Open access • Journal Article • DOI:10.1097/MNM.0000000000000604

Clinical evaluation of a block sequential regularized expectation maximization reconstruction algorithm in 18F-FDG PET/CT studies. — Source link

Bert-Ram Sah, Paul Stolzmann, Gaspar Delso, Gaspar Delso ...+8 more authors

Institutions: University of Zurich, GE Healthcare

Published on: 01 Jan 2017 - Nuclear Medicine Communications (Nucl Med Commun)

Topics: Ordered subset expectation maximization

Clinical evaluation of a block sequential regularized expectation maximization reconstruction algorithm in 18F-FDG PET/CT studies

Sah, Bert-Ram ; Stolzmann, Paul ; Delso, Gaspar ; Wollenweber, Scott D ; Hüllner, Martin ; Hakami, Yahya A ; Queiroz, Marcelo A ; De Galiza Barbosa, Felipe ; von Schulthess, Gustav K ; Pietsch, Carsten ; Veit-Haibach, Patrick

Abstract: PURPOSE To investigate the clinical performance of a block sequential regularized expectation maximization (BSREM) penalized likelihood reconstruction algorithm in oncologic PET/computed tomography (CT) studies. METHODS A total of 410 reconstructions of 41 fluorine-18 fluorodeoxyglucose-PET/CT studies of 41 patients with a total of 2010 lesions were analyzed by two experienced nuclear medicine physicians. Images were reconstructed with BSREM (with four different values) or ordered subset expectation maximization (OSEM) algorithm with/without time-of-flight (TOF/non-TOF) corrections. OSEM reconstruction postfiltering was 4.0 mm full-width at half-maximum; BSREM did not use postfiltering. Evaluation of general image quality was performed with a five-point scale using maximum intensity projections. Artifacts (category 1), image sharpness (category 2), noise (category 3), and lesion detectability (category 4) were analyzed using a four-point scale. Size and maximum standardized uptake value (SUVmax) of lesions were measured by a third reader not involved in the image evaluation. RESULTS BSREM-TOF reconstructions showed the best results in all categories, independent of different body compartments. In all categories, BSREM non-TOF reconstructions were significantly better than OSEM non-TOF reconstructions ($P<0.001$). In almost all categories, BSREM non-TOF reconstruction was comparable to or better than the OSEM-TOF algorithm ($P<0.001$ for general image quality, image sharpness, noise, and $P=1.0$ for artifact). Only in lesion detectability was OSEM-TOF significantly better than BSREM non-TOF ($P<0.001$). Both BSREM-TOF and BSREM non-TOF showed a decreasing SUVmax with increasing values ($P<0.001$) and TOF reconstructions showed a significantly higher SUVmax than non-TOF reconstructions ($P<0.001$). CONCLUSION The BSREM reconstruction algorithm showed a relevant improvement compared with OSEM reconstruction in PET/CT studies in all evaluated categories. BSREM might be used in clinical routine in conjunction with TOF to achieve better/higher image quality and lesion detectability or in PET/CT-systems without TOF-capability for enhancement of overall image quality as well.

DOI: https://doi.org/10.1097/MNM.0000000000000604

Haibach, Patrick (2017). Clinical evaluation of a block sequential regularized expectation maximization reconstruction algorithm in 18F-FDG PET/CT studies. Nuclear medicine communications, 38(1):57-66. DOI: https://doi.org/10.1097/MNM.0000000000000604

Original article

Nuclear Medicine Communications

Clinical evaluation of a block sequential regularized expectation maximization reconstruction algorithm in ^{18}F-FDG PET/CT studies

Bert-Ram Sah[a,b,d], Paul Stolzmann[a,c,d], Gaspar Delso[a,e], Scott D. Wollenweber[e], Martin Hüllner[a,c,d], Yahya A. Hakami[a], Marcelo A. Queiroz[a], Felipe de Galiza Barbosa[a,d], Gustav K. von Schulthess[a,d], Carsten Pietsch[a,d] and Patrick Veit-Haibach[a,b,d]

Purpose **To investigate the clinical performance of a block sequential regularized expectation maximization (BSREM) penalized likelihood reconstruction algorithm in oncologic PET/computed tomography (CT) studies.**

Methods **A total of 410 reconstructions of 41 fluorine-18 fluorodeoxyglucose-PET/CT studies of 41 patients with a total of 2010 lesions were analyzed by two experienced nuclear medicine physicians. Images were reconstructed with BSREM (with four different β values) or ordered subset expectation maximization (OSEM) algorithm with/without time-of-flight (TOF/non-TOF) corrections. OSEM reconstruction postfiltering was 4.0 mm full-width at half-maximum; BSREM did not use postfiltering. Evaluation of general image quality was performed with a five-point scale using maximum intensity projections. Artifacts (category 1), image sharpness (category 2), noise (category 3), and lesion detectability (category 4) were analyzed using a four-point scale. Size and maximum standardized uptake value (SUV_{max}) of lesions were measured by a third reader not involved in the image evaluation.**

Results **BSREM-TOF reconstructions showed the best results in all categories, independent of different body compartments. In all categories, BSREM non-TOF reconstructions were significantly better than OSEM non-TOF reconstructions ($P < 0.001$). In almost all categories, BSREM non-TOF reconstruction was comparable to or better than the OSEM-TOF algorithm ($P < 0.001$ for general image quality, image sharpness, noise, and $P = 1.0$ for artifact). Only in lesion detectability was OSEM-TOF significantly better than BSREM non-TOF ($P < 0.001$). Both BSREM-TOF and BSREM non-TOF showed a decreasing SUV_{max} with increasing β values ($P < 0.001$) and TOF reconstructions showed a significantly higher SUV_{max} than non-TOF reconstructions ($P < 0.001$).**

Conclusion **The BSREM reconstruction algorithm showed a relevant improvement compared with OSEM reconstruction in PET/CT studies in all evaluated categories. BSREM might be used in clinical routine in conjunction with TOF to achieve better/higher image quality and lesion detectability or in PET/CT-systems without TOF-capability for enhancement of overall image quality as well.** ***Nucl Med Commun*** **00:000–000**

Nuclear Medicine Communications 2016, **00**:000–000

Keywords: image reconstruction, penalized likelihood, PET, reconstruction algorithm

Departments of [a]Nuclear Medicine, [b]Diagnostic and Interventional Radiology, [c]Neuroradiology, University Hospital Zurich, [d]University of Zurich, Zurich, Switzerland and [e]PET Clinical Science, GE Healthcare, Waukesha, Wisconsin, USA

Correspondence to Bert-Ram Sah, MD, Department of Nuclear Medicine, University Hospital of Zurich, Raemistrasse 100, CH-8091 Zurich, Switzerland
Tel: +41 44 255 11 11; fax: +41 44 255 44 14; e-mail: bert-ram.sah@usz.ch

Received 25 January 2016 **Revised** 9 August 2016
Accepted 13 September 2016

Introduction

PET is a powerful imaging device, which enables imaging and semiquantitative measurement of tracer activity *in vivo*, and thereby visualizes physiologic and pathophysiologic processes in different organs [1]. After PET/computed tomography (PET/CT) was used successfully in the primary staging of lung cancer and lymphoma, several new indications in different malignant diseases and therapy response were introduced in clinical routine [2], as well as investigation of nonmalignant diseases such as infections [3, 4]. Its main clinical indication continues to be the detection and staging of neoplastic disease [1,5].

Considerable evolutions within the last decade, such as the development of several promising tracers, and new hardware features such as time-of-flight (TOF) [6–12] were translated from research into clinical routine. Additional improvements in image quality and several technical imaging-based parameters were facilitated by advanced reconstruction methods as well [13]. Since the first PET scanners were introduced into clinical use, several different reconstruction algorithms have been used [13,14]. Analytical methods such as filtered

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website (*www.nuclearmedicinecomm.com*).

DOI: 10.1097/MNM.0000000000000604

back projection were used early on and profited from their simple robustness and low computational time and costs [15]. Later, iterative reconstructions led to an improvement of noise and artifacts [16,17]. With the considerably improved computational resources available, they were consequently introduced into clinical imaging, in PET as well as in CT [15]. In the last decade, mainly maximum likelihood reconstructions with ordered subset principles such as OSEM [5,10,16,18,19] were investigated and available on commercial scanners. OSEM algorithms successfully accelerated reconstruction processes, but are not globally or locally convergent [18]. Owing to their somewhat slow convergence in cold regions and close to hot objects [20], OSEM algorithms are challenged with increasing image noise per iteration and subset [14,17,21], especially in systems using TOF [22]. Thus, improvement of contrast with a higher number of iterations will result in higher noise [11,23]. This constitutes a limitation for clinical image reading, and for lesion quantification and lesion detection properties. Thus, the algorithm needs to be halted way before full convergence [22].

Despite their known potential for improved lesion quantification compared with OSEM algorithms, the widespread use of edge-preserving penalized-likelihood methods such as BSREM was precluded by the visual properties of the resulting images, such as blocky background noise textures, piecewise-constant appearances of organs, and relative noise strengths in high-activity and low-activity regions [20]. On the basis of different improvements, BSREM was recently identified preclinically as a useful reconstruction algorithm. For example, Asma *et al.* [20] inserted lesions with known activity into clinically acquired data sets (hybrid data sets). The authors noted promising results in terms of quantification performance, whereas visual image properties similar to OSEM could be maintained [20]. Extending this study, Ahn *et al.* [24] evaluated the quantification accuracy of the new penalized-likelihood method using phantom, hybrid, and clinical data sets. Their results confirmed the first study, showing significant improvement in BSREM in lesion quantification [24]. Another recent study from Teoh and coworkers found that such algorithms can deliver an increase in maximum standardized uptake value (SUV_{max}), signal-to-background, and signal-to-noise ratios compared with OSEM [25]. However, although the recent studies explored more quantitative approaches, only little information is available on clinical reader perception.

Therefore, the aim of our study was to evaluate and compare overall image quality, artifacts, image sharpness, noise, and lesion detectability in clinical oncological PET/CT studies reconstructed with BSREM compared to OSEM.

Methods

This single-center observational cohort study was approved by the review board of our institution, and all patients provided signed informed consent before the examinations.

Patients and image acquisition

All patients were imaged with a full-ring TOF 64-slice PET/CT scanner (Discovery PET/CT 690 VCT; GE Healthcare, Waukesha, Wisconsin, US). The PET data were acquired in the three-dimensional TOF mode with a scan duration of 2 min per bed position, an overlap of bed positions of 23%, an axial field of view of 153, and a 700 mm diameter field of view. The emission data were corrected for attenuation using the low-dose CT and iteratively reconstructed [matrix size 256 × 256, VUE Point FX (three-dimensional TOF-OSEM) with three iterations, 18 subsets] (GE Healthcare, Waukesha, Wisconsin, USA). Images were filtered in image space using an in-plane Gaussian convolution kernel with a full-width at half-maximum (FWHM) of 4.0 mm, followed by a standard axial filter with a three-slice kernel. This procedure has been used in this standard way in other studies as well [26].

Imaging studies were consecutively obtained between February and December 2012. Forty-one consecutive patients were analyzed (20 female and 21 male patients, median age: 61 years, range 38–82 years). Patients had lung cancer ($n = 13$), breast cancer ($n = 6$), head and neck cancer ($n = 5$), gastrointestinal cancer ($n = 4$), skin cancer ($n = 5$), urological cancer ($n = 2$), thyroid cancer ($n = 1$), esophageal cancer ($n = 1$), pleura mesothelioma ($n = 1$), lymphoma ($n = 1$), sarcoma ($n = 1$), and carcinoma of unknown primary ($n = 1$). Patients fasted at least 4 h before injection of tracer. Body weight, height, and blood glucose level were measured before injection of fluorine-18 fluorodeoxyglucose (^{18}F-FDG). Blood glucose level less than 8 mmol/l were accepted for imaging. Patients were intravenously administered 3–3.5 MBq of ^{18}F-FDG per kilogram of body weight [297 MBq ± 26 MBq (mean ± standard deviation), range: 238–366 MBq]. The PET scan was acquired 1-h after tracer administration.

Image processing

Raw data sets were reconstructed with 10 different reconstruction settings. In a preanalysis of the new BSREM regularization setting, a wide range of the regularization parameter β was evaluated in seven patients to define a more narrow range of β for further evaluation. The semiquantitative preanalysis with image reconstruction sets with a β of 300, 350, and 400 showed the best results with a β of 350 and 400 (data not shown) [27]. Thus, for further detailed analysis, data sets with/without TOF (non-TOF) information and a regularization setting with β of 325, 350, 375, and 400 were reconstructed using BSREM (Q-Clear), to date, a proprietary reconstruction mode of GE Healthcare. For comparison, one set using the standard OSEM method with TOF and non-TOF were reconstructed, respectively. OSEM-TOF and

Table 1 Image grading

Categories	General image quality	Artifact	Image sharpness	Noise	Lesion detectability
1	Excellent	Excellent, no artifacts	Clear, excellent images	Almost none	Very good
2	Very good	Good, some diagnostically irrelevant artifacts	Diagnostically irrelevant image blurring	Diagnostically irrelevant	Good
3	Good	Average, diagnostically relevant artifacts	Diagnostically relevant image blurring	Diagnostically relevant	Average
4	Reasonable	Inadequate, marked artifacts	Inadequate image with blurring	Marked	Poor
5	Poor	NA	NA	NA	NA

NA, not available.

non-TOF used the point spread function and three iterations, 18 subsets.

The penalized likelihood function is written as follows:

$$\hat{x} = \arg\ \max_{x \geq 0} \sum_{i=1}^{n_d} y_i\ \log[Px]_i - [Px]_i - \beta R(x),$$

where y_i represents the measured PET coincidence data, x is the image estimate, and P is the system geometry matrix, $R(x)$ is a penalty to control noise, and β controls the relative strength of the regularizing term relative to the data statistics.

The relative difference penalty, which has the advantage of providing activity dependent noise control, is then given by:

$$R(\underline{x}) = \sum\nolimits_{j=1}^{n_y} \sum_{k \in N_j} w_j w_k \frac{(x_j - x_k)^2}{(x_j + x_k) + \gamma |x_j - x_k|},$$

where w_j and w_k are the relative weights for different components of the function and γ is a tunable parameter that controls edge preservation [28].

Image evaluation

A total of 410 reconstructed PET data sets (41 patient studies with overall 10 different reconstructions) and 2010 lesions (201 lesions with 10 different reconstructions) were evaluated in random order by two experienced nuclear medicine physicians (with 6 and 7 years of experience interpreting PET/CT, respectively) blinded to the reconstruction method used. The two-reader setup was chosen to prove the reliability of quantitative image analyses. For general image quality (GIQ), data sets were viewed using maximum intensity projection of the PET and axial views for reformatted sections were performed before for PET-image quality evaluation [26]. The two readers subjectively evaluated GIQ of each PET data using a five-point scale and evaluated the criteria artifact, image sharpness (IS), noise, and lesion detectability (LD) using a four-point scale. The criteria used for these grades are summarized in Table 1 and are based on previously published studies assessing image quality [29–31]. For further analysis, lesions were grouped into compartments according to their location [28 (14%) cervical, 33 (16%) pulmonary, 75 (37%) mediastinal, 37 (18%) in the bone, 25 (12%) abdomen, and 3 (1.5%) in the limbs]. Lesions were selected independent of their size. Per patient, all suspicious PET-positive lesions up to a maximum of five lesions per compartment were chosen. If more than five lesions were present in one compartment, five target lesions were defined for further analysis, covering a range of sizes and subsegments of the compartments (e.g. different lung segments). The size and SUV_{max} of lesions were measured by a third reader not involved in the image evaluation. Size was measured in the longest distance of the lesion. Image evaluation was performed using the 'PET/CT COMPARE' algorithm of the AW Workstation, version 4.5 (GE Healthcare Biosciences, Pittsburgh, Pennsylvania, USA).

Statistical analysis

Continuous variables were expressed as means ± SD and categorical variables as qualitative parameters as frequencies (percentages).

For qualitative parameters, we compared the 10 reconstruction techniques with respect to GIQ and to four different IQ parameters (artifacts, IS, noise, LD) using the nonparametric Friedman test for multiple samples and the Wilcoxon signed-ranks test for paired samples (results of the latter one presented in the Supplementary Tables).

For quantitative parameters, SUV_{max} and lesion size were compared both among all reconstruction techniques as well as for TOF and non-TOF reconstructions among different β values separately using analysis of variances for repeated measures.

Multivariate linear regression analysis was carried out to assess independent predictors (i.e. reconstruction, lesion size, location) of quantitative parameters (i.e. LD and SUV_{max}).

Inter-reader agreement was assessed using receiver operating characteristic curves plotting reader ratings against the consensus overall diagnostic quality of the study (1 2 3 4 5 vs. 1 2). Overall diagnostic quality was rated to be adequate for diagnostic purposes if GIQ was rated by both readers with a score less than 4 and was of nondiagnostic quality if at least one reader assigned a score more than 3. Data analysis was carried out using commercially available software (SPSS statistics 21, release 21.0.0; SPSS Inc., Chicago, Illinois, USA). A P-value less than 0.05 indicated statistical significance.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or the national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This study was approved by the review board of our institution (KEK-ZH-Nr. 2010-0235).

Results

Image quality

Rating of GIQ showed significant differences between reconstructions for both readers ($P < 0.001$). For reader 1, GIQ was rated best in BSREM-TOF reconstructions compared with all other reconstructions (Table 2 for mean rating, Figs 1 and 2, Supplementary Table 1a, Supplemental digital content 1, *http://links.lww.com/NMC/A83*, for pairwise comparison and Supplementary Fig. 1a, Supplemental digital content 2, *http://links.lww.com/NMC/A84*). OSEM-TOF and BSREM non-TOF reconstructions did not show any significant differences from each other (for all β values of BSEM non-TOF).

For reader 2, GIQ was best in BSREM reconstructions and also showed significantly better results than OSEM reconstructions (Table 2 for mean rating, Supplementary Table 1b, Supplementary digital content 3, *http://links.lww.com/NMC/A85*, for pairwise comparison and Supplementary Fig. 1b, Supplemental digital content 4, *http://links.lww.com/NMC/A86*). The BSREM non-TOF reconstructions ($\beta = 325$ and 350) showed significantly better results than best BSREM-TOF reconstructions ($\beta = 375$ and 400).

Artifacts

Rating of artifacts showed significant differences between reconstructions for both readers ($P < 0.001$). For reader 1, artifacts were less prominent in BSREM-TOF reconstructions compared with all other reconstructions (Table 2 for mean rating, Fig. 3, Supplementary Table 2a, Supplemental digital content 5, *http://links.lww.com/NMC/A87* for pairwise comparison and Supplementary Fig. 2a, Supplemental digital content 6, *http://links.lww.com/NMC/A88*). BSREM non-TOF ($\beta = 325$, 350, and 400) were not rated significantly different from OSEM-TOF reconstructions.

For reader 2, artifacts were less apparent in BSREM-TOF reconstructions (Table 2 for mean rating, Supplementary Table 2b, Supplemental digital content 7, *http://links.lww.com/NMC/A89* for pairwise comparison and Supplementary Fig. 2b, Supplemental digital content 8, *http://links.lww.com/NMC/A90*); the BSREM-TOF reconstruction with a regularization parameter of $\beta = 400$ was significantly better than all the others. OSEM-TOF and BSREM non-TOF reconstructions were not rated significantly different.

Table 2 Results of different categories of image quality assessment and maximum standardized uptake value measurement

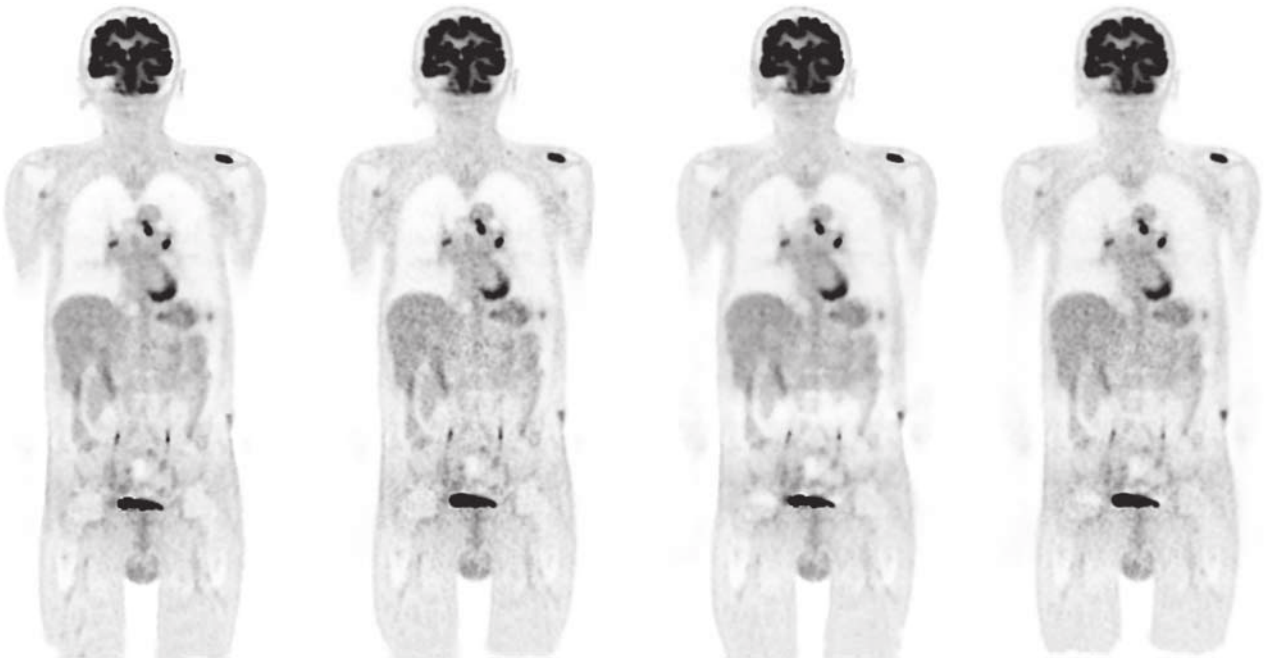
	General image quality				Artifacts				Image sharpness			
	Reader 1		Reader 2		Reader 1		Reader 2		Reader 1		Reader 2	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
BSREM-TOF 325	1.18	0.4	2.95	0.6	1.10	0.3	1.75	0.6	1.28	0.5	2.03	0.6
BSREM-TOF 350	1.18	0.4	2.58	0.6	1.10	0.3	1.73	0.6	1.28	0.5	1.73	0.6
BSREM-TOF 375	1.33	0.5	1.63	0.5	1.15	0.4	1.65	0.6	1.43	0.5	1.30	0.5
BSREM-TOF 400	1.33	0.5	1.65	0.5	1.15	0.4	1.65	0.5	1.45	0.6	1.30	0.5
OSEM-TOF	2.68	0.5	4.83	0.4	2.30	0.5	2.10	0.6	2.53	0.6	3.55	0.5
BSREM non-TOF 325	2.55	0.5	1.85	0.5	2.30	0.5	2.25	0.6	2.68	0.7	1.43	0.5
BSREM non-TOF 350	2.58	0.5	1.88	0.5	2.33	0.5	2.25	0.6	2.75	0.5	1.40	0.5
BSREM non-TOF 375	2.78	0.5	1.05	0.2	2.40	0.5	2.18	0.6	2.98	0.4	1.40	0.5
BSREM non-TOF 400	3.03	0.7	1.03	0.2	2.60	0.5	2.15	0.6	3.20	0.5	1.50	0.5
OSEM non-TOF	3.65	0.5	3.53	0.8	3.00	0.5	2.95	1.0	3.60	0.5	3.03	0.7

	Noise				Lesion detectability				SUV_{max}	
	Reader 1		Reader 2		Reader 1		Reader 2		Reader 3	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
BSREM-TOF 325	1.10	0.3	2.43	0.5	1.16	0.4	1.11	0.3	7.21	5.3
BSREM-TOF 350	1.13	0.3	1.98	0.4	1.16	0.4	1.12	0.3	7.05	5.2
BSREM-TOF 375	1.23	0.4	1.65	0.5	1.26	0.5	1.54	0.5	6.86	5.2
BSREM-TOF 400	1.25	0.4	1.38	0.5	1.28	0.5	1.58	0.5	6.72	5.1
OSEM-TOF	3.10	0.3	3.98	0.2	2.01	0.6	2.05	0.7	6.95	4.9
BSREM non-TOF 325	2.13	0.3	1.70	0.5	2.62	0.7	2.49	0.6	6.04	5.0
BSREM non-TOF 350	2.13	0.3	1.35	0.5	2.65	0.7	2.49	0.6	5.89	4.9
BSREM non-TOF 375	2.20	0.4	1.15	0.4	2.80	0.8	3.21	0.7	5.78	4.9
BSREM non-TOF 400	2.38	0.5	1.13	0.3	3.01	0.8	3.22	0.7	5.59	4.8
OSEM non-TOF	3.43	0.5	3.00	0.3	3.28	0.8	3.63	0.6	5.95	4.7

Means ± SD.

BSREM, block sequential regularized expectation maximization; non-TOF, without time-of-flight; OSEM, ordered subset expectation maximization; SUV_{max}, maximum standardized uptake value; TOF, time-of-flight.

Fig. 1



Coronal slice of PET images in four different reconstructions, showing a 64-year-old patient with disseminated soft tissue and lymph node metastases of a squamous cell carcinoma. BSREM, block sequential regularized expectation maximization; non-TOF, without time-of-flight; OSEM, ordered subset expectation maximization; TOF, time-of-flight.

Image sharpness

Rating of IS showed significant differences between reconstructions for both readers ($P < 0.001$). For reader 1, IS was better in BSREM-TOF reconstructions compared with all other reconstructions (Table 2 for mean rating, Fig. 4, Supplementary Table 3a, Supplemental digital content 9, *http://links.lww.com/NMC/A91* for pairwise comparison and Supplementary Fig. 3a, Supplemental digital content 10, *http://links.lww.com/NMC/A92*). OSEM-TOF and the BSREM non-TOF ($\beta = 325$ and 350) reconstructions did not show any significant difference.

For reader 2, IS was significantly better in BSREM reconstructions (Table 2 for mean rating, Supplementary Table 3b, Supplemental digital content 11, *http://links.lww.com/NMC/A93* for pairwise comparison and Supplementary Fig. 3b, Supplemental digital content 12, *http://links.lww.com/NMC/A94*), without significant differences between BSREM-TOF ($\beta = 375$ and 400) and BSREM non-TOF (all β values).

Noise

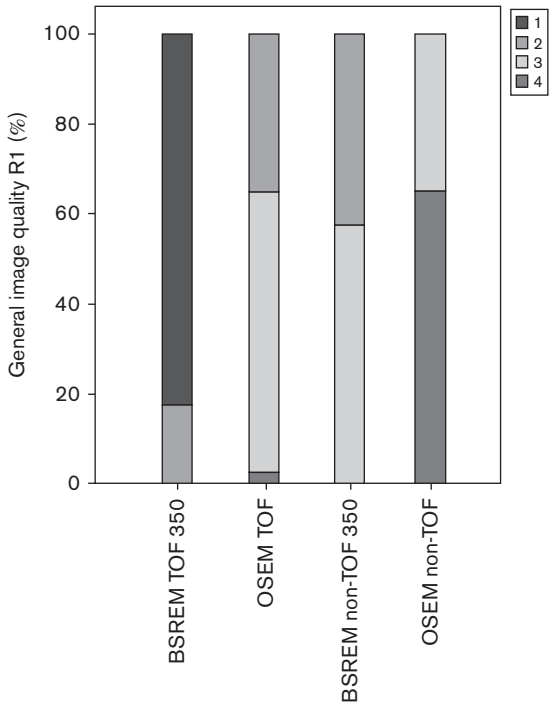
Rating of noise showed significant differences between reconstructions for both readers ($P < 0.001$). For reader 1, noise was best in BSREM-TOF reconstructions compared with all other reconstructions (Table 2 for mean rating, Fig. 5, Supplementary Table 4a, Supplemental digital content 13, *http://links.lww.com/NMC/A95* for pairwise comparison and Supplementary Fig. 4a, Supplemental digital content 14, *http://links.lww.com/NMC/A96*). BSREM non-TOF reconstructions showed significantly better results than both OSEM reconstructions.

For reader 2, noise was best in BSREM reconstructions and slightly better in BSREM non-TOF reconstructions (not significant for TOF $\beta = 400$ vs. non-TOF $\beta = 400$; Table 2 for mean rating, Supplementary Table 4b, Supplemental digital content 15, *http://links.lww.com/NMC/A97* for pairwise comparison and Supplementary Fig. 4b, Supplemental digital content 16, *http://links.lww.com/NMC/A98*).

Lesion detectability

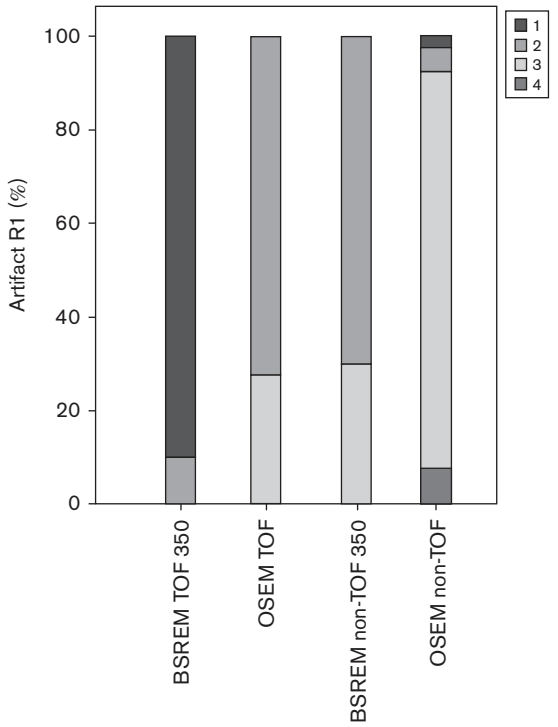
Rating of LD showed significant differences between reconstructions for both readers ($P < 0.001$). LD was significantly better in BSREM-TOF reconstructions compared with all other reconstructions (Table 2 for mean rating; Figs 6 and 7 and Supplementary Table 5a and b, Supplemental digital content 17, *http://links.lww.com/NMC/A99*, and Supplemental digital content 18, *http://links.lww.com/NMC/A100*, for pairwise comparison; and Supplementary Fig. 5a and b, Supplemental digital content 19, *http://links.lww.com/NMC/A101*, Supplemental digital content 20, *http://links.lww.com/NMC/A102*, and Supplementary Fig. 6a and b, Supplemental digital content 21, *http://links.lww.com/NMC/A103*, and Supplemental digital content 22, *http://links.lww.*

Fig. 2



General image quality. R1, reader 1; 1–5, score value 1–5 (or 1–4, see Table 1). BSREM, block sequential regularized expectation maximization; non-TOF, without time-of-flight; OSEM, ordered subset expectation maximization; TOF, time-of-flight.

Fig. 3



Artifact. For abbreviations, see Fig. 2.

com/NMC/A104). LD for OSEM-TOF was rated significantly better than the BSREM non-TOF reconstructions.

Maximum standardized uptake value

Analysis of SUV_{max} of the 201 selected lesions showed significant differences between different reconstructions ($P < 0.001$). The highest SUV_{max} per lesion was measured in BSREM-TOF and OSEM-TOF reconstructions (Table 2 for mean rating, Supplementary Table 6, Supplemental digital content 23, *http://links.lww.com/NMC/A105* for pairwise comparison and Supplementary Fig. 7, Supplemental digital content 24, *http://links.lww.com/NMC/A106*). Both BSREM-TOF and BSREM non-TOF showed a decreasing SUV_{max} with increasing β values ($P < 0.001$) and TOF reconstructions showed a significantly higher SUV_{max} than non-TOF reconstructions ($P < 0.001$).

Multivariate analysis

Multivariate analysis did not show an influence of the body compartment on the LD ($P = 0.09$ for reader 1 and $P = 0.50$ for reader 2). LD was significantly different depending on lesion size ($P < 0.001$ for both readers). SUV_{max} showed similar results and was significant depending on lesion size but not the location ($P < 0.001$ for size; $P = 0.31$ for location).

Fig. 4

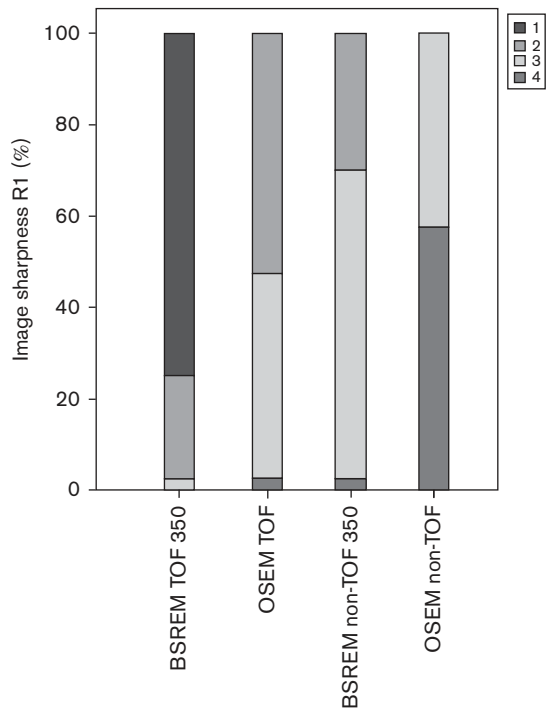
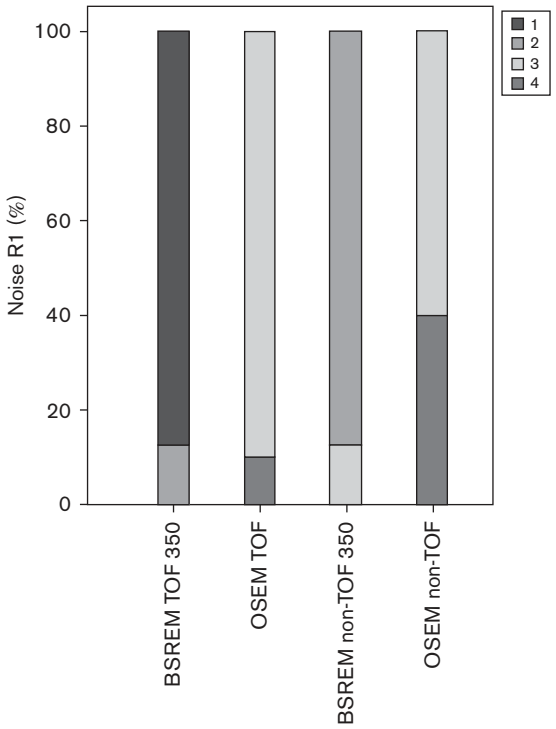


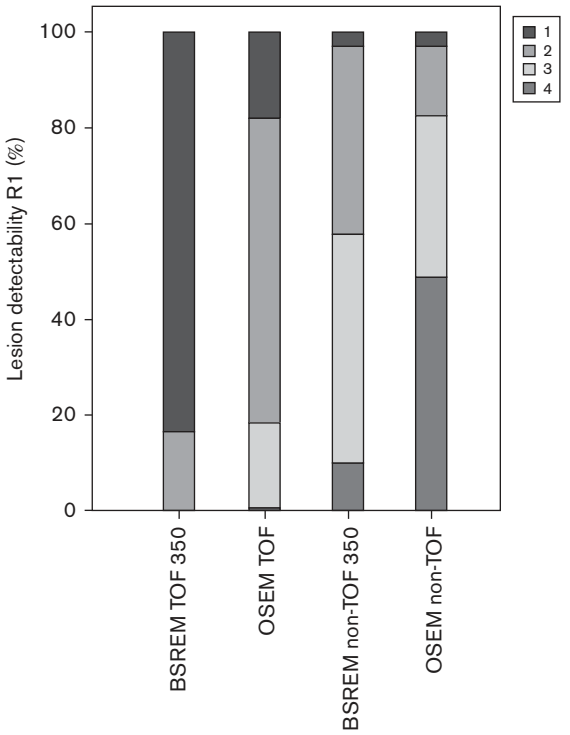
Image sharpness. For abbreviations see Fig. 2.

Fig. 5



Noise. For abbreviations see Fig. 2.

Fig. 6



Lesion detectability. For abbreviations see Fig. 2. Bars in figures show the single rated score values as percent of all rated values. Each bar represents one reconstruction method. Figures in this manuscript represent the results of reader 1 and for BSREM reconstructions with a regularization parameter of $\beta = 350$. All the results are shown in the Supplementary Figures.

Inter-reader agreement

Area under the curve (AUC) of both readers was good (>0.80 each) ($AUC_{reader\ 1} = 0.81 \pm 0.03$, $AUC_{reader\ 2} = 0.89 \pm 0.03$; Supplemental digital content 25, *http://links.lww.com/NMC/A107*).

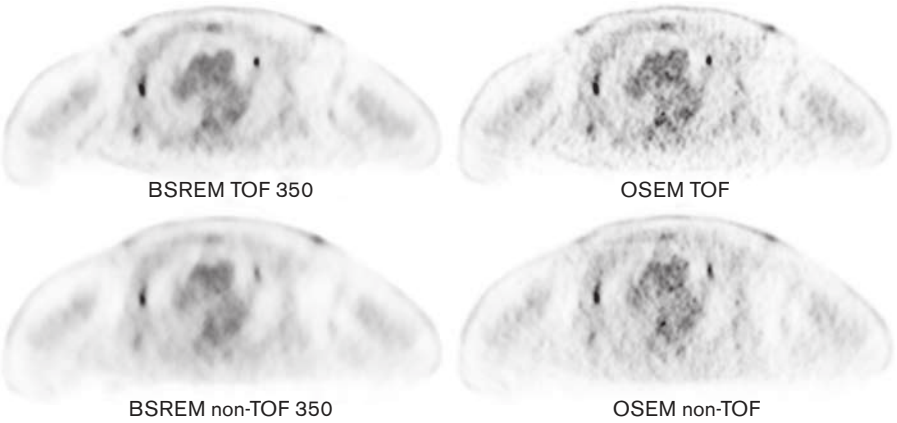
Discussion

Overall, BSREM reconstructions showed a significant improvement over the OSEM reconstruction algorithm. BSREM-TOF showed best results in almost all categories and BSREM non-TOF reconstructions showed significantly better results than those using OSEM non-TOF. Except in the LD rating (mean of 2.1 vs. 2.5 and 2.0 vs. 2.6), BSREM non-TOF reconstructions showed results comparable to or even superior than OSEM-TOF images.

The potential of penalized reconstruction methods such as BSREM in lesion evaluation and detection was published by De Pierro *et al.* [32] and Ahn *et al.* [18], who showed that BSREM is a fast and globally convergent algorithm. First evaluations of lesion quantification of this penalized-likelihood method showed promising results [20,24,25,33]. For example, it was already found that BSREM algorithms enhance the quantification accuracy of lesions [24]. Moreover, it can significantly increase SUV_{max} and increase signal-to-background/noise of lung lesions [25]. However, such technical advantages do not necessarily always translate into an obvious improvement in clinical routine. In this study, we have chosen observer performance assessment over a quantitative approach to enhance the clinical transferability. We showed that several aspects of clinical routine reading are actually enhanced with BSREM.

Despite all the improvements that OSEM reconstruction brought into PET imaging, one major disadvantage is its considerable noise, especially when combined with TOF imaging and even more so at higher numbers of iterations. Increasing the number of iterations provides a higher contrast, but at the cost of higher noise [11,23]. However, higher iterations would allow for a more accurate quantification of the standardized uptake value (SUV), which is desired in clinical imaging, especially when following up oncological patients. Exact SUV measurement is particularly important in body areas with a high background activity such as the liver parenchyma, the neck, or the mediastinum. On the basis of its technical properties and depending on the applied regularization parameter β, BSREM is expected to significantly improve this challenge by 'smoothing' the areas with higher background and at the same time emphasizing hot

Fig. 7



Axial slice of PET images, showing a patient with breast cancer and pleural metastasis in the right and a lung metastasis in the left lobe.

lesions. This was supported by a group around Parvizi *et al.* [34] who compared the properties of 42 liver metastases reconstructed with BSREM-TOF and OSEM-TOF. They reported a higher SUV_{max} of the lesion without an increased image noise using the new penalized algorithm [34]. An improvement in lesion detection is also expected especially in cold regions (e.g. lung) and close to hot objects because of OSEMs' slow convergence in these regions. This was confirmed in a recently published study by Teoh *et al.* [25]. They analyzed 121 histologically proven lung nodules in BSREM-TOF and OSEM-TOF data sets and showed a significant increase in signal-to-noise and signal-to-background measures [25]. Our study showed that lesion detection with BSREM-TOF reconstructions was rated excellent by both readers, being significantly better than OSEM-TOF. This improvement was actually found in all body compartments.

Reduction of noise also offers the possibility of further dose reduction. As shown in a study by Geismar *et al.* [35] a decreased signal-to-noise ratio in the liver parenchyma is one of the main limitations for dose reduction. Therefore, the radiation burden could potentially be reduced by using BSREM reconstructions. Besides the improvement in lesion detection with a penalized algorithm, relevant progress might also be made in non-TOF imaging. BSREM non-TOF represents a superior alternative to OSEM non-TOF and – even more importantly – showed comparable or partly better results than OSEM-TOF in general image quality as well as in terms of artifacts, image sharpness, and noise. Hence, updating PET/CT scanners non-TOF-capabilities with BSREM might therefore enhance the image quality also on these systems. Besides the expected better diagnostic image quality, upgrading scanners non-TOF-capability could prolong the life cycle of older PET/CT-systems and might therefore improve the cost effectiveness for healthcare institutions (perhaps not for the vendor of new systems, though). However, it has to be mentioned here that this was not part of the presented study as imaging data were acquired on the same machine and only different reconstruction algorithms were tested. Results for imaging data acquired on other systems may vary.

The SUV_{max} depends considerably on the reconstruction method [36]. Therefore, all clinically used systems show limitations in the number of iterations, subsets, and convergence. As discussed above, OSEM-reconstructed images are becoming somewhat noisy and finally non-diagnostic at high iteration numbers. Thus, iterations have to be stopped relatively early in the reconstruction process and as a consequence, the SUV is generally underestimated.

As expected, reconstructions using BSREM and TOF showed a higher SUV_{max} and are therefore closer to the 'true' SUV. Our results are in line with the previously published studies that provided a more quantitative evaluation compared with the study presented here [24,25,33, 34]. Differences in the SUV-values on different scanners represent a problem in imaging, both in clinical routine as well as in research studies [2,21,36]. SUV measurements are not directly comparable if patients are examined with different scanners, be it in larger institutions with several PET/CT scanners or if being referred to other hospitals. Another issue is related to PET imaging methods, in which SUV are compared with a normal database, for example, in ^{18}F-FDG brain studies. In those studies, the metabolic activity is expressed as *Z*-scores, showing a difference compared with a healthy population. In addition, such variations in SUV represent a problem in studies depending on quantification of metabolic activity, for example, in multicenter studies. Finally, consistent measurements of activity are important for defining cut-off values/thresholds for tumor-specific therapy response. Therefore, it is highly desirable to improve the reconstruction quality to achieve

an SUV measurement that is reliable and reproducible on different PET-scanners and that reflects tracer activity within tissues most realistically. On the basis of this clinical investigation, BSREM might be one important step toward arriving at true tracer activity.

Limitations

We did not investigate the clinical significance of our results, for example, whether reader confidence was improved. Also, it was not tested whether more lesions would have been detected with/out BSREM reconstructed images. One of the reasons is that we chose patients who already presented an advanced stage of disease and thus, there might be some selection bias in our cohort. However, several other parameters, which are not just important in the evaluation of malignant lesions, were also tested here. Analysis of SUV was restricted to measurement of SUV_{max}, which is the main parameter, assessed in our clinical care. Further evaluation of corrected SUV-values would possibly increase accuracy. Furthermore, lesions were evaluated by two readers; however, more readers might have balanced out personal preferences of reading. The power of observer assessment is limited as only two readers participated in the readout. However, for both readers, the majority of the results point toward the direction of improvement in image quality parameters on the basis of BSREM-reconstruction. A clinical reader assessment like the one presented in here might always be influenced by some bias as readers are not totally blind to the 'appearance' of different reconstruction algorithms. However, in our study, the readers were blinded to 10 different reconstruction sets (eight with BSREM und two with OSEM), which minimizes this bias. Finally, this analysis was carried out on oncology whole-body ^{18}F-FDG PET examinations and results may vary depending on the tracer and the imaging technique.

Conclusion

The BSREM reconstruction algorithm shows relevant improvement in image quality compared with OSEM reconstruction in PET/CT studies. BSREM-TOF reconstruction showed improved results in lesion detection, independent of the body region. Furthermore, BSREM non-TOF offers comparable or even better results than OSEM-TOF in GIQ, IS, noise, and artifacts. Upgrading PET/CT-systems with BSREM reconstruction capability could enhance the image quality even in older systems without the need to purchase a new scanner. According to our preanalysis and recently published results [24,33], the regularization parameter β should generally be between 350 and 400. Both BSREM-TOF and BSREM non-TOF showed a decreasing SUV_{max} with increasing β values and TOF reconstructions showed a significantly higher SUV_{max} than non-TOF reconstructions.

Acknowledgements

Financial support was provided by an institutional grant from GE Healthcare.

Conflicts of interest

Patrick Veit-Haibach received investigator-initiated study grants from Bayer Healthcare, Siemens Healthcare, Roche Pharmaceuticals, and GE Healthcare, and speaker's fees from GE Healthcare. Gustav von Schulthess is a grant recipient from GE Healthcare. For the remaining authors there are no conflicts of interest.

References

1 Hess S, Blomberg BA, Zhu HJ, Hoilund-Carlsen PF, Alavi A. The pivotal role of FDG-PET/CT in modern medicine. *Acad Radiol* 2014; **21**:232–249.
2 Kinahan PE, Fletcher JW. Positron emission tomography-computed tomography standardized uptake values in clinical practice and assessing response to therapy. *Semin Ultrasound CT MR* 2010; **31**:496–505.
3 Keidar Z, Pirmisashvili N, Leiderman M, Nitecki S, Israel O. ^{18}F-FDG uptake in noninfected prosthetic vascular grafts: incidence, patterns, and changes over time. *J Nucl Med* 2014; **55**:392–395.
4 Sah BR, Husmann L, Mayer D, *et al.* Diagnostic performance of ^{18}F-FDG-PET/CT in vascular graft infections. *Eur J Vasc Endovasc Surg* 2015; **49**:455–464.
5 Qi J, Huesman RH. Penalized maximum-likelihood image reconstruction for lesion detection. *Phys Med Biol* 2006; **51**:4017–4029.
6 El Fakhri G, Surti S, Trott CM, Scheuermann J, Karp JS. Improvement in lesion detection with whole-body oncologic time-of-flight PET. *J Nucl Med* 2011; **52**:347–353.
7 Karp JS, Surti S, Daube-Witherspoon ME, Muehllehner G. Benefit of time-of-flight in PET: experimental and clinical results. *J Nucl Med* 2008; **49**:462–470.
8 Surti S. Update on time-of-flight PET imaging. *J Nucl Med* 2015; **56**:98–105.
9 Surti S, Scheuermann J, El Fakhri G, *et al.* Impact of time-of-flight PET on whole-body oncologic studies: a human observer lesion detection and localization study. *J Nucl Med* 2011; **52**:712–719.
10 Akamatsu G, Ishikawa K, Mitsumoto K, *et al.* Improvement in PET/CT image quality with a combination of point-spread function and time-of-flight in relation to reconstruction parameters. *J Nucl Med* 2012; **53**:1716–1722.
11 Lois C, Jakoby BW, Long MJ, *et al.* An assessment of the impact of incorporating time-of-flight information into clinical PET/CT imaging. *J Nucl Med* 2010; **51**:237–245.
12 Kadrmas DJ, Casey ME, Conti M, Jakoby BW, Lois C, Townsend DW. Impact of time-of-flight on PET tumor detection. *J Nucl Med* 2009; **50**:1315–1323.
13 Vandenberghe S, D'Asseler Y, van de Walle R, *et al.* Iterative reconstruction algorithms in nuclear medicine. *Comput Med Imaging Graph* 2001; **25**:105–111.
14 Riddell C, Carson RE, Carrasquillo JA, Libutti SK, Danforth DN, Whatley M, Bacharach SL. Noise reduction in oncology FDG PET images by iterative reconstruction: a quantitative assessment. *J Nucl Med* 2001; **42**:1316–1323.
15 Chen H, Lei X, Yao D. An improved ordered subsets expectation maximization positron emission computerized tomography reconstruction. *Comput Biol Med* 2007; **37**:1780–1785.
16 Shepp LA, Vardi Y. Maximum likelihood reconstruction for emission tomography. *IEEE Trans Med Imaging* 1982; **1**:113–122.
17 Boellaard R, van Lingen A, Lammertsma AA. Experimental and clinical evaluation of iterative reconstruction (OSEM) in dynamic PET: quantitative characteristics and effects on kinetic modeling. *J Nucl Med* 2001; **42**:808–817.
18 Ahn S, Fessler JA. Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms. *IEEE Trans Med Imaging* 2003; **22**:613–626.
19 Hudson HM, Larkin RS. Accelerated image reconstruction using ordered subsets of projection data. *IEEE Trans Med Imaging* 1994; **13**:601–609.
20 Asma E, Ahn S, Ross SG, Chen A, Manjeshwar RM. Accurate and consistent lesion quantitation with clinically acceptable penalized likelihood images. Paper presented at Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC); 2012 IEEE; 27 October 2012 to 3 November 2012.

21 Jaskowiak CJ, Bianco JA, Perlman SB, Fine JP. Influence of reconstruction iterations on ^{18}F-FDG PET/CT standardized uptake values. *J Nucl Med* 2005; **46**:424–428.

22 Morey AM, Kadrmas DJ. Effect of varying number of OSEM subsets on PET lesion detectability. *J Nucl Med Technol* 2013; **41**:268–273.

23 Lee YS, Kim JS, Kim KM, Kang JH, Lim SM, Kim HJ. Performance measurement of PSF modeling reconstruction (True X) on Siemens Biograph TruePoint TrueV PET/CT. *Ann Nucl Med* 2014; **28**:340–348.

24 Ahn S, Ross SG, Asma E, Miao J, Jin X, Cheng L, *et al*. Quantitative comparison of OSEM and penalized likelihood image reconstruction using relative difference penalties for clinical PET. *Phys Med Biol* 2015; **60**:5733–5751.

25 Teoh EJ, McGowan DR, Bradley KM, Belcher E, Black E, Gleeson FV. Novel penalised likelihood reconstruction of PET in the assessment of histologically verified small pulmonary nodules. *Eur Radiol* 2016; **26**:576–584.

26 Queiroz MA, Hullner M, Kuhn F, Huber G, Meerwein C, Kollias S, *et al*. PET/MRI and PET/CT in follow-up of head and neck cancer patients. *Eur J Nucl Med Mol Imaging* 2014; **41**:1066–1075.

27 Sah BR, Veit-Haibach P, Delso G, Wollenweber S, Tarrade X, Licato P, *et al*. Clinical evaluation of a new block sequential regularized expectation maximization (BSREM) reconstruction algorithm in PET/CT studies. *J Nucl Med Meeting Abstracts* 2014; **55**:2097.

28 Ross SQ. *Clear whitepaper*. Waukesha, Wisconsin: GE Healthcare; 2014.

29 Queiroz MA, Delso G, Wollenweber S, Deller T, Zeimpekis K, Huellner M, *et al*. Dose optimization in TOF-PET/MR compared to TOF-PET/CT. *PLoS One* 2015; **10**:e0128842.

30 Everaert H, Vanhove C, Lahoutte T, Muylle K, Caveliers V, Bossuyt A, Franken PR. Optimal dose of ^{18}F-FDG required for whole-body PET using an LSO PET camera. *Eur J Nucl Med Mol Imaging* 2003; **30**:1615–1619.

31 Stansfield EC, Sheehy N, Zurakowski D, Vija AH, Fahey FH, Treves ST. Pediatric ^{99m}Tc-MDP bone SPECT with ordered subset expectation maximization iterative reconstruction with isotropic 3D resolution recovery. *Radiology* 2010; **257**:793–801.

32 De Pierro AR, Beleza Yamagishi ME. Fast EM-like methods for maximum 'a posteriori' estimates in emission tomography. *IEEE Trans Med Imaging* 2001; **20**:280–288.

33 Teoh EJ, McGowan DR, Macpherson RE, Bradley KM, Gleeson FV. Phantom and clinical evaluation of the Bayesian penalized likelihood reconstruction algorithm Q.Clear on an LYSO PET/CT system. *J Nucl Med* 2015; **56**:1447–1452.

34 Parvizi N, Franklin JM, McGowan DR, Teoh EJ, Bradley KM, Gleeson FV. Does a novel penalized likelihood reconstruction of ^{18}F-FDG PET-CT improve signal-to-background in colorectal liver metastases? *Eur J Radiol* 2015; **84**:1873–1878.

35 Geismar JH, Stolzmann P, Sah BR, Burger IA, Seifert B, Delso G, *et al*. Intra-individual comparison of PET/CT with different body weight-adapted FDG dosage regimens. *Acta Radiol Open* 2015; **4**:2047981614560076.

36 Adams MC, Turkington TG, Wilson JM, Wong TZ. A systematic review of the factors affecting accuracy of SUV measurements. *AJR Am J Roentgenol* 2010; **195**:310–320.