semi-quantitative measurements like standardized uptake value (SUV) are being used for assessment of the efficacy of treatments [1]. Semi-quantitative measurements are also used for the metabolic characterization of lesions, but they have limited applicability owing to the lack of an agreed-upon SUV. This is partly due to technical issues that impact the precision and repeatability of SUV measurements, such as the image reconstruction techniques employed [2]. Positron emission tomography technology has advanced over the last ten years with the addition of new hardware features such as time-of-flight (TOF) collection and sophisticated point spread function (PSF) image reconstruction techniques, leading to a significant improvement in PET imaging [3].

Currently, the most popular technique for PET image reconstruction is ordered subset expectation maximization (OSEM), an expedited variation of expectation maximization (EM). In this method, the projection data are split into subgroups, or subsets, and are examined one after the other in each iteration of this process [4]. The trade-off between contrast recovery (CR) and signal-to-noise ratio (SNR) is improved by both TOF and PSF; non-

etheless, all conventional reconstruction techniques have the same fundamental drawback, which is that achieving appropriate CR by increasing the number of iterations or subsets would inevitably result in lower SNR.Iterative reconstruction techniques that are regularized have been developed recently. The block sequential regularized expectation maximization (BSREM) algorithm Q.Clear (GE Healthcare, Waukesha, WI, USA) is one reconstruction approach that does this [5]. To achieve full convergence, BSREM adds a penalty function to the probability function that simultaneously maintains edges and controls noise depending on activity. A relative difference penalty that depends on both the difference between neighboring voxels and their sum is included in Bayesian penalized likelihood (BPL). By acting as a noise suppression term, this penalty function permits more iterations without the typical noise detected in OSEM [6].

In the literature, studies mostly use phantoms to compare the effectiveness of different reconstruction algorithms [7, 8]. There were also clinical studies comparing different algorithms in terms of lesion detection [9, 10]. However, these studies are generally designed for the detection of small lesions and signal-to-noise ratio (SNR) comparisons. Another problem is the variability of lesion semi-quantitative PET parameters and reference organ SUV values independent of size in patients undergoing treatment response assessment imaging on devices using different reconstruction algorithms and there are only a few studies investigating this issue [11]. Therefore, it is considered that there is a demand for studies investigating semi-quantitative parameters in different sizes and organs similar to real clinical scenarios.

This study was carried out to understand whether the reconstruction algorithms Q.Clear and OSEM used in ^{18}F-FDG PET/CT applications and parameters such as TOF and point spread function (PSF) cause different results in semi-quantitative measurements of lesions with different sizes and organs.

Subjects and Methods

Study protocol and patient selection

Patients who were referred to Prof. Dr. Cemil Taşcıoğlu City Hospital Nuclear Medicine Clinic for ^{18}F-FDG PET/CT imaging between 24.10.2022 and 23.01.2023, were enrolled in this prospective observational study. The inclusion criteria of the patients to the study were as follows:

1. Patients who were referred to ^{18}F-FDG PET/CT imaging with the purpose of evaluation of known malignant disease or identification of malignancy of unknown primary; and patients with rheumatologic diseases, granulomatous diseases like sarcoidosis, tuberculosis, etc. for whom PET/CT is indicated by the clinician.
2. Patients greater than 18 years old.
3. Patients who underwent imaging with ^{18}F-FDG PET/CT using the standard imaging protocol, including the vertex-thigh upper 1/3 of the body.

The study was initiated after obtaining a signed informed consent form from each patient. Patients were excluded from the study if PET/CT imaging, which is routinely used in our clinic, failed to detect a lesion that could be clearly distinguished from background structures. This study, which was planned over a series of 300 patients, was completed by obtaining data from the images of 264 patients after the examinations.

Patient preparation

Fluorine-18-FDG PET/CT imaging of the patients was performed after a six-hour fasting period. According to the routine imaging protocol of the clinic, patients were given an oral contrast agent (Iohexol 350mg/mL) in 1.5 liters of water for 6 hours before imaging to ensure positive contrasting of the intestinal system. Patients with a blood glucose level of 200mg/dL or less before the injection were injected intravenously with ^{18}F-FDG at an activity of 0.09-0.14 millicurie per kilogram. Following the injection, the patients were imaged after a waiting period of approximately 60 minutes in a calm and body temperature-controlled manner. Imaging was performed with the patients positioned in the device in the supine position so that the imaging area covered the vertex-upper thigh level. Non-diagnostic low-dose CT scanning was performed for anatomic correlation and attenuation correction. Positron emission tomography imaging was then performed for the same area, with 2 minutes for each bed. General Electric Discovery MI 3 Ring PET/CT (GE Healthcare, Milwaukee, USA) was used for imaging.

Application of reconstruction algorithms

In order to apply different reconstruction parameters, raw images were used as 4 repetitions/16 subgroups and 3 repetitions/16 subgroups in different reconstruction groups in the OSEM study. Additionally, TOF (VUE Point FX®), block sequential regularized expectation maximization (BSREM) using Bayesian penalization (β factor) (VUE Point HD®), and PSF (SharpIR®) image reconstruction algorithms were used. Images were also reconstructed using 5mm post-filtering. Accordingly, seven different reconstruction parameters were determined in the following order:

1. β factor 1200 (β 1200)
2. β factor 700 + ToF (β 700+ToF) (used routinely in the clinic)
3. β factor 350 (β 350)
4. β factor 350 + ToF (β 350+ToF) (advised by the company)
5. OSEM+ ToF (4 iteration 16 subset) (OSEM+TOF4iter16ss)
6. OSEM + ToF + PSF (4 iterations 16 subset) (OSEM+ TOF+ PSF)
7. OSEM +ToF (3 iterations 16 subsets) (ToF+OSEM3iter16ss)

Among these reconstruction parameters, the use of β factor 350 + ToF (β 350+ToF) is recommended by GE Healthcare company, but β factor 700 + ToF (β 700+ToF) is routinely being used in our clinic.

After examining the images, the β factor 1200 reconstruction algorithm was excluded from the statistical analysis during metabolic tumor volume (MTV) and total lesion glycolysis (TLG) evaluation due to its inability to clearly distinguish lesions from background structures, especially in small

lesions, and inability to perform semi-automatic measurements accurately.

Obtaining data from PET/CT images

General Electric Advanced Workstation (AWS) version 3.2 was used for image evaluation. Fluorine-18-FDG PET/CT images were analyzed as maximum intensity projection (MIP) and 3D cross-sectional images. Images were analyzed after the workstation was set in a design where the same image could be viewed in 7 different windows at the same section and point at the same time (Figure 1) In our study, a single lesion was identified for each patient by examining the images of the patients by the investigators. While no algorithm was used to determine lesion localization, lesions from different regions were included in the study as much as possible to make comparisons with background structures. Patients with lesions that were not distinguishable from background structures were excluded from the study. The dimensions of the lesions were recorded from the CT component of the PET/CT images. In order to obtain the outputs of the lesions selected for MTV and TLG calculation, the lesions were localized in axial, sagittal, and coronal slices with the "auto contour" region of interest ROI method, which can automatically draw contours, and it was manually checked that all lesion borders were within the region of interest. In the calculation of lesion MTV and TLG, the SUVmax threshold value was calculated as 42%. From PET images, the SUVmax, SUVmean, MTV, and TLG parameters of these lesions and the SUVmax and SUVmean values of the background structures where the lesions are located were calculated by measuring the non-lesion part of the same tissue up to 1cm from the auto-contoured borders of the lesions and the lesion/background SUVmax and SUVmean values were calculated separately. In addition, SUVmax and SUVmean values were determined by measuring 4cm^2 (2× 2cm elliptical ROI) from the non-lesion area in the right lobe of the liver and 1cm^2 (1×1cm elliptical ROI) from the mediastinal blood pool (MBP) in the thoracic aorta.

Determination of selection characteristics of lesions

Lesions are classified as benign or malignant, according to the characteristics on PET/CT and conventional radiological imaging; or by pathology whenever it is available.

Statistical Methods

Statistical studies were performed with MedCalc version 22.016. In order to investigate the normality distribution of continuous variables, the Kolmogorov-Smirnoff test, kurtosis and skewness values, and histogram curves of the variables were examined. In order to compare the differences between dependent groups, differences (variances) were calculated and normality distributions of these results were also examined. Non-parametric tests were applied for comparisons of variables that did not show normality distribution and parametric tests were applied if normality was ensured. If there was a statistically significant difference in the main

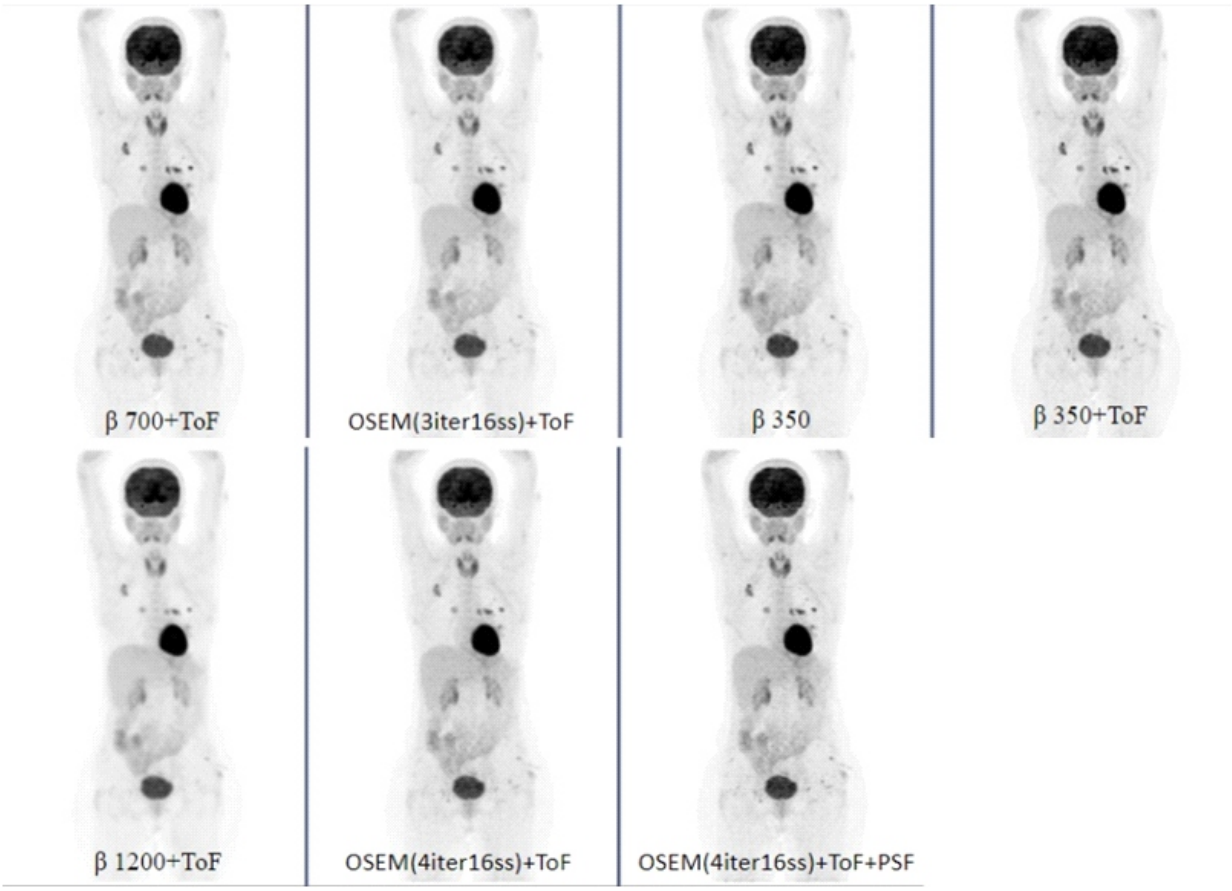

Figure 1. ^{18}F-FDG PET/CT images of a patient reconstructed by different algorithms.

group for variables in comparative analyses, paired subgroup analyses were performed to investigate this difference. Marker graphs were used to visually indicate the comparisons of the measurements of the reconstruction algorithms. In these graphs, median values were marked, and the 25th-75th percentiles were visualized as the upper and lower limits of the values of each algorithm in the graph. The MTV variable was divided into subgroups according to lesion size and shown with Dot-Plot graphs. Comparative ROC curves were drawn to investigate the threshold values of malignant lesions of the method with the highest SUVmax values among different reconstruction algorithms and the method routinely used. Youden J index was used to statistically determine the threshold value. Areas under the curve (AUC), sensitivity, and specificity values were determined. For general analyses, results with an alpha value of less than 5% were considered statistically significant, whereas Bonferroni correction was made for the comparison of variables in multiple groupings, and the P threshold value according to this result was considered statistically significant.

All procedures were performed in compliance with relevant laws and institutional guidelines and the study has been approved by the institutional review board of our hospital with the number 48670771-514.99. Signed written informed consent was obtained from all patients.

Results

In this prospective observational study, 264 patients were included. Of these patients, 156 were female and 108 were male. In the sub-analysis excluding B 1200, in which all PET parameters were evaluated, data from 252 patients were available; 151 of these patients were female and 101 were male. The mean age of these patients was 61.2 years. Descriptive statistics of the patients are summarized as mean± standard deviation and minimum-maximum (median) in Table 1.

After ^{18}F-FDG PET/CT imaging, the semi-quantitative parameters of lesions, the background structures in which the lesions were located, the physiologic areas of the MBP and liver tissues; and lesion size measurements of the lesions are performed. Wilcoxon paired two-sample test was performed between these variables and reconstruction algorithms in paired groups. Accordingly, the β1200+ ToF algorithm was found to have statistically significantly lower uptake values against all other algorithms in all variables except the liver SUVmean variable (P<0.001). Findings after this stage were made without including this method.

Comparison of semi-quantitative PET parameters of reference structures

In the comparative analysis, the OSEM+TOF4iter16ss variable was found to be statistically significantly higher than all other reconstruction algorithms with a mean value of 4.16 for liver SUVmax (P<0.001). For the same parameter, the β700+TOF reconstruction algorithm had statistically significantly lower results with a mean value of 3.04 (P<0.001).

When the comparative analysis was performed for the liver SUVmean, the reconstruction algorithm with the highest mean was β350 with a value of 2.67, while the lowest mean was β700+ToF, but there was no statistically significant difference between this algorithm and ToF 3iteration 16 ss (P=0.001-P=0.135, respectively).

In the comparison of SUVmax values obtained as a result of MBP measurements, the variable with the highest mean value was determined as OSEM+ToF4iter16ss reconstruction algorithm with a mean of 2.88, while the lowest reconstruction was determined as β700+TOF with 2.37 (P<0.001-P<0.001, respectively). In the comparison of MBP SUVmean values, the algorithm with the highest mean was β350 (mean: 2.08) (P<0.001). While the algorithm with the lowest mean was ToF+OSEM 3iter16ss, a statistically significant difference was detected between this variable and only the β350 variable (P<0.001).

Comparison of semi-quantitative PET parameters of lesions

In the analyses performed according to the lesion and lesion-background SUVmax and SUVmean, it was determined that β350 + ToF reconstruction algorithm had statistically significantly higher for the lesion SUVmax, lesion SUVmean, L/BGSUVmax and L/BGSUVmean, (means 8.6-5.16-7.56-7.88, respectively) (P<0.001) while ToF+OSEM3iter16ss algorithm had the lowest mean for lesion SUVmax and SUVmean (mean=6.60-mean=3.95) (P<0.001- P<0.001), OSEM+ToF4iter16ss algorithm had the lowest mean for L/BGSUVmax (mean=5.14) (P<0.001) and the algorithm with the lowest mean value for L/BGSUVmean was determined as ToF+OSEM3iter16ss algorithm (mean=5.52) and no statistically significant difference was detected between this algorithm and β350 and ToF+OSEM4iter16ss (P=0.916-P=0.302).

In order to evaluate all continuous variables comparatively with different reconstruction algorithms, the Friedman test was performed to investigate whether there was a statistically significant difference between repeated measurements for dependent variables. As a result of the evaluation, a statistically significant difference was found for all variables. Wilcoxon paired two-sample test was performed for statistical comparison between subgroups.

The algorithm with the highest mean value for MTV was β factor 350 (mean 12.10cm^3). In the Wilcoxon test, this algorithm showed a statistically significant difference between the other four algorithms (P<0.001), whereas no statistically significant difference was detected when compared with the ToF3iter16SS variable (P=0.829). The reconstruction algorithm found to have the lowest values for MTV analysis was the β 350+ToFalgorithm (mean 9.07cm^3) and this algorithm was found to have statistically significantly lower values compared to all other reconstruction algorithms (P< 0.001).

In the analysis of TLG, the reconstruction algorithm with the highest mean value was found to be β 350 (mean: 49.08). In the Wilcoxon test analysis performed to evaluate the pairwise comparison of this algorithm and the others for TLG, similar to the MTV, no statistically significant difference was detected with ToF3iter16SS (P=0.023) and a statistically sig-

Table 1. *^{18}F-FDG PET/CT continuous variables.*

	Reconstruction Groups						
	B1200+TOF	**OSEM+TOF**	**OSEM+TOF+PSF**	**B700+TOF**	**TOF+OSEM3iter16ss**	**B350**	**B350+TOF**
SUVmax	2,91±0,59 1,5-4,6 (2,9)	4,16±1,02 1,5-8 (4)	3,85±0,95 1,3-8,1 (3,8)	3,04±0,64 1,2-4,9 (3)	3,38±0,77 1,3-5,5 (3,4)	3,38±0,71 1,6-5,2 (3,4)	3,53±0,84 1,2-6,9 (3,5)
Liver SUVmean	2,45±0,52 1-5,4 (2,4)	2,50±0,52 1-4,5 (2,5)	2,47±0,54 1-5,4 (2,4)	2,38±0,49 1,1-4,4 (2,3)	2,39±0,50 1-4,7 (2,4)	2,55±0,54 1,4-4,3 (2,5)	2,41±0,50 1-4,5 (2,4)
Mediastinum blood pool SUVmax	2,29±0,49 1,2-4 (2,2)	2,88±0,67 1,5-5,3 (2,8)	2,78±0,66 0,4-5,9 (2,7)	2,37±0,49 1,2-4 (2,3)	2,50±0,56 0,8-5 (2,5)	2,64±0,57 1,4-4,7 (2,6)	2,66±0,58 1,4-4,5 (2,6)
Mediastinum blood pool SUVmean	2,03±0,46 0,7-3,6 (2)	1,94±0,44 0,8-3,4 (1,9)	1,94±0,44 0,2-3,3 (1,9)	1,95±0,43 0,9-3,3 (1,9)	1,92±0,44 0,7-3,3 (1,9)	2,08±0,47 1-3,7 (2)	1,95±0,44 0,9-3,3 (1,9)
Background SUVmax	1,47±1,44 0,1-14 (1,2)	1,73±1,40 0,2-14,2 (1,2)	1,63±1,42 0-15,5 (1)	1,43±1,32 0,09-15,7 (1)	1,49±1,28 0,1-14 (1,1)	1,55±1,40 0,1-16,2 (1,2)	1,58±1,47 0,1-17,5 (1,2)
Background SUVmean	1,21±1,15 0,1-10,5 (0,85)	1,10±1,00 0-9,7 (0,8)	1,09±1,02 0,04-10,1 (0,7)	1,10±1,04 0,05-10,6 (0,8)	1,11±1,04 0,05-10 (0,8)	1,15±1,09 0,08-11 (0,8)	1,08±1,04 0,04-10,6 (0,7)
Lesion SUVmax	4,87±4,97 0,4-31,7 (3,45)	6,76±4,91 0-30,7 (5,9)	7,56±5,50 1,2-34,3 (6,4)	7,14±5,77 1-36,4 (5,6)	6,60±5,03 1-32,2 (5,5)	7,08±6,23 0,7-37,8 (5,25)	8,60±6,39 1,3-40,4 (6,8)
Lesion SUVmean	3,05±3,10 0-20 (2,05)	4,03±3,07 0,5-20,3 (3,35)	4,52±3,42 0,6-21,3 (3,7)	4,34±3,64 0,6-22,4 (3,2)	3,95±3,12 0,5-20 (3,1)	4,28±3,86 0,4-23,9 (3,05)	5,16±3,95 0,5-24,2 (4)
L/BG SUVmax	4,31±5,00 0,07-41,75 (2,59)	5,14±4,48 0-42 (3,8)	6,38±6,06 0-62,33 (4,63)	6,95±7,04 1, 07-62,33 (4,61)	5,96±5,15 0,94-41,5 (4,21)	6,13±6,43 0, 43-9,57 (3,91)	7,56±6,73 1,13-52,5 (5,40)
SUVmean	3,55±4,42 0-30,7 (2)	5,56±6,00 0-47,5 (3,76)	6,69±7,41 0,92-52 (4,5)	6,17±7,02 0,8-52 (4)	5,52±6,03 0-47,5 (3,6)	5,60±6,60 0,71-46,5 (3,18)	7,88±8,65 0,94-57,5 (5,41)
Size (mm)	18,74±15,45 3-138 (13,7)	18,74±15,45 3-138 (13,7)	18,74±15,45 3-138 (13,7)	18,74±15,45 3-138 (13,7)	18,74±15,45 3-138 (13,7)	18,74±15,45 3-138 (13,7)	18,74±15,45 3-138 (13,7)

nificant difference was detected with all other algorithms (P<0.001).

In the comparison of β 350 and β 350+TOF variables, no statistically significant difference was detected between these two algorithms for the MBP SUVmax (P=0.720). However, the same situation was different in the analysis of MTV and TLG values and β 350 had the highest mean and median value for these variables, while β 350+TOF had a statistically significant low value (P<0.001).

Another finding was that the mean and median SUVmax values of the reference structures in the measurement results of the OSEM+ToF and OSEM+TOF+PSF algorithms were higher than the other algorithm, while the SUVmax and SUVmean values of the lesions were statistically significantly lower than the β 350+ToF algorithm (P<0.001).

Comparison of PET algorithms according to characterization of lesions

According to these analyses, the β 350 + ToF algorithm had statistically significantly higher results for lesion SUVmax and SUVmean, L/BG SUVmax and L/BG SUVmean compared to all other algorithms (mean benign: 5.2-3.06-5.39-5.77) and (mean malignant: 11,03-6,66-9-9,34) respectively (P< 0,001).

In the comparison of MTV and TLG values of the lesions, the results in the benign-malignant subgroups were similar to the overall group for the lowest value. For both MTV and TLG, the β 350 + ToF algorithm had statistically significantly lower outputs than all other parameters in both benign and malignant lesions (mean(benign) for MTV: 1.86-mean(malignant): 11.05cm^3), (mean(benign): 3.91-mean(malignant): 50.46 for TLG) (P<0.001-P<0.001, respectively). The reconstruction parameters with the highest values in benign and malignant groups were β 350 for MTV (mean:12.56 for malignant and mean:10.243 for benign) (malignant P=0.108 vs. benign P=0.047, respectively). For the TLG variable, while β 350 was the algorithm with the highest results in malignant lesions (mean: 60.79), ToF+OSEM 3iter 16ss was determined in benign lesions (mean: 23.58). In the Wilcoxon test performed in paired groups for statistical comparison of these data, there was no statistically significant difference between β 350 and ToF+OSEM 3iter16ss variables in malignant lesions (P=0.048), while a statistically significant difference was detected with all other algorithms (P< 0.001). In addition, the Wilcoxon test was performed to compare paired groups in the examination of the data of benign lesions and it was determined that while there was no statistically significant relationship between OSEM+ToF 4iter16ss and only OSEM+ToF 4iter16ss (P=0.005), statistically significant higher results were found compared to all other algorithms.

Comparison of PET algorithms according to the size of lesions

At last, the selected lesions were divided into subgroups according to whether they were smaller or larger than 1cm in size, and subgroup analyses were performed. For all of the SUVmax, SUVmean, L/BGSUVmax, and SUVmean variables of the lesions, β 350+ToF algorithm had higher result values than all other reconstruction algorithms in Wilcoxon paired sample comparative test results (P<0.001). These values were found to be 6.64- 4.12- 6.68-6.89 for lesions below 1cm (mean: 6.64- 4.12- 6.68-6.89, respectively) and 9.46-5.64-7.91-8.29 for lesions above 1cm (mean: 9.46-5.64-7.91-8.29, respectively). For subcentimetric lesions, a statistically significant difference was found between the β 350 reconstruction algorithm and only OSEM+ToF+PSF and β 350+ToF variables (P<0.001). In the analysis for the variables in which the ratio of lesions to background structures was calculated, β 350+ToF was the reconstruction algorithm with the highest ratios for L/BG SUVmax and SUVmean for both lesions below 1cm and above 1cm (mean L/BG SUVmax above 1cm: 7.91) (mean L/BGSUVmean above 1 cm: 8.29) (mean SUVmax below 1cm: 6.68) (mean L/BGSUVmean below 1cm: 6.89). For MTV and TLG, the reconstruction algorithm with the lowest mean and median results for both variables was determined as β 350+ToF. In the Wilcoxon test analysis, a statistically significant difference was found according to all variables (P= 0.002).

In the size-dependent analysis of volumetric variables, the parameter with the highest MTV and TLG median values was the β 350 algorithm (mean: 3.98-7.14 below 1cm, respectively) (mean: 14.37-60.79 above 1cm, respectively). In subgroup analyses, there was no statistically significant difference between this algorithm and ToF+OSEM 3iter16ss algorithm for MTV and TLG variables (P=0.693-P=0.049) and β 700, β 700+TOF and OSEM+ToF 4iter16ss for MTV variable (P= 0.018-P=0.053-P=0.009, respectively) in lesions larger than 1cm, while a statistically significant difference was detected with other variables (P<0.001). For TLG, a statistically significant difference was detected with all other variables (P= 0.002). For lesion size less than 1cm, there was no statistically significant difference between the β 350 algorithm and OSEM+ToF 4iter16ss and ToF+OSEM 3iter16ss variables, while statistically significant higher values were found with all other algorithms (P<0.001). Receiver operating characteristic curve analyses were performed for all variables for the β 350 algorithm, which has the highest SUVmax and SUVmean values among all algorithms, and for the reconstruction algorithm β 700+TOF, which is routinely used for lesion selection. For β 350, lesion SUVmax was 81.3% (AUC) and the threshold was 5.5 with 84% sensitivity and 64% specificity, while for β 700+ToF, the threshold was 5.1 with 82.1% AUC and 77% sensitivity and 72% specificity.

In the analysis performed for L/liver SUVmax and L/liver SUVmean groups formed by the ratio of lesion SUVmax and SUVmean parameters and semi-quantitative parameters of liver SUVmax and SUVmean, a statistically significant difference was detected for L/liver SUVmax by Friedman test (P< 0.00001). The algorithm with the lowest mean value for the L/liver SUVmax variable was OSEM+ToF 4iter16ss (mean= 1.76), while the algorithm with the highest mean value was β 350+ToF (mean=2.26). In the comparative analysis, a statistically significant difference was found between OSEM+ToF 4iter16ss and β 350+ToF algorithms and all other algorithms (P<0.001). For the L/liver SUVmean, the variable with the lowest mean value was OSEM+TOF+PSF (mean= 1.71). The variable with the highest mean value was β 350 +ToF (mean=2.29). In the comparative analysis, β 350 +ToF

was statistically higher than all other algorithms.

Discussion

In general, iterative reconstruction techniques in PET studies are more often employed than analytical algorithms due to their higher signal-to-noise ratios. Among these algorithms, OSEM nowadays, is the most widely used algorithm for PET image reconstruction. It depends on repeated iterations to obtain complete convergence of pictures during their reconstruction; however, as the number of iterations increases, so does the background noise. However, because the reconstruction of the images is constrained to reduce noise, the accuracy and quality of PET images are diminished because full convergence is not attained. Bayesian penalized likelihood, uses a penalty function to suppress the noise and this permits more iterations without the typical noise encountered in OSEM [4,12]. Bayesian penalized likelihood permits the achievement of effective convergence in images, possibly yielding a more accurate SUV, but also increases the SUVmax of findings and the signal-to-noise ratio [4]. In our study, β350 + ToF yielded higher measurement results than all other variables for all of the lesion SUVmax, lesion SUVmean, L/BG SUVmax, and L/BG SUVmean parameters which are concordant with the findings in the literature [3, 6,13,14]. These results suggest that, when Q Clear is used as a reconstruction algorithm, the present interpretive standards based on quantitative evaluation will need to be modified or adjusted to prevent an incorrect estimation of the illness load. In the present study, the β1200+ ToFalgorithm was found to have statistically significantly lower uptake values against all other algorithms in all variables except the liver SUVmean variable, and this raises the question of which β value will give optimal results. In a study made with phantoms, it has been shown that there is a linear correlation between β values and the overall detectability of the lesions up to a certain point (β500) beyond which a plateau is reached [15]. The optimal values of β reported are between 300 and 400 based on the signal-to-noise ratios [4]. In the present study, OSEM+ToF and OSEM+TOF+PSF algorithms yielded higher mean and median SUVmax values for the reference structures (liver and mediastinum), and for lesions SUVmax and SUVmean values were statistically significantly lower than the β 350+ToF algorithm (P<0.001). The algorithm with the lowest mean value for the L/Liver SUVmax variable was OSEM+ToF 4iter16ss (mean=1.76),while the algorithm with the highest mean value was β 350+ToF (mean=2.26). These findings imply that treatment response evaluation algorithms like the five-point scale (5-PS) Deauville, which evaluates with reference to the liver-mediastinum blood pool, may result in errors in lesion-based evaluation if the same devices are not used in comparative images, especially in diseases with heterogeneous responses and ^{18}F-FDG affinity. In a study by Subesinghe et al. (2023), performed to assess the effect of a BPL reconstruction algorithm on the 5-PS score, after interim ^{18}F-FDG PET/CT (iPET-CT) to guide treatment in classical Hodgkin's lymphoma (HL); the authors reported significantly higher 5-PS scores with BPL compared with OSEM; 34/81 (42%) cases were categorized as 5-PS score 4 with BPL, compared with 23/81 (28.3%) with OSEM. They found that, if BPL reconstruction had been utilized instead of OSEM, an intensification of chemotherapy regimen would had been indicated in 11/81 (13.6%) patients due to a change in response categorization corresponds to 25% (11/44) of patients with score 3 on OSEM [14]. In a recent study Genç et al. (2023), investigated the effect of BPL reconstruction algorithm in comparison with OSEM on quantitative parameters of ^{18}F-FDG PET/BT and Deauville 5-PS score in lymphoma patients. They reported that, SUVmax, SUVpeak, and SUVmean, measurements of the lesions were significantly higher with BPL compared to OSEM, regardless of the size (P<0.001) and in 30 patients out of 255 Deauville 5-PS increased to 4-5 which were evaluated as 1-2-3 with OSEM [16].

According to our analyses, the β 350 + ToF algorithm had statistically significantly higher results for lesion SUVmax and SUVmean, L/BG SUVmax and L/BG SUVmean variables compared to all other parameters in malignant lesions which may be accounted for the higher metabolic activities in these lesions. Matti et al. (2019) demonstrated that BPL improved the SUV values in comparison with the OSEM in each scan, especially for high SUVmax values; they reported an average variation of 22% in SUVmax values ≥10, and, 11,5 % variation when SUVmax is below 10 [4]. Murphy et al. (2019) conducted a study to assess the effect of a BPL PET reconstruction on the assessment of solitary pulmonary nodule ^{18}F-FDG uptake and estimation of malignancy risk. They found that mean nodule SUVmax was significantly higher on the images of BPL with an established penalization factor (beta) of 400, compared to OSEM for the overall cohort for malignant nodules but not for benign nodules [17].

When we investigated the performances of the reconstruction algorithms according to the size of the lesions, we found that β 350+ToF was the reconstruction algorithm with the highest ratios for L/BGSUVmax and SUVmean for both lesions below 1cm and above 1cm, which is a finding concordant with the literature. Wang et al. (2022) compared the differences between BPL and OSEM algorithms in terms of image quality and effects on clinical diagnostics and quantification of lymphoma among 70 patients. They reported that, compared with OSEM + TOF + PSF reconstruction, the BPL algorithm (β=570) can improve the diagnostic accuracy, SUVmax, SBR, and MTV of lymph nodes of less than 2cm, while there had been no impact on SUVmax of lesions that were larger than 2cm [3]. Kurita et al. (2020) evaluated the efficacy of BPL reconstruction for improving lesion conspicuity of malignant lung tumors as compared with the OSEM reconstruction incorporating the TOF model and PSF correction and they demonstrated that BPL reconstruction was superior to OSEM + TOF + PSF reconstruction in the detection of small pulmonary nodules in patients with suspected primary and metastatic lung cancer [18].

In the analysis performed for MTV and TLG variables in the subgroups made according to the lesion size, the reconstruction algorithm with the lowest mean and median results for both variables was determined as β 350+ToF. We observed the same situation in both benign and malignant lesions; in both MTV and TLG, the β 350 + ToF algorithm had statistically significantly lower outputs than all other parameters. These

findings are in concordance with the studies in the literature reporting significantly increased SUVmax, SUVmean, and SBR and decreased MTV of tumor lesions, especially in small or relatively hypometabolic lesions with the BPL algorithm reconstruction algorithm [19, 20].

This fact may be accounted for by the filtering of scattered counts and the lower uptake in the less involved parts of the lesions, consequently resulting in the sharpening of the activity in the lesions with a lower volume.

In conclusion, our study demonstrated that β350 + ToF reconstruction algorithm yielded higher SUVmax, lesion SUVmean, L/BG SUVmax, and L/BG SUVmean values, which results in an improvement in image quality of PET images but in turn might cause a necessity of reevaluating the interpretation criteria. When the higher SUVmax values obtained from the reference structures with OSEM+ToF and OSEM+TOF+PSF algorithms are taken into account; we also concluded that like the five-point scale (5-PS) Deauville, which evaluates with reference to the liver-mediastinum blood pool, may give different results when different algorithms are used. We conclude that the BPL algorithm has an advantage in detecting smaller lesions and is more effective in visualizing malignant lesions.

Bibliography

1. Kwee TC, Cheng G, Lam MG et al. SUVmax of 2.5 should not be embraced as a magic threshold for separating benign from malignant lesions. *Eur J Nucl Med Mol Imaging* 2013; 40(10): 1475-7.
2. Teoh EJ, McGowan DR, Macpherson RE et al. Phantom and Clinical Evaluation of the Bayesian Penalized Likelihood Reconstruction Algorithm Q.Clear on an LYSO PET/CT System. *J Nucl Med* 2015; 56(9): 1447-52.
3. Wang Y, Lin L, Quan W et al. Effect of Bayesian penalty likelihood algorithm on ^{18}F-FDG PET/CT image of lymphoma. *Nucl Med Commun* 2022; 43(3): 284-91.
4. Matti A, Lima GM, Pettinato C et al. How Do the More Recent Reconstruction Algorithms Affect the Interpretation Criteria of PET/CT Images? *Nucl Med Mol Imaging* 2019; 53(3): 216-22.
5. S R. Q.Clear 2014 Available from: https://www.gehealthcare.com/-/media/739d885baa59485aaef5ac0e0eeb44a4.pdf.
6. Teoh EJ, McGowan DR, Bradley KM et al. ^{18}F-FDG PET/CT assessment of histopathologically confirmed mediastinal lymph nodes in non-small cell lung cancer using a penalised likelihood reconstruction. *Eur Radiol* 2016; 26(11): 4098-106.
7. Rogasch JM, Suleiman S, Hofheinz F 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.
8. Miwa K, Wagatsuma K, Nemoto R et al. Detection of sub-centimeter lesions using digital TOF-PET/CT system combined with Bayesian penalized likelihood reconstruction algorithm. *Ann Nucl Med* 2020; 34(10): 762-71.
9. Lohaus N, Enderlin F, Skawran S et al. Impact of Bayesian penalized likelihood reconstruction on quantitative and qualitative aspects for pulmonary nodule detection in digital 2-^{18}F-FDG-PET/CT. *Sci Rep* 2022; 12(1): 8308.
10. Otani T, Hosono M, Kanagaki M et al. Evaluation and Optimization of a New PET Reconstruction Algorithm, Bayesian Penalized Likelihood Reconstruction, for Lung Cancer Assessment According to Lesion Size. *Am J Roentgenol* 2019; 213(2): W50-W6.
11. Wyrzykowski M, Siminiak N, Kazmierczak M et al. Impact of the Q.Clear reconstruction algorithm on the interpretation of PET/CT images in patients with lymphoma. *EJNMMI Res* 2020; 10(1): 99.
12. Teoh EJ, McGowan DR, Bradley KM et al. Novel penalised likelihood reconstruction of PET in the assessment of histologically verified small pulmonary nodules. *Eur Radiol* 2016; 26(2): 576-84.
13. Parvizi N, Franklin JM, McGowan DR et al. Does a novel penalized likelihood reconstruction of ^{18}F-FDG PET-CT improve signal-to-back-ground in colorectal liver metastases? *Eur J Radiol* 2015; 84(10): 1873-78.
14. Subesinghe M, Ilyas H, Dunn JT et al. The frequency of change in five-point scale score with a Bayesian penalised likelihood PET reconstruction algorithm on interim ^{18}F-FDG PET-CT and its potential implications for therapy decisions in Hodgkin's lymphoma. *Clin Radiol* 2023; 78(2): e89-e98.
15. Reynés-Llompart G, Gámez-Cenzano C, Vercher-Conejero JL et al. Phantom, clinical, and texture indices evaluation and optimization of a penalized-likelihood image reconstruction method (Q.Clear) on a BGO PET/CT scanner. *Med Phys* 2018; 45(7): 3214-22.
16. Genc M, Yildirim N, Coskun N et al. The variation of quantitative parameters and Deauville scores with different reconstruction algorithms in ^{18}F-FDG PET/CT imaging of lymphoma patients. *Rev Esp Med Nucl Imagen Mol (Engl Ed)* 2023; 42(6): 388-92.
17. Murphy DJ, Royle L, Chalampalakis Z et al. The effect of a novel Bayesian penalised likelihood PET reconstruction algorithm on the assessment of malignancy risk in solitary pulmonary nodules according to the British Thoracic Society guidelines. *Eur J Radiol* 2019; 117: 149-55.
18. Kurita Y, Ichikawa Y, Nakanishi T et al. The value of Bayesian penalized likelihood reconstruction for improving lesion conspicuity of malignant lung tumors on ^{18}F-FDG PET/CT: comparison with ordered subset expectation maximization reconstruction incorporating time-of-flight model and point spread function correction. *Ann Nucl Med* 2020; 34(4): 272-9.
19. Liu Y, Gao MJ, Zhou J et al. Changes of 1^{8}F-FDG-PET/CT quantitative parameters in tumor lesions by the Bayesian penalized-likelihood PET reconstruction algorithm and its influencing factors. *BMC Med Imaging* 2021; 21(1): 133.
20. Vallot D, Caselles O, Chaltiel L et al. A clinical evaluation of the impact of the Bayesian penalized likelihood reconstruction algorithm on PET FDG metrics. *Nucl Med Commun* 2017; 38(11): 979-84.