

18F-FDG PET/CT assessment of histopathologically confirmed mediastinal lymph nodes in non-small cell lung cancer using a penalised likelihood reconstruction

Eugene J. Teoh^{1,2} · Daniel R. McGowan^{2,3} · Kevin M. Bradley¹ · Elizabeth Belcher⁴ · Edward Black⁴ · Alastair Moore⁵ · Annemarie Sykes⁵ · Fergus V. Gleeson^{1,2}

Received: 27 May 2015 / Revised: 23 December 2015 / Accepted: 26 January 2016 / Published online: 25 February 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract

Purpose To investigate whether using a Bayesian penalised likelihood reconstruction (BPL) improves signal-to-background (SBR), signal-to-noise (SNR) and SUV_{max} when evaluating mediastinal nodal disease in non-small cell lung cancer (NSCLC) compared to ordered subset expectation maximum (OSEM) reconstruction.

Materials and methods 18F-FDG PET/CT scans for NSCLC staging in 47 patients (112 nodal stations with histopathological confirmation) were reconstructed using BPL and compared to OSEM. Node and multiple background SUV parameters were analysed semi-quantitatively and visually.

Results Comparing BPL to OSEM, there were significant increases in SUV_{max} (mean 3.2–4.0, $p < 0.0001$), SBR (mean 2.2–2.6, $p < 0.0001$) and SNR (mean 27.7–40.9, $p < 0.0001$). Mean background SNR on OSEM was 10.4 (range 7.6–14.0), increasing to 12.4 (range 8.2–16.7, $p < 0.0001$). Changes in background SUVs were minimal (largest mean difference 0.17 for liver SUV_{mean} , $p < 0.001$). There was no

significant difference between either algorithm on receiver operating characteristic analysis ($p = 0.26$), although on visual analysis, there was an increase in sensitivity and small decrease in specificity and accuracy on BPL.

Conclusion BPL increases SBR, SNR and SUV_{max} of mediastinal nodes in NSCLC compared to OSEM, but did not improve the accuracy for determining nodal involvement.

Key Points

- Penalised likelihood PET reconstruction was applied for assessing mediastinal nodes in NSCLC.
- The new reconstruction generated significant increases in signal-to-background, signal-to-noise and SUV_{max} .
- This led to an improvement in visual sensitivity using the new algorithm.
- Higher SUV_{max} thresholds may be appropriate for semi-quantitative analyses with penalised likelihood.

Keywords Mediastinal nodes · Lung cancer staging · PET reconstruction · Signal-to-noise · Bayesian · PET-CT

✉ Daniel R. McGowan
daniel.mcgowan@oncology.ox.ac.uk

¹ Department of Radiology, Churchill Hospital, Oxford University Hospitals NHS Foundation Trust, Oxford OX3 7LE, UK

² Department of Oncology, University of Oxford, Oxford OX3 7DQ, UK

³ Radiation Physics and Protection, Churchill Hospital, Oxford University Hospitals NHS Foundation Trust, Oxford OX3 7LE, UK

⁴ Department of Thoracic Surgery, John Radcliffe Hospital, Oxford University Hospitals NHS Foundation Trust, Oxford OX3 9DU, UK

⁵ Department of Respiratory Medicine, Churchill Hospital, Oxford University Hospitals NHS Foundation Trust, Oxford OX3 7LE, UK

Introduction

18F-FDG PET/CT is used in the staging of patients with non-small cell lung cancer (NSCLC) due to its superior accuracy in the detection of nodal involvement and metastatic disease compared to CT [1].

Meta-analyses of nodal involvement per patient and per node in NSCLC demonstrated pooled sensitivities of 0.76 and 0.65, and pooled specificities of 0.88 and 0.95 [2], suggesting that there is scope for improvement in both sensitivity and specificity. There is also a lack of evidence for the use of semi-quantitative analysis, with variable maximum standardised uptake value (SUV_{max}) thresholds used for differentiating benign from malignant involvement [2], in part

due to the limitations of different image reconstruction parameters, which are known to affect the accuracy of standardised uptake value (SUV) measurements [3, 4].

Iterative methods are commonly used for the reconstruction of PET data because of improved signal-to-noise (SNR) ratios [3–5]. The most widely used iterative algorithm is ordered subset expectation maximisation (OSEM) [6]. OSEM aims to find the most likely image through repeated iterations with each iteration giving an image with a greater likelihood of describing the measured data. However, it is not possible to run this algorithm to full convergence as the image noise increases with each iteration, becoming visually unacceptable well before full convergence is reached [3, 5]. Due to this OSEM is stopped after a stipulated number of iterations resulting in an under-converged image and underestimation of SUVs.

Bayesian penalised likelihood (BPL) is an iterative PET reconstruction which includes point spread function (PSF) modelling, recently developed by GE Healthcare (Q.Clear, GE Healthcare, Milwaukee, WI, USA) [7]. As described elsewhere [7–10], BPL includes a relative difference penalty [11] which is a function of the difference between neighbouring voxels as well as a function of their sum [12]. This penalty function acts as a noise suppression term, which allows an increased number of iterations without the noise usually seen in OSEM [7]. It is controlled by the penalisation factor (beta), which is the only user-input variable. Modified block sequential regularised expectation maximization is used as an optimiser for this BPL algorithm, which, due to the penalty function, allows effective convergence to be achieved in images, potentially providing a more accurate SUV [12, 13]. We have previously shown the improvement this algorithm provides over OSEM in phantom studies [8], lung nodules [9] and colorectal cancer liver metastases [10].

Aim

The aim of this study was to test whether using BPL increases signal-to-background (SBR), SNR and SUV_{max} when evaluating nodal disease in patients with lung cancer compared to OSEM.

Materials and methods

Patient selection

All patients who underwent ^{18}F -FDG PET/CT at our institution between October 2011 and April 2013 for the staging of NSCLC, with subsequent nodal station histopathological diagnosis, were retrospectively identified. Institutional review board approval is not required for retrospective analyses of this nature in our hospital.

^{18}F -FDG PET/CT imaging protocol

PET/CT scans were performed on a 3D mode GE Discovery 690 PET/CT system (GE Healthcare). The patients were fasted for at least 6 h prior to their scan. Their blood glucose was measured and recorded on the radiology information system prior to intravenous injection, with 4 MBq/kg of ^{18}F -FDG. Imaging commenced approximately 90 min post-injection (mean \pm one standard deviation for this patient group 92 ± 6 min) and covered the skull base to upper thighs. The PET/CT images were acquired under normal tidal respiration for 4 min per bed position. The CT was performed using a pitch of 0.984, 120 kV, automA with a noise index of 25.

PET reconstructions

PET images were reconstructed using two different algorithms with the same normalisation correction factors and both using the CT scan for attenuation correction. The standard of care PET reconstruction algorithm used is time of flight (ToF) OSEM (VPFX, GE Healthcare). This was used with two iterations, 24 subsets and 6.4-mm Gaussian filter. The sinograms generated at the time of scanning were retrospectively processed using the ToF BPL reconstruction algorithm with a penalisation factor (beta) of 400, the only user-input variable for this algorithm, as this has been shown to optimise signal detection [8].

Imaging analysis

Semi-quantitative analysis

Analysis of the pre-existing PET images (reconstructed using OSEM) and new PET images reconstructed using BPL, was performed with the PET images fused with the CT component of the original study (on mediastinal windows) for analysis.

If there was more than one node in a single station, the analysed node was chosen based on the highest FDG-avidity on OSEM reconstruction, being the standard of care reconstruction used at time of analysis. If all the nodes within the station demonstrated background FDG uptake, the largest node was chosen.

The SUV_{max} of each node was recorded using a standard volume of interest (VOI) tool. Background SUVs were measured in the right lobe of the liver and descending aorta at the level of the carina, with 3.0-cm and 1.0 diameter spherical VOIs respectively. SUV_{max} , SUV_{mean} and standard deviation (SUV_{sd}) within the VOI were recorded for both reference organs. In addition, liver SUV_{peak} was also recorded. Signal-to-background ratio (SBR) for each node was calculated as node SUV_{max} divided by the descending aorta SUV_{mean} . SNRs for both node and background were calculated using the SUV_{sd} on a reference VOI as a measure of noise. The node

SNR was defined as node SUV_{max} divided by descending aorta SUV_{sd} and the liver background SNR as liver SUV_{mean} divided by liver SUV_{sd} .

To assess the effect of BPL on nodes with background FDG uptake (on OSEM), nodes were classified as FDG-positive (above-background) or FDG-negative (at-background). Background uptake was patient-specific and set as the descending aorta SUV_{mean} on the OSEM images.

Visual analysis

Visual analysis of the OSEM and BPL PET images, fused with the CT component of the original study (viewed on mediastinal windows), was performed by a senior radiology resident with four years of radiology (including one year of PET/CT) experience, on the same workstation. Nodes were scored according to the degree of FDG uptake (above-background or at-background). The reference organ for background uptake was the descending aorta. The scorer reviewed the cases in a randomised order, blinded to the clinical outcome and outcome of prior scoring for each case. The scorer was not blinded to the nature of the reconstruction algorithm used.

Statistical analysis

Statistical analyses were performed using R [14] and IBM SPSS Statistics 22.0 (IBM Corporation, New York, NY, USA) with p values less than 0.05 considered as statistically significant. Differences in background SUV_{mean} , SUV_{max} and SNR across the entire cohort were analysed using paired t -tests. Differences in node SUV_{max} , SBR and SNR were analysed using Wilcoxon rank-sum tests. The percentage difference in node SUV_{max} ($\% \Delta SUV_{max}$) was also calculated. Percentage difference in node SUV_{max} between histopathologically positive and negative nodes were analysed using the Mann-Whitney U test.

Diagnostic performance

The performance of both algorithms to detect malignant nodes was assessed using both semi-quantitative and visual criteria. For semi-quantitative criteria, receiver operating characteristic (ROC) curves were plotted, and area under the curve (AUC) values calculated. The areas under both ROC curves were compared using the method described by DeLong et al. [15]. The optimal SUV threshold for the diagnosis of malignancy was defined as the point on the curve closest to the upper left corner of the ROC space. Sensitivity, specificity and accuracy for malignancy detection were calculated for these thresholds, and an SUV_{max} threshold of 2.5. For visual criteria, nodes scored as above-background were designated malignant and

nodes at-background were designated benign. Sensitivity, specificity and accuracy for malignancy detection were then calculated.

Results

Clinical characteristics

Forty-seven patients (29 male, 18 female, mean age 69 years, range 36–82 years) met the inclusion criteria. Within the cohort, 27 had squamous cell carcinoma, 18 had adenocarcinoma and two had adenosquamous carcinoma. A total of 112 nodal stations were included for analysis, of which 25 stations in 18 patients were histopathologically positive. Histopathological diagnosis was obtained by surgical or mediastinoscopic sampling in the majority of the stations ($n=97$), the remainder by transbronchial needle aspiration ($n=15$). The mean nodal short-axis diameter was 9 mm (range 3–27 mm). When stratified according to histopathological status, the mean short-axis diameter was 11 mm (range 5–27 mm) in positive stations and 8 mm (range 3–15 mm) in negative stations.

Background analysis

The average background SNR on OSEM was 10.4 (range 7.6–14.0), increasing to 12.4 on BPL (range 8.2–16.7, $p<0.0001$). There was no statistically significant difference in liver SUV_{max} and descending aorta SUV_{max} between OSEM and BPL ($p=0.35$ and 0.07, respectively), and very small albeit statistically significant differences in liver SUV_{mean} , liver SUV_{peak} , liver SUV_{sd} descending aorta SUV_{mean} and descending aorta SUV_{sd} (Table 1). The largest difference was in liver SUV_{mean} with a mean difference of 0.17 (95 % confidence interval (CI) 0.11–0.22).

Semi-quantitative analysis

SUV_{max}, SNR and SBR

On comparison of BPL with OSEM, there was a statistically significant difference in node SUV_{max} (mean difference 0.8, $p<0.0001$), SNR (mean difference 13.2, $p<0.0001$) and SBR (mean difference 0.4, $p<0.0001$). The $\% \Delta SUV_{max}$ was 16 %. The results of the node analysis are summarised in Table 2.

Analysis by histopathology

In the 25 histopathologically positive nodes, there was a statistically significant difference in node SUV_{max} (mean difference 1.8, $p<0.0001$) and SBR (mean difference 0.9, $p<0.001$).

Table 1 Background standardised uptake value (SUV) analysis

SUV parameter	OSEM mean (range)	BPL mean (range)	Paired t-test (BPL vs. OSEM)	
			Mean difference (95 % CI)	p-value
Liver SUV _{mean}	2.2 (1.3–3.2)	2.4 (1.3–3.4)	0.17 (0.11–0.22)	<0.001
Liver SUV _{max}	2.9 (1.7–4.4)	3.0 (1.7–4.2)	0.04 (–0.05–0.13)	0.349
Liver SUV _{peak}	2.4 (1.4–3.5)	2.6 (1.4–3.7)	0.17 (0.11–0.23)	<0.001
Liver SUV _{sd}	0.22 (0.13–0.34)	0.19 (0.12–0.27)	–0.022 (–0.013–0.032)	<0.001
D. aorta SUV _{mean}	1.5 (1.0–2.2)	1.6 (0.9–2.7)	0.07 (0.02–0.13)	0.006
D. aorta SUV _{max}	1.7 (1.2–2.7)	1.8 (1.2–3.1)	0.04 (–0.003–0.09)	0.066
D. aorta SUV _{sd}	0.13 (0.06–0.24)	0.11 (0.05–0.23)	–0.018 (–0.007–0.028)	0.001
SNR	10.4 (7.6–14.0)	12.4 (8.2–16.7)	2.0 (1.5–2.4)	<0.001

BPL Bayesian penalised likelihood reconstruction, OSEM ordered subset expectation maximum, CI confidence interval

The mean $\% \Delta \text{SUV}_{\text{max}}$ was 23.7 %. There was a relatively lower increment in SUV_{max} (mean difference 0.5, mean percentage difference 14.3 %) and SBR (mean difference 0.2) in histopathologically negative nodes although the differences were also statistically significant. The results are summarised in Table 2 and Fig. 1. The differences in $\% \Delta \text{SUV}_{\text{max}}$ between positive and negative nodes were also statistically significant (p=0.032).

Visual analysis of FDG uptake

On visual analysis of FDG uptake on OSEM compared to BPL (Table 3), scores were concordant in 100 nodes (89 %). All of the nodes with discordant scores had a higher score on BPL (two histopathology positive, ten negative).

With regard to semi-quantitative analysis, there was concordance in classification of FDG uptake in 110 nodes (98 %).

Table 2 Summary of maximum standardised uptake value (SUV_{max}), signal-to-noise (SNR), signal-to-background (SBR) and percentage difference in SUV_{max} across the entire cohort and classified according to histopathology

	All (n=112)		Positive (n=25)		Negative (n=87)	
	OSEM	BPL	OSEM	BPL	OSEM	BPL
SUV_{max}						
Mean	3.2	4.0	5.2	7.0	2.6	3.1
Median	2.4	2.9	3.4	4.4	2.3	2.8
Range	1.0 – 17.0	1.0 – 25.3	1.3 – 17.0	1.3 – 25.3	1.0 – 6.9	1.0 – 8.3
Wilcoxon p value	<0.0001		<0.0001		<0.0001	
SNR						
Mean	27.7	40.9	43.4	73.7	23.2	31.5
Median	21.2	27.4	36.0	38.4	20.7	25.8
Range	6.0 – 131.0	5.2 – 281.4	9.1 – 131.0	16.1 – 281.4	6.0 – 67.5	5.2 – 106.5
Wilcoxon p value	<0.0001		<0.0001		<0.0001	
SBR						
Mean	2.2	2.6	3.6	4.5	1.8	2.0
Median	1.8	1.9	2.9	3.0	1.7	1.9
Range	0.6 – 10.2	0.6 – 18.1	0.8 – 10.2	0.8 – 18.1	0.6 – 4.1	0.6 – 5.4
Wilcoxon p value	<0.0001		<0.001		<0.0001	
$\% \Delta \text{SUV}_{\text{max}}$						
Mean	16.4 %		23.7 %		14.3 %	
Median	14.6 %		24.2 %		13.4 %	
Range	–11.8 – 87.0 %		–3.0 – 87.0 %		–11.8 – 48.4 %	
Spearman p value (correlation vs size)	0.0001		<0.0001		0.125	
Mann-Whitney U p value	-		0.032		-	

BPL Bayesian penalised likelihood reconstruction, OSEM ordered subset expectation maximum

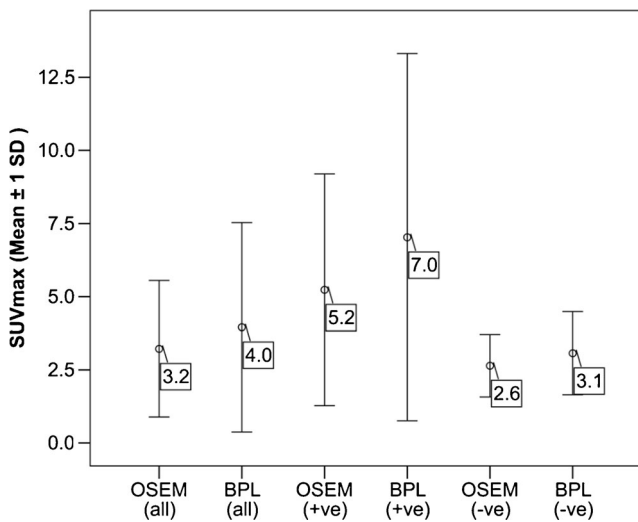


Fig. 1 Error bar chart plotting the mean maximum standardised uptake value (SUV_{max}) ± 1 SD on both reconstructions in the entire cohort and according to histopathology

The remaining two nodes had negative histopathology and were FDG-positive (above-background) on OSEM, but classified as FDG-negative (at-background) on BPL (Table 3, footnote).

Diagnostic performance

ROC curves were plotted to evaluate the usefulness of OSEM and BPL to detect histopathologically positive nodes. The AUC values were 0.711 (p=0.001) and 0.697 (p=0.003), respectively (Fig. 2), with no statistically significant difference between the two algorithms (p=0.256).

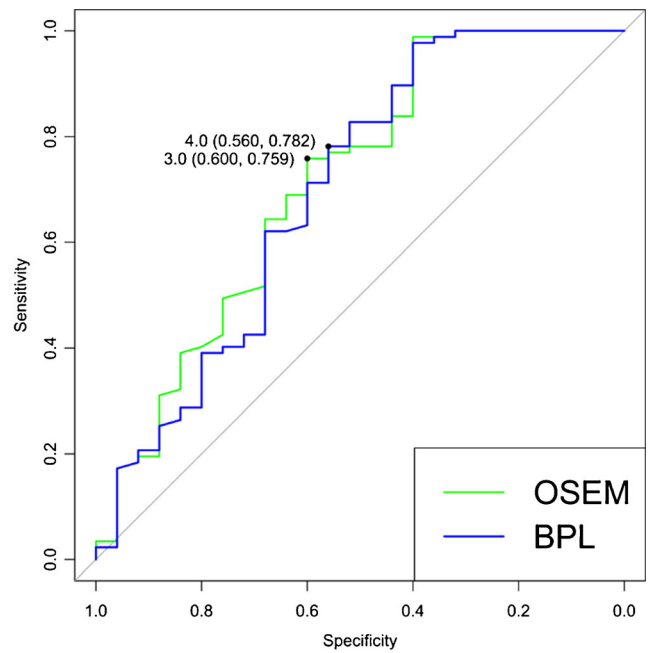


Fig. 2 Receiver operating characteristic (ROC) curves for detection of histopathologically positive lymph nodes with Bayesian penalised likelihood reconstruction (BPL) compared to ordered subset expectation maximum (OSEM). Optimum maximum standardised uptake value (SUV_{max}) thresholds for malignancy detection are indicated as SUV_{max} (sensitivity, specificity)

The optimum SUV_{max} threshold for detection of malignancy was 3.0 and 4.0 for OSEM and BPL, respectively (Fig. 2). The sensitivities, specificities and accuracies at these thresholds as an entire cohort are summarised in Table 4. Across these groups, a minor decrease in sensitivity (by 4.0 %) and increase in specificity (by 2.3 %) between OSEM and BPL

Table 3 Results of visual analysis of FDG uptake compared to semi-quantitative criteria

	Visual Above (FDG +ve) / at background (FDG -ve)	Semi-quantitative
OSEM	52 / 60	104 / 8
BPL	64 / 48	102 / 10
Concordant overall	100 (89 %)	110 (98 %)
Concordant with OSEM	52 (100 %)/48 (80 %)	102 (98 %)/8 (80 %)
Increased score / became FDG-positive	12 (11 %) (2 histopathology positive, 10 negative)	0
Decreased score / became FDG-negative	0	2 (2 %) (2 histopathology negative)

*Node 1: OSEM SUV_{max} 1.5, Background SUV_{mean} 1.4

BPL SUV_{max} 1.4 Background SUV_{mean} 1.4

*Node 2: OSEM SUV_{max} 1.5 Background SUV_{mean} 1.4

BPL SUV_{max} 1.6 Background SUV_{mean} 1.6

BPL Bayesian penalised likelihood reconstruction, OSEM ordered subset expectation maximum, SUV standardised uptake value

Table 4 Diagnostic performance of ordered subset expectation maximum (OSEM) and Bayesian penalised likelihood reconstruction (BPL) in detecting malignant nodes based on (A) semi-quantitative analysis using a maximum standardised uptake value (SUV_{max}) threshold of 2.5, (B) optimum SUV_{max} threshold (3.0 and 4.0, respectively) and (C) visual analysis

	(A) SUV _{max} 2.5		(B) Optimum SUV _{max}		(C) Visual	
	OSEM	BPL	OSEM (3.0)	BPL (4.0)	OSEM	BPL
All (n=112)						
Sensitivity	68.0 %	72.0 %	60.0 %	56.0 %	68.0 %	76.0 %
Specificity	56.3 %	40.2 %	75.9 %	78.2 %	59.8 %	48.3 %
Accuracy	58.9 %	47.3 %	72.3 %	73.2 %	61.6 %	54.5 %

was observed, with little change in accuracy (increase by 0.9 %). Conversely, using either an SUV_{max} threshold of 2.5 or visual criteria, a divergence in sensitivity and specificity between OSEM and BPL was observed, with sensitivity increasing and specificity decreasing with BPL (Table 4).

To investigate if nodal size had an influence over diagnostic performance, the sensitivities, specificities and accuracies at the optimum SUV_{max} thresholds of 3.0 and 4.0 for OSEM and BPL, respectively, were repeated with the dataset dichotomised into two groups: >10 mm and ≤10 mm (Table 5). The results in nodes >10 mm (n=24) were identical between these two groups. In nodes ≤10 mm (n=88), there was a drop in sensitivity by 7.2 %, increase in specificity by 2.7 % and accuracy by 1.2 %. These marginal changes reflect the prior results when analysed as a cohort.

Discussion

This study demonstrated a significant increase in SBR and SUV_{max} of mediastinal nodes using BPL compared to OSEM in patients with NSCLC, examples are illustrated in Figs. 3 and 4. While SBR and SUV_{max} increased in both histopathologically positive and negative nodes, the %ΔSUV_{max} in histopathologically positive nodes was significantly higher (p=0.033, SUV_{max} mean difference 1.8 vs. 0.4, respectively).

There were also significant increases in node SNR using BPL, which is of relevance as OSEM algorithms can

underestimate lesion activity when located in a relatively FDG-avid background [7]. In the context of mediastinal nodes, the surrounding blood pool may render faintly FDG-avid abnormalities less visually conspicuous. The resultant effect of BPL is demonstrated by the increase in sensitivity based on visual analysis (76 % from 68 %, Table 4). Some of the increase in SUV_{max} can be attributed to PSF modelling being included within the BPL and not within OSEM. PSF modelling incorporates information about the PET detector response into the reconstruction algorithm which leads to an improved image, especially

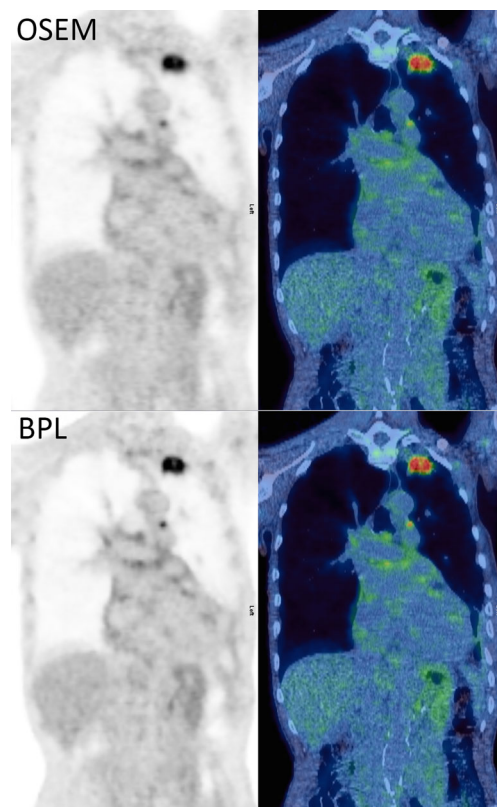
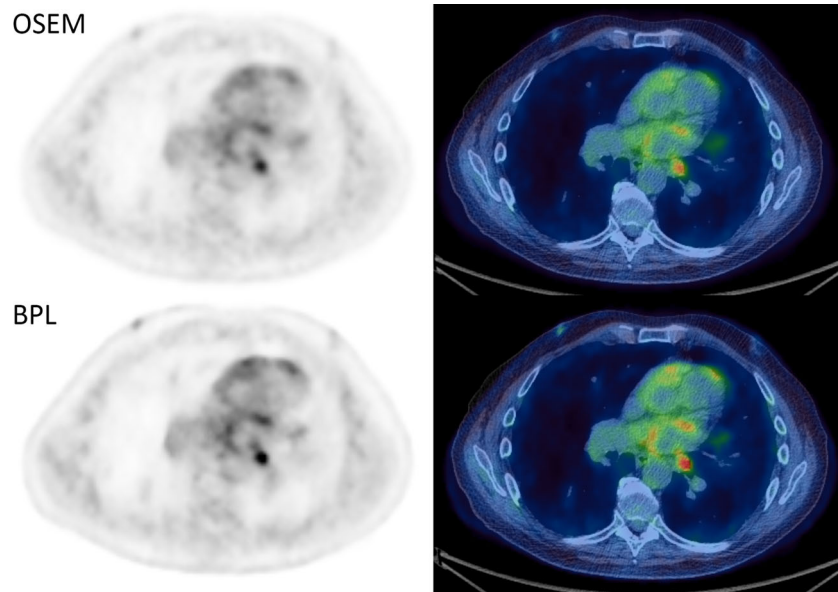


Fig. 3 A histopathologically proven involved 8-mm node in station 5 in adenocarcinoma of the left upper lobe. Ordered subset expectation maximum (OSEM) maximum standardised uptake value (SUV_{max}) 3.8, increasing to 5.0 on Bayesian penalised likelihood reconstruction (BPL). The blood pool SUV_{mean} difference was 0.1. All positron emission tomography (PET) images are displayed on SUV scale 0–6

Table 5 Diagnostic performance of ordered subset expectation maximum (OSEM) and Bayesian penalised likelihood reconstruction (BPL) in detecting malignant nodes based on size, using optimum maximum standardised uptake value (SUV_{max}) threshold (3.0 and 4.0, respectively)

	Nodes > 10mm (n=24)		Nodes ≤ 10mm (n=88)	
	OSEM	BPL	OSEM	BPL
Sensitivity	100.0 %	100.0 %	28.6 %	21.4 %
Specificity	23.1 %	23.1 %	85.1 %	87.8 %
Accuracy	58.3 %	58.3 %	76.1 %	77.3 %

Fig. 4 A histopathologically proven involved 13-mm node in station 10L in a patient with squamous cell carcinoma. Ordered subset expectation maximum (OSEM) maximum standardised uptake value (SUV_{max}) 5.4, increasing to 7.7 on Bayesian penalised likelihood reconstruction (BPL), with a visually appreciable decrease in noise of the mediastinal blood pool. All positron emission tomography (PET) images are displayed on SUV scale 0–6



for small lesions. While comparison between OSEM with PSF modelling (SharpIR on GE systems) and BPL might seem more appropriate as BPL includes PSF modelling, OSEM with PSF has not been adopted as standard of care in our centre. This is due to the increased intervoxel covariance seen with the PSF modelling [9, 16] which causes images to appear very heterogeneous. We have seen here that background metrics (Table 1, Figs. 3 and 4) remain very similar when moving from OSEM to BPL despite the addition of PSF modelling.

Despite improved definition of nodal FDG uptake and a relatively higher SUV_{max} increment in histopathologically positive nodes using BPL, ROC analysis of the usefulness of SUV_{max} as a single semi-quantitative parameter between both algorithms showed that BPL did not significantly improve the performance for diagnosing nodal disease (for example, accuracy 72.3 % (OSEM) to 73.2 % (BPL) at optimum SUV_{max} thresholds, Table 4). The observation of two OSEM FDG-negative histopathologically positive nodes remaining FDG-negative on BPL lends weight to this although the numbers are small (these two nodes were also sub-centimetre).

Interestingly, performing the same analysis dichotomising the dataset according to nodal size (>10 mm, ≤10 mm), demonstrated the same pattern of (small) change in diagnostic performance from OSEM to BPL only in nodes ≤10 mm (Table 5). This may suggest that the difference in performance is size dependent, although the effect may be due to the considerably larger number of nodes ≤10 mm (n=88) compared to nodes >10 mm (n=24). It is important to stress that the difference is clinically negligible so conclusions as to whether BPL confers any advantage or vice versa to evaluation of subcentimetre

nodes using semi-quantitative analysis cannot be drawn from this observation.

On applying the widely used SUV_{max} threshold of 2.5 [17] to both reconstructions, there was an expected increase in sensitivity and decrease in specificity using BPL compared to OSEM. On visual analysis, there was increased detection where a small, but relatively greater, proportion of nodes were ‘upgraded’ on BPL, which also resulted in a divergence of sensitivity and specificity. There was also a small decrease in accuracy from 62 % to 55 %.

Importantly the diagnostic performance was not significantly improved with BPL, even though the increments in SBR, SNR, and SUV_{max} contributed to increased visual detection. This is likely to be due to the increase in visualisation and semi-quantitative measurements occurring independent of their aetiology; glucose utilisation is better detected in both malignant and granulomatous nodes. Furthermore, when a semi-quantitative method of analysis is applied, a higher SUV_{max} threshold than is commonly used may be appropriate when using BPL, although there is no universally agreed level at present for OSEM [18–20]. Similarly, a different threshold will have to be adopted when using a visual analysis. As PET technology (hardware and software) evolves, for example with BPL reconstruction, radiologists and physicians will need to adapt and re-learn to account for the improving quality in images. The need to potentially amend current parameters for disease detection and reporting is occurring in other areas of imaging due to improved imaging technology. Improved nodule detection using low dose CT, as seen in the National Lung Screening Trial, where low-dose CT demonstrated high sensitivity but low positive predictive values in lung cancer detection, has resulted in the definition of a positive screening result being refined [21].

There are a number of limitations of this study. Firstly, there is inherent sampling error as direct radiological-pathological correlation was not possible for individual nodes and the unit of analysis had to be based on each nodal station. Where there were multiple nodes in a single station on CT, a single node was chosen for analysis based on the degree of FDG-avidity followed by size, and both of these criteria may contribute to false-positive results. This was thought not to be of overall significance as primary evaluation of the data was centred on the difference due to the methods of PET reconstruction.

Finally, the cohort was skewed with a majority proportion of histopathologically negative nodes and small absolute number of positive nodes. This may explain the relatively small differences in semi-quantitative diagnostic performance between BPL and OSEM. The small number of positive nodes may also in part explain the relatively poor AUC values derived from semi-quantitative analysis. This was based on SUV_{max} as a standalone predictor of nodal positivity but was to some extent an expected finding. It is also concordant with the wider observations of 20–25 % false-negative and false-positive rates in PET for mediastinal nodal involvement [22, 23], confirming the importance of tissue sampling in mediastinal staging to determine nodal involvement [24].

Conclusion

BPL, an iterative reconstruction technique using a Bayesian penalised likelihood reconstruction algorithm, increases SBR, SNR and SUV_{max} of mediastinal nodes in NSCLC, compared to OSEM, the current standard of care. This led to an improvement in visual sensitivity using BPL. However, this did not improve the accuracy for determining nodal involvement, and suggests that the limitations of 18F-FDG PET/CT in nodal analysis in NSCLC are due to the inherent non-specificity in the reasons for FDG-avidity, and are likely to be unchanged by improvements in techniques in its detection.

Acknowledgments The scientific guarantor of this publication is Prof. Fergus Gleeson. The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article. The authors thank Alice Asseraf and Hugo Arques at GE Healthcare for their assistance in the processing of sinograms using BPL. Daniel McGowan is funded by a National Institute for Health Research (NIHR) / Health Education England (HEE) Healthcare Scientist Chief Scientific Officer Doctoral Award. The research was supported by the NIHR Oxford Biomedical Research Centre Programme. This paper presents independent research funded by the NIHR and HEE. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, HEE, or the Department of Health.

No complex statistical methods were necessary for this paper. Institutional Review Board approval was not required because it is not required for retrospective reviews of this nature in our institution. Written informed consent was not required for this study because it is not required

for retrospective reviews of this nature in our institution. Methodology: retrospective, diagnostic or prognostic study, performed at one institution.

Open Access This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Lardinois D, Weder W, Hany TF et al (2003) Staging of non-small-cell lung cancer with integrated positron-emission tomography and computed tomography. *N Engl J Med* 348:2500–2507
- Lv YL, Yuan DM, Wang K et al (2011) Diagnostic performance of integrated positron emission tomography/computed tomography for mediastinal lymph node staging in non-small cell lung cancer: A bivariate systematic review and meta-analysis. *J Thorac Oncol* 6: 1350–1358
- Adams MC, Turkington TG, Wilson JM, Wong TZ (2010) A systematic review of the factors affecting accuracy of SUV measurements. *AJR Am J Roentgenol* 195:310–320
- Kinahan PE, Fletcher JW (2010) Positron emission tomography-computed tomography standardized uptake values in clinical practice and assessing response to therapy. *Semin Ultrasound CT MR* 31:496–505
- Tong S, Alessio AM, Kinahan PE (2010) Image reconstruction for PET/CT scanners: Past achievements and future challenges. *Imaging Med* 2:529–545
- Hudson HM, Larkin RS (1994) Accelerated image reconstruction using ordered subsets of projection data. *IEEE Trans Med Imaging* 13:601–609
- Ross S (2013) *Q.clear (GE healthcare white paper)*. http://www3.gehealthcare.com/~media/documents/us-global/products/pet-ct/whitepaper/q%20clear/ge-healthcare-white-paper_qclear.pdf. Accessed 8 December 2015
- Teoh EJ, McGowan DR, Macpherson RE, Bradley KM, Gleeson FV (2015) Phantom and clinical evaluation of the bayesian penalized likelihood reconstruction algorithm Q.clear on an LYSO PET/CT system. *J Nucl Med* 56:1447–1452
- Teoh EJ, McGowan DR, Bradley KM, Belcher E, Black E, Gleeson FV (2016) Novel penalised likelihood reconstruction of PET in the assessment of histologically verified small pulmonary nodules. *Eur Radiol* 26:576–584 doi:10.1007/s00330-015-3832-y
- Parvizi N, Franklin JM, McGowan DR, Teoh EJ, Bradley KM, Gleeson FV (2015) Does a novel penalized likelihood reconstruction of 18F-FDG PET-CT improve signal-to-background in colorectal liver metastases? *Eur J Radiol* 84:1873–1878
- Nuyts J, Beque D, Dupont P, Mortelmans L (2002) A concave prior penalizing relative differences for maximum-a-posteriori reconstruction in emission tomography. *IEEE Trans Nucl Sci* 49:56–60
- Asma E, Ahn S, Ross SG, Chen A, Manjeshwar RM (2012) Accurate and consistent lesion quantitation with clinically acceptable penalized likelihood images [abstract]. 2012 I.E. Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC):4062–4066. Doi:10.1109/NSSMIC.2012.6551928
- Ahn S, Fessler JA (2003) Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms. *IEEE Trans Med Imaging* 22:613–626
- R Core Team (2014) R: A language and environment for statistical computing

15. DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* 44:837–845
16. Rahmim A, Qi J, Sossi V (2013) Resolution modeling in PET imaging: Theory, practice, benefits, and pitfalls. *Med Phys* 40: 064301
17. Schmidt-Hansen M, Baldwin DR, Hasler E, Zamora J, Abreira V, Roque I, Figuls M (2014) PET-CT for assessing mediastinal lymph node involvement in patients with suspected resectable non-small cell lung cancer. *Cochrane Database Syst Rev* 11, CD009519. doi: [10.1002/14651858.CD009519.pub2](https://doi.org/10.1002/14651858.CD009519.pub2)
18. Kwee TC, Cheng G, Lam MG, Basu S, Alavi A (2013) SUVmax of 2.5 should not be embraced as a magic threshold for separating benign from malignant lesions. *Eur J Nucl Med Mol Imaging* 40: 1475–1477
19. Bryant AS, Cerfolio RJ, Klemm KM, Ojha B (2006) Maximum standard uptake value of mediastinal lymph nodes on integrated FDG-PET-CT predicts pathology in patients with non-small cell lung cancer. *Ann Thorac Surg* 82:417–22, discussion 422–3
20. Vansteenkiste JF, Stroobants SG, De Leyn PR et al (1998) Lymph node staging in non-small-cell lung cancer with FDG-PET scan: A prospective study on 690 lymph node stations from 68 patients. *J Clin Oncol* 16:2142–2149
21. Aberle DR, DeMello S, Berg CD et al (2013) Results of the two incidence screenings in the national lung screening trial. *N Engl J Med* 369:920–931
22. Gould MK, Kuschner WG, Rydzak CE et al (2003) Test performance of positron emission tomography and computed tomography for mediastinal staging in patients with non-small-cell lung cancer: A meta-analysis. *Ann Intern Med* 139:879–892
23. Cerfolio RJ, Bryant AS, Ojha B (2006) Restaging patients with N2 (stage IIIa) non-small cell lung cancer after neoadjuvant chemoradiotherapy: A prospective study. *J Thorac Cardiovasc Surg* 131: 1229–1235
24. De Leyn P, Doooms C, Kuzdzal J et al (2014) Revised ESTS guidelines for preoperative mediastinal lymph node staging for non-small-cell lung cancer. *Eur J Cardiothorac Surg* 45:787–798