European Journal of Radiology 117 (2019) 149–155

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

Research article

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

D.J. Murphy[a,b,*], L. Royle[c], Z. Chalampalakis[a], L. Alves[a], N. Martins[a], P. Bassett[d], R. Breen[e], A. Nair[c], A. Bille[f], S. Chicklore[a,b], G.J. Cook[a,b], M. Subesinghe[a,b]

[a] King's College London & Guy's and St. Thomas' PET Centre, St. Thomas' Hospital, London, UK
[b] Department of Cancer Imaging, School of Biomedical Engineering and Imaging Sciences, King's College London, London, UK
[c] Department of Radiology, Guy's and St. Thomas' NHS Foundation Trust, London, UK
[d] Statsconsultancy Ltd., London, UK
[e] Department of Respiratory Medicine, Guy's and St. Thomas' NHS Foundation Trust, London, UK
[f] Department of Cardiothoracic Surgery, Guy's and St. Thomas' NHS Foundation Trust, London, UK

ARTICLE INFO

Keywords:
Solitary pulmonary nodule
Fluorodeoxyglucose F18
Positron emission tomography computed tomography
Image reconstruction
Risk assessment

ABSTRACT

Purpose: British Thoracic Society (BTS) guidelines advocate using FDG PET-CT with the Herder model to estimate malignancy risk in solitary pulmonary nodules (SPNs). Qualitative and semi-quantitative assessment of SPN uptake is based upon analysis of Ordered Subset Expected Maximisation (OSEM) PET images. Our aim was to assess the effect of a Bayesian Penalised Likelihood (BPL) PET reconstruction on the assessment of SPN FDG uptake and estimation of malignancy risk (Herder score).

Methods: Subjects with SPNs who underwent FDG PET-CT between 2014–2017, with histological confirmation of malignancy or histological/imaging follow-up confirmation of benignity were included. Two blinded readers independently classified SPN uptake on both OSEM and BPL (BTS score; 1 = none; 2 = ≤ mediastinal blood pool (MBP); 3 = > MBP but ≤ 2x liver; 4 = > 2x liver), with resultant calculation of the Herder score (%) for both reconstructions.

Results: 97 subjects with 75 (77%) malignant SPNs were included. BPL increased the BTS score in 25 (26%) SPNs; 9 SPNs (7 malignant) increased from BTS score 2 to 3, 16 (13 malignant) from BTS score 3 to 4, with a mean Herder score increase of 18 ± 22%. The mean Herder score for all SPNs with BPL was higher than OSEM (73 ± 29 vs 68 ± 32%, p = 0.001). There was no difference in Herder model diagnostic performance between BPL and OSEM, with similar areas under the curve (0.84 vs 0.83, p = 0.39).

Conclusion: BPL increases the Herder score in 26% of SPNs compared to OSEM but does not alter the diagnostic performance of the Herder model.

1. Introduction

Recent guidance issued by the British Thoracic Society (BTS) in 2015 [1] on the investigation and management of solitary pulmonary nodules (SPNs) advocates the use of clinico-radiological risk prediction models to guide clinical decision making for patients who have a SPN identified and confirmed on computed tomography (CT) imaging.

Following an initial CT scan, an estimate of malignancy risk (< 10% or > 10%) is made using the Brock model [2], which encompasses findings on the CT scan (nodule size, type, location, count, spiculation, emphysema) along with clinical parameters (age, gender, family history of lung cancer). Patients with a greater than 10% risk of malignancy following their CT scan are recommended to undergo a 2-deoxy-2-[^{18}F] fluoro-D-glucose (FDG) positron emission tomography (PET)-CT scan for further risk stratification (< 10%, 10–70% or > 70% risk of malignancy), using the Herder model [3], which incorporates the degree of

Abbreviations: BPL, Bayesian penalised likelihood; BTS, British Thoracic Society; CT, computed tomography; FDG, 2-deoxy-2-[^{18}F]fluoro-D-glucose; ICC, intra-class correlation; MBP, mediastinal blood pool; OSEM, ordered subset expected maximisation; PET, positron emission tomography; PSF, point spread function; ROC, receiver operating characteristic; SPN, solitary pulmonary nodule; SUV, standardised uptake value

* Corresponding author:Present address at: St Vincent's University Hospital, Elm Park, Dublin 4, Ireland.
E-mail address: david.murphy@st-vincents.ie (D.J. Murphy).

https://doi.org/10.1016/j.ejrad.2019.06.005
Received 14 January 2019; Received in revised form 21 May 2019; Accepted 9 June 2019

Table 1
The BTS ordinal scale for classification of FDG uptake in SPNs.

SPN uptake	BTS score	Definition of FDG uptake in SPN	Definition of SPN SUV_{max}
Absent	1	≤ background lung tissue	= background lung tissue SUV_{max}
Faint	2	≤ MBP	> background lung tissue SUV_{max} but ≤ MBP SUV_{max}
Moderate	3	> MBP	> MBP SUV_{max} but ≤ 2 times hepatic SUV_{max}
Intense	4	Markedly > MBP	> 2 times hepatic SUV_{max}

BTS = British Thoracic Society; FDG = 2-deoxy-2-[^{18}F]fluoro-D-glucose; SPN = solitary pulmonary nodule; MBP = mediastinal blood pool; SUV_{max} = maximum standardised uptake value.

FDG uptake within a SPN, along with other clinico-radiological parameters (age, smoking status, history of extra-thoracic cancer, nodule size, location and spiculation). The incorporation of FDG uptake adds incremental benefit to diagnostic accuracy, increasing the area under the receiver operating characteristic (ROC) curve from 0.79 to 0.92 [3].

The Herder model, an accurate model incorporating PET for predicting malignancy in SPNs, has been validated in a UK population [4]. Its application is dependent on the accurate classification of FDG uptake in a SPN using qualitative assessment with an ordinal scale as follows; absent, faint, moderate, or intense. Herder et al. [3] did not provide detail regarding the definition of these levels of FDG uptake, although other studies have [5,6]. The BTS guideline development group adapted and derived a 4-point scale based on the above, to enable use with the Herder model (Table 1, Fig. 1).

Importantly, the degree of FDG uptake within SPNs is based upon analysis of PET images reconstructed using a standard PET reconstruction algorithm, Ordered Subset Expected Maximisation (OSEM). OSEM is an iterative reconstruction algorithm, with each successive iteration producing an image estimate closer to the true image, i.e. convergence. However, with each iteration, image noise increases and eventually reaches unacceptable levels whereby noise dominates. For this reason, OSEM is terminated after a set number of iterations resulting in an underconverged image with quantification underestimation and resultant underestimation of standardised uptake values (SUVs) [7,8].

In recent years, novel PET reconstruction algorithms have been introduced to improve image quality through increasing signal-to-noise ratio. GE Healthcare have introduced a Bayesian Penalised Likelihood (BPL) iterative PET reconstruction algorithm, Q.Clear, which is available on modern GE Healthcare PET-CT scanners [9]. This reconstruction permits an increased number of iterations without an accompanying increase in noise due to the use of a penalty function, which acts as a noise suppressor, with the level of penalisation controlled by the penalisation factor beta, set by the user. The Q.Clear algorithm also includes point spread function (PSF) modelling. Studies have shown that BPL increases the apparent level of FDG uptake particularly in small lesions compared to OSEM [10–12]. This may potentially have a significant impact on the classification of FDG uptake within SPNs according to the BTS guidelines by increasing FDG uptake within a SPN to a greater degree than that of reference soft tissues, e.g. mediastinal blood pool (MBP), with resultant increases in the estimated risk of malignancy generated by the Herder model. The aim of this study is to investigate the effect of BPL, in comparison with OSEM, on the qualitative assessment of FDG uptake and resultant estimation of malignancy risk (Herder score) in SPNs assessed using the Herder model according to the current BTS guidelines.

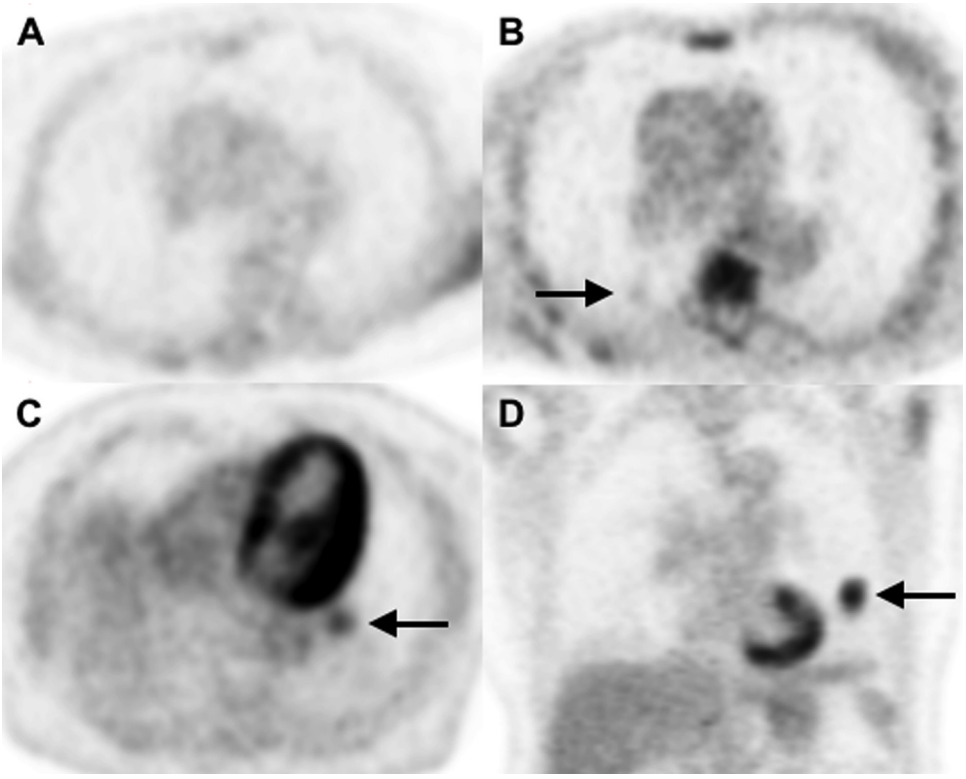

Fig. 1. Sample OSEM PET images for the BTS score (SUV scale 0–6 g/ml).
1A. BTS score 1 = No uptake. Axial image showing uniform uptake in the lung parenchyma, with no increased uptake seen in the 10 mm right upper lobe (RUL) nodule (not visible)
1B. BTS score 2 = Faint. Axial image showing faint FDG uptake in a 12 mm right lower lobe (RLL) nodule (arrow). Nodule SUV_{max} = 1.4 g/ml; Mediastinal blood pool SUV_{max} = 1.9 g/ml.
1C. BTS score 3 = Moderate. Axial image showing moderate FDG uptake in a 18 mm left lower lobe (LLL) nodule (arrow). Nodule SUV_{max} = 2.6 g/ml; Mediastinal blood pool SUV_{max} = 1.8 g/ml.
1D. BTS score 4 = Intense. Coronal image showing intense FDG uptake in a 22 mm LLL nodule (arrow). Nodule SUV_{max} = 6.9 g/ml; Liver SUV_{max} = 2.9 g/ml.

2. Methods

2.1. Inclusion criteria

Patients who underwent FDG PET-CT assessment of their SPN between January 2014-December 2017 were identified through institutional databases.

Eligible patients fulfilled the following criteria:

- FDG PET-CT examination performed on a GE 710 PET-CT system (GE Healthcare, Chicago, IL, USA) with assessable OSEM and BPL PET reconstructions.
- SPNs measuring ≥ 5 mm and ≤ 30 mm in size.
- Histological confirmation obtained either via percutaneous biopsy or surgical resection for all malignant nodules.
- Histological confirmation, unchanged morphological appearances on CT follow-up (either by mean diameter for a minimum of 2 years, or by volumetric analysis for a minimum of 1 year) or resolution on follow-up imaging for all benign nodules.

An institutional board review approval was obtained for this retrospective analysis.

2.2. FDG PET-CT imaging protocol

All FDG PET-CT examinations were performed on one of two institutional GE 710 PET-CT systems. Patients were fasted for a minimum of 6 h. A standard PET acquisition from skull base to upper thighs was acquired post-injection of 350 M Bq (± 10%) of FDG with an axial field view of 15.7 cm and an 11 slice overlap at 3 min per bed position. The mean tracer uptake time was 84 ± 10 min (range 59–99 min). Non-attenuation corrected and attenuation corrected datasets were reconstructed. The low-dose unenhanced CT component was performed with patients maintaining normal shallow respiration, using a standardised protocol with the following settings: 140 kV, pitch 1.375, 40 mm beam collimation in the z-axis corresponding with the detector bank length (64 x 0.625 mm) with an effective slice collimation of 0.625 mm, 0.5 s rotation time, Auto mA (15–100 mA, noise index 40). CT images were reconstructed with a slice thickness of 2.5 mm.

2.3. PET reconstructions

The standard clinical PET reconstruction was a Time of Flight OSEM reconstruction using 2 iterations, 24 subsets and a 6.4 mm Gaussian filter (VPFX, GE Healthcare). The archived raw data sinograms were retrospectively reconstructed using the novel BPL algorithm (Q.clear, GE Healthcare) with the established penalisation factor (beta) of 400, which has been identified to be the optimal penalty factor for FDG PET-CT imaging in oncology from previous studies [10–13]. PET images were reconstructed with a slice thickness of 3.27 mm and pixel size of 4.7 mm.

2.4. FDG PET-CT imaging review

Image analysis was undertaken independently by a PET-CT fellow (DJM) and a consultant radionuclide radiologist (MS) with 2 and 7 years of experience in PET-CT reporting, respectively. Qualitative and semi-quantitative analysis was performed on both PET reconstructions by both readers in a random order, blinded to the clinical information. Qualitative assessment of FDG uptake within SPNs was assigned a score according to the BTS score (Table 1, Fig. 1). Although not explicitly mandated by Herder et al. [3] or BTS guidelines [1], semi-quantitative measures of FDG uptake, i.e. SUV_{max}, were used to help confirm the qualitative assessment and classification of FDG uptake in SPNs using the BTS scale. A freehand ROI was outlined in the arch of the aorta, careful to avoid the aortic walls, to define SUV_{max} MBP, whilst a large freehand ROI was outlined in the right lobe of the liver to define SUV_{max} hepatic activity, ignoring any voxels that contained spuriously elevated SUV_{max} values. We chose to measure MBP and hepatic SUV_{max}, as SUV_{max} is easily available, has good inter-reader reproducibility, and is relatively unaffected by partial volume averaging [14], and is recommended in the lymphoma literature as standard methodology regarding the Deauville criteria [15]. Scans with disagreement in BTS score between the two readers had consensus reads for both reconstructions. Once FDG uptake in all SPNs was classified, the estimated risk of malignancy was calculated for each SPN using the Herder model [3,16] for both PET reconstructions, as follows:

Herder model = $1/(1 + e^{-x})$,

where x = -4.739 + 3.691 (% probability by the model of Swensen et al [17]) + 2.322 (BTS score 2) + 4.617 (BTS score 3) + 4.771 (BTS score 4); e is the base of natural logarithms; the level of nodule FDG uptake (BTS score) is denoted as 1 for present and 0 for absent. The Swenson model is as follows:

Swensen model = $1/(1 + e^{-x})$,

where x = -6.8272 + 0.0391 (age in years) + 0.7917 (cigarettes) + 1.3388 (cancer history) + 0.1274 (diameter) + 1.0407 (spiculation) + 0.7838 (upper lobe); cigarettes is 1 if current/former smoker, otherwise 0; cancer history is 1 if the patient has a history of extra-thoracic cancer; spiculation is 1 if the nodule is spiculated and upper lobe is 1 if the nodule has an upper lobe location.

2.5. Statistical analysis

Continuous variables are presented as mean ± standard deviation (SD). Categorical variables are presented as frequencies and percentages. Group comparison of continuous variables was performed using the Student's t-test. Group comparisons of categorical variables were performed using the Wilcoxon matched pairs test. A value of $P < 0.05$ was considered significant. Inter-observer agreement for each PET reconstruction was assessed as follows. Firstly, the agreement between the categorised BTS scores was examined. Due to the ordinal nature of these categories, the weighted kappa method was used to assess agreement as follows: 0.81–1.00 = very good agreement; 0.61-0.80 = good agreement, 0.41–0.60 = moderate agreement; 0.21–0.4 = fair agreement [18]. Secondly, the agreement between observers for the continuous Herder scores was examined using the intra-class coefficient (ICC) method. Bootstrapping methods were used to calculate a confidence interval for the difference in the level of agreement between the two PET reconstructions. For each PET reconstruction, the diagnostic performance of the continuous Herder score was determined by calculating the area under the ROC curve (AUC) [19]. The AUCs from the two PET reconstructions were compared using methods of comparing correlated ROC curves [20]. ROC curves were also generated for SPN SUV_{max} for each PET reconstruction using the averaged SPN SUV_{max} between the two readers. The ROC curve results were analysed to determine the sensitivity, specificity, positive and negative predictive values for a series of pre-defined Herder score cut-off points (0.1, 0.5 and 0.7), and to determine the optimum SPN SUV_{max} cut-off value. Confidence intervals were calculated using the exact binomial method.

3. Results

3.1. Clinical and nodule characteristics (Table 2)

Ninety-seven patients (47 female, 50 male, mean age 69 ± 10 years) with 97 SPNs met the inclusion criteria. Of the 97 SPNs, 75 (77%) were malignant, and 22 (23%) benign. Malignant SPNs were predominantly non-small cell lung cancers (n = 67, 89%), with extra-thoracic metastasis (n = 7, 9%) and small cell lung cancer (n = 1, 1%)

Table 2
Basic patient and nodule characteristics.

Patient demographics (n = 97)	Overall n ± SD (%)	Benign (n = 22, 23%)	Malignant (n = 75, 77%)	P value
Mean Age (years)	69 ± 10	70.9 ± 9.4	67.1 ± 14.8	0.25
Female	47 (48%)	9 (41%)	38 (51%)	0.47
Current/former smoker	81 (84%)	16 (73%)	65 (87%)	0.19
Previous malignancy	27 (28%)	5 (23%)	22 (29%)	0.6
Nodule characteristics				
Mean nodule diameter (mm)	16 ± 6	11.4 ± 5.3	17.8 ± 5.5	0.0001
Upper lobe	51 (53%)	12 (55%)	39 (52%)	1
Spiculated	54 (56%)	5 (23%)	49 (65%)	0.0005

comprising the remainder. Benign aetiologies included infection, mycobacterial granulomata, granulomatous disease and intrapulmonary lymph nodes. The mean SPN size was 16 ± 6 mm (range 5–29 mm), with 20 SPNs (21%) measuring ≤ 10 mm in size. Malignant SPNs were significantly larger than benign SPNs (17.8 ± 5.5 mm vs. 11.4 ± 5.3 mm, p = 0.0001), and were more likely to be spiculated (49 vs. 5, p = 0.0005).

3.2. BTS nodule uptake score

There was a significant difference in FDG uptake in SPNs classified using the BTS score, between the two PET reconstruction methods, with significantly higher BTS scores observed with BPL compared to OSEM (Fig. 2, p < 0.001). 84% of SPNs showed BTS 3 or 4 uptake on BPL, compared to 75% for OSEM. There was concordance in BTS score between the two PET reconstructions in 74% of SPNs (n = 72) (Fig. 3); BTS score 1 = 5 SPNs, BTS score 2 = 10 SPNs, BTS score 3 = 26 SPNs, BTS score 4 = 31 SPNs. BPL increased the BTS score in 25 (26%) SPNs (20 malignant, 5 benign) (Fig. 4); differences in BTS score were only by a single category with 9 SPNs (7 malignant) increasing from BTS score 2 to 3, and 16 (13 malignant) increasing from BTS score 3 to 4 (Fig. 2). No SPN was categorised as a higher BTS score using OSEM. There was no disagreement between BPL and OSEM in the categorisation of BTS score 1 uptake. The mean diameter of the 25 SPNs that demonstrated an increase in BTS score with BPL was 13.8 ± 3.9 mm (range 8–24 mm), significantly smaller than the mean diameter of 17.3 ± 6.4 mm (range 5–29 mm) in the remainder of the SPNs that were unchanged (p = 0.015).

3.3. Herder score (Table 3)

Commensurate with the overall increase in the level of SPN FDG uptake associated with BPL, the overall mean Herder score was higher for BPL compared to OSEM (73 ± 29% vs. 68 ± 32%, p = 0.001). The Herder score for malignant SPNs was higher for BPL than OSEM (83 ± 19% vs. 78 ± 25%, p = 0.004), but not for benign SPNs (42 ± 35% vs. 37 ± 34%, p = 0.07). For the 9 SPNs that increased from a BTS score 2 to 3, the mean Herder score increase was 48 ± 3%, whilst for the 16 SPNs increasing from a BTS score 3 to 4, the mean Herder score increase was only 2 ± 1%. Adenocarcinomas comprised 57% (n = 43) of malignant SPNs and were the only histological subtype to demonstrate a significantly higher Herder score on BPL compared to OSEM images (84 ± 16% vs. 76 ± 27%, p = 0.005); these results are summarised in Table 3.

3.4. Nodule SUV_{max} (Table 3)

Mean nodule SUV_{max} was significantly higher on the BPL images compared to OSEM for the overall cohort (7 ± 5.2 g/ml vs 4.9 ± 3.5 g/ml respectively, p = 0.001), for malignant nodules (8.3 ± 5.1 g/ml vs 5.8 ± 3.5 g/ml respectively, P = 0.0004), but not for benign nodules (2.4 ± 1.9 g/ml vs 1.8 ± 1.3 g/ml respectively, p = 0.07). There was no significant difference between the two PET reconstructions for measurement of mean MBP SUV_{max}, or mean liver SUV_{max}.

3.5. Inter-observer agreement

The inter-observer agreement using the BTS score was very good for both BPL and OSEM, with weighted kappas of 0.89 (95% CI: 0.74, 1.0) and 0.85 (95% CI: 0.71, 0.99), respectively. There was also very good inter-observer agreement using the continuous Herder score for both BPL and OSEM, with an ICC value of 0.98 (95% CI: 0.97, 0.99) for BPL and 0.93 (95% CI: 0.90, 0.95) for OSEM.

3.6. Diagnostic performance

The diagnostic performance of the continuous Herder score (Fig. 5a)

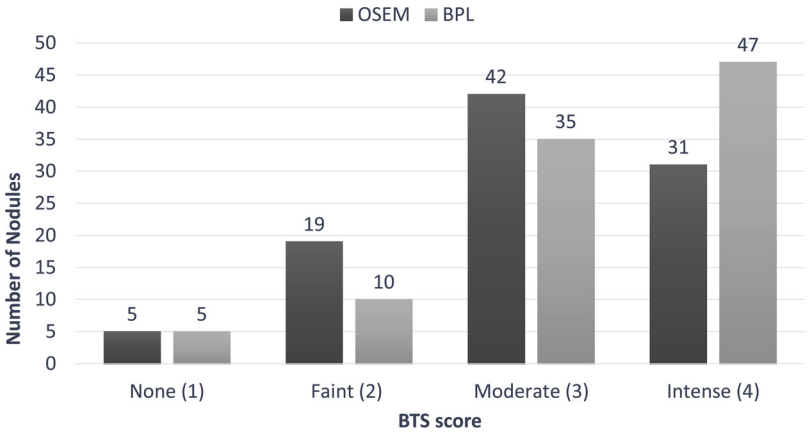

Fig. 2. Comparison of SPN FDG uptake levels (BTS score) between OSEM and BPL reconstructions.
BTS = British Thoracic Society; OSEM = Ordered Subset Expected Maximisation; BPL = Bayesian Penalisation Likelihood

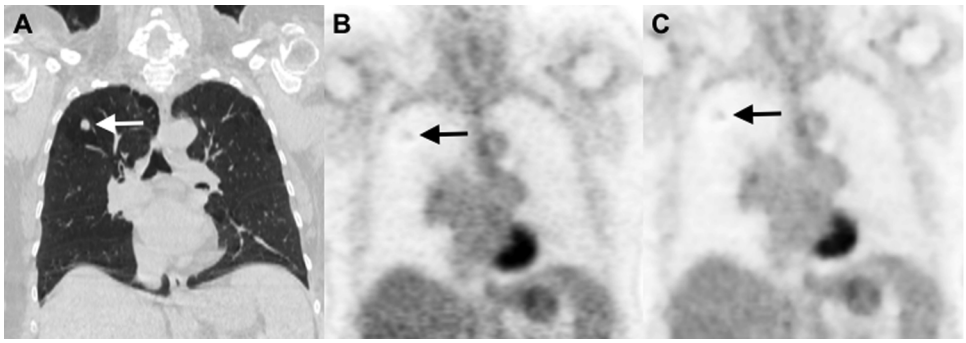

Fig. 3. Example of a SPN which showed the same level of uptake on both OSEM and BPL reconstructed PET images.
3A. Coronal CT image on lung windows showing a 13 mm RUL nodule (arrow).
3B. Coronal OSEM PET image showing faint FDG uptake in the RUL nodule (arrow); BTS score 2. Nodule SUV_{max} = 1.5 g/ml; Mediastinal blood pool SUV_{max} = 2.0 g/ml (SUV scale 0–6 g/ml).
3C. Coronal BPL PET image showing faint FDG uptake in the RUL nodule (arrow); BTS score 2. Nodule SUV_{max} = 1.6 g/ml; Mediastinal blood pool SUV_{max} = 1.9 g/ml (SUV scale 0–6 g/ml).

revealed a good performance for both methods, with an AUC of 0.84 for BPL (95% CI: 0.75, 0.93), and 0.83 for OSEM (95% CI 0.74, 0.92) with no statistically significant difference (p = 0.39). Using a Herder score cut-off of 10%, as proposed by the BTS guidelines, resulted in similar sensitivity and specificity for both techniques (Table 4). A Herder score of 50% was more sensitive for BPL than OSEM, but less specific, and a cut-off of 70% yielded similar results between the two methods. The diagnostic performance of SPN SUV_{max} **(**Fig. 5b**)** for predicting malignancy, using each PET reconstruction, revealed a good performance of each method, with an AUC of 0.87 for BPL (95% CI: 0.79, 0.94) and 0.88 for OSEM (95% CI: 0.81, 0.96), with no statistically significant difference between the two PET reconstructions (p = 0.23). There was no statistically significant difference between the diagnostic performance of the Herder score and SPN SUV_{max} for either OSEM or BPL (p = 0.09 and 0.35 respectively). The optimum SPN SUV_{max} cut-off value for BPL was 2.8 g/ml (sensitivity 0.87, specificity 0.68, PPV 0.9, NPV 0.6), and for OSEM 2.9 g/ml (sensitivity 0.78, specificity 0.82, PPV 0.94, NPV 0.53).

Table 3
Mean Herder scores & nodule SUV_{max} for BPL and OSEM reconstructions.

Mean Herder scores (% risk malignancy)			
Grouping (n)	**OSEM (% ± SD)**	**BPL (% ± SD)**	**P value**
Overall (97)	68 ± 32	73 ± 29	0.0011
Malignant (75)	78 ± 25	83 ± 19	0.0039
Benign (22)	37 ± 34	42 ± 35	0.07
Malignant nodules (n, %)			
Adenocarcinoma (43, 57%)	76 ± 27	84 ± 16	0.005
Squamous (15, 20%)	91 ± 6	92 ± 6	0.3
Carcinoid (5, 7%)	66 ± 13	68 ± 13	0.2
NSCLC, NOS (4, 5%)	82 ± 10	82 ± 10	1
Small cell (1, 1%)	80	80	1
Metastasis (7, 9%)	67 ± 38	67 ± 39	0.3
Mean Nodule SUV_{max}	**OSEM (g/ml ± SD)**	**BPL (g/ml ± SD)**	**P value**
Overall (95)	4.9 ± 3.5	7 ± 5.2	0.001
Malignant (75)	5.8 ± 3.5	8.3 ± 5.1	0.0004
Benign (22)	1.8 ± 1.3	2.4 ± 1.9	0.2

OSEM = Ordered Subset Expected Maximisation; BPL = Bayesian Penalisation Likelihood; NSCLC = non-small cell lung cancer; NOS = not otherwise specified.

4. Discussion

BPL PET reconstruction increased the BTS score by a single category in comparison with OSEM in 25 (26%) SPNs (20 malignant, 5 benign). SPNs whose BTS score increased with BPL were significantly smaller than those that remained unchanged (mean SPN diameter 13.8 mm vs. 17.3 mm respectively). Both Teoh et al. [11] and Howard et al. [21] found the greatest increase in SUV_{max} was in smaller SPNs, particularly those in the 8–15 mm range.

BPL enables effective convergence of the image through an increased number of iterations by application of a penalisation factor to suppress image noise [13], which improves quantification and increases SUV_{max}, particularly in smaller lesions [22]. However, due to PSF modelling in Q.clear, the likelihood of artefacts increases [22]. This phenomenon, known as a Gibb's artefact, is an artefactual increase in uptake at opposite edges of lesions, which in smaller lesions get closer together and overlap at a critical diameter, resulting in an artificial overestimation of intensity of uptake in the centre of the lesion. Yamaguchi et al. [23] showed that although BPL suppresses such artefacts, overestimation of uptake continues in smaller lesions, to a greater degree than OSEM with PSF modelling, resulting in concern that incorrect diagnoses based on SUVs might increase, particularly with lesions ˜10 mm.

Although the increase in BTS score with BPL translated into an overall significant increase in the Herder score, this did not result in a change in diagnostic accuracy (0.84) compared to OSEM (0.83). Reviewing the 25 SPNs that increased their BTS score with BPL; 16 increased from BTS score 3 to BTS score 4. This results in only a small change in the weighting factor associated with level of FDG uptake in the Herder model equation [3], resulting in an average final increase in the Herder score of only 2%. In practice, this is unlikely to result in management change, as nodules with BTS score 3 are already likely to be considered high risk for malignancy, given that BTS guidance advocates further investigation for any SPN with a Herder score of > 10% [1]. The 9 SPNs that increased from BTS score 2 to BTS score 3 with BPL did result in a significant increase in the Herder score of 48%, due to an

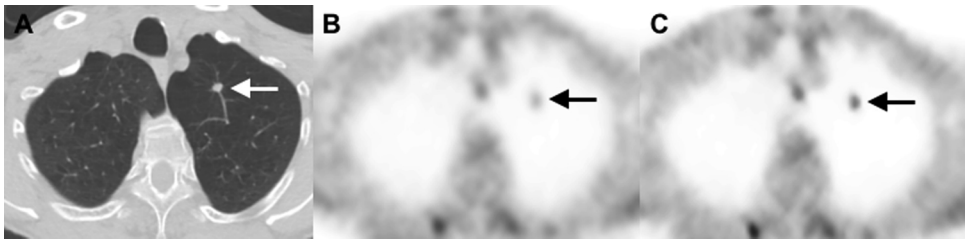

Fig. 4. Example of a nodule which showed increased uptake on BPL compared to OSEM PET reconstructed images.
4A. Axial CT image on lung windows showing an 8 mm LUL nodule (arrow).
4B. Axial OSEM PET image showing faint uptake in the LUL nodule (arrow); BTS score 2. Nodule SUV_{max} = 0.9 g/ml, mediastinal blood pool SUV_{max} = 1.2 g/ml (SUV scale 0–6 g/ml).
4C. Axial BPL PET image showing moderate uptake in the LUL nodule (arrow); BTS score 3. Nodule SUV_{max} = 1.9 g/ml, mediastinal blood pool SUV_{max} = 1.1 g/ml (SUV scale 0–6 g/ml).

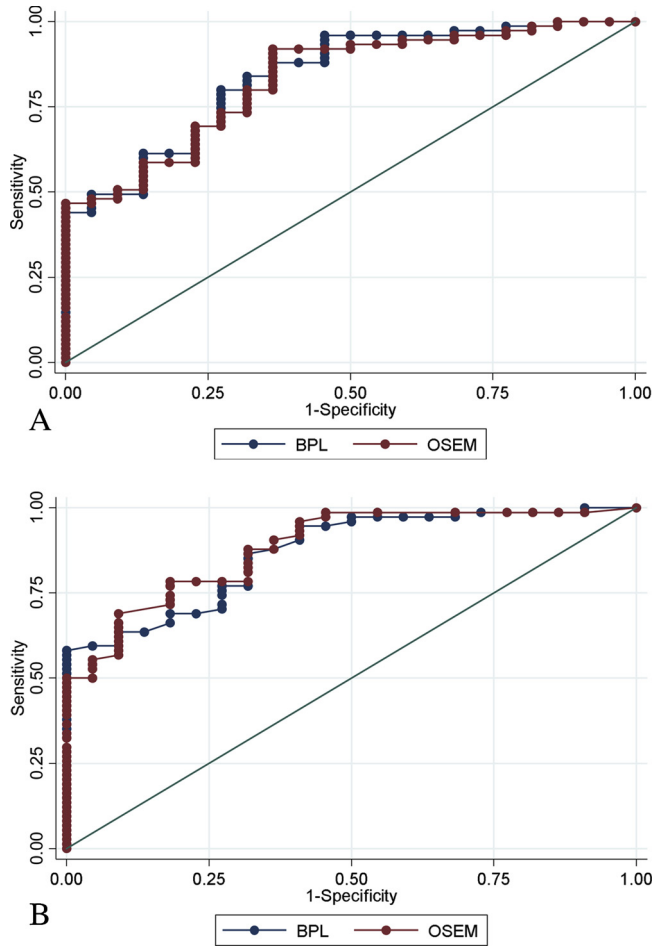

Fig. 5. (a) ROC curves for the diagnostic performance of the Herder score for predicting malignancy in SPNs, using both BPL and OSEM PET reconstructions. OSEM = Ordered Subset Expected Maximisation; BPL = Bayesian Penalisation Likelihood. (b) ROC curves for the diagnostic performance of SPN SUV_{max} for predicting malignancy using both BPL and OSEM PET reconstructions. OSEM = Ordered Subset Expected Maximisation; BPL = Bayesian Penalisation Likelihood

Table 4
Comparison of diagnostic abilities of OSEM and BPL reconstructions using predefined Herder score cut-off points, =.

Herderscore (%)	Statistic	OSEMEstimate (95% CI)	BPLEstimate (95% CI)
10	Sensitivity	0.99 (0.93, 1.00)	0.99 (0.93, 1.00)
	Specificity	0.18 (0.05, 0.40)	0.18 (0.05, 0.40)
	PPV	0.80 (0.71, 0.88)	0.80 (0.71, 0.88)
	NPV	0.80 (0.28, 1.00)	0.80 (0.28, 1.00)
50	Sensitivity	0.87 (0.77, 0.93)	0.96 (0.89, 0.99)
	Specificity	0.64 (0.41, 0.83)	0.55 (0.32, 0.76)
	PPV	0.89 (0.80, 0.95)	0.88 (0.79, 0.94)
	NPV	0.58 (0.37, 0.78)	0.80 (0.52, 0.96)
70	Sensitivity	0.77 (0.66, 0.86)	0.83 (0.72, 0.90)
	Specificity	0.68 (0.45, 0.86)	0.68 (0.45, 0.86)
	Positive PV	0.89 (0.79, 0.96)	0.90 (0.80, 0.96)
	NPV	0.47 (0.29, 0.65)	0.54 (0.34, 0.73)

OSEM = Ordered Subset Expected Maximisation; BPL = Bayesian Penalisation Likelihood; PPV = positive predictive value; NPV = negative predictive value.

higher weighting factor in the Herder model equation [3]. A near 50% increase in Herder score has the potential to change management, resulting in more aggressive and invasive management of SPNs, i.e. excision or non-surgical management (+/- image guided biopsy). The overall increase in Herder scores with BPL across the cohort explains the higher sensitivities compared to OSEM using the Herder score cut-offs of 0.5 and 0.7. The BTS guidelines propose a Herder score cut-off of 10%, above which further investigation or biopsy is recommended [1], and both OSEM and BPL demonstrated similar high sensitivities and low specificities at this cut-off of 0.99 and 0.18 respectively. Reassuringly, there was complete agreement between the two methods in determining BTS score 1. Distinguishing between BTS 2 and BTS 3 score in metabolically active SPNs is the most important PET-related factor determining the final Herder score.

There was very good inter-observer agreement using the BTS scale. Herder et al. classified nodule uptake into faint, mild, moderate and intense, but did not define these cut-offs [3]. Al-Ameri et al. assigned SUV_{max} cut-offs to help classification; no uptake; faint uptake ($SUV_{max} \leq 2.5$ g/ml); moderate uptake (SUV_{max} 2.6–10 g/ml) intense uptake ($SUV_{max} > 10$ g/ml) [4]. Assigning strict SUV_{max} cut-off values can be problematic due to the myriad of factors that can affect SUV_{max} readings other than lesional metabolic activity, including blood glucose levels, radiotracer uptake time, scanner type and PET reconstruction algorithm [7]. Internal reference controls, i.e. MBP and hepatic uptake, can overcome some of these reproducibility issues and have been validated in other malignancies [24]. The 2015 BTS guidelines defined the uptake levels proposed by Herder et al. with reference to background lung and MBP uptake [1]. We incorporated SUV_{max} rather than SUV_{mean} in these reference regions to help confirm the qualitative BTS scale classification because although SUV_{mean} may be less influenced by image noise than SUV_{max}, its reproducibility is more dependent on standardising the location and size of the region of interest than $SUV_{max,}$ limiting consistency between readers [24].

In our cohort we found slightly higher AUCs for the diagnostic performance of SPN SUV_{max} on both OSEM and BPL compared with the Herder model, although the differences were not significant. This may be explained by the high prevalence of malignancy in our cohort of 77% compared with that of 57.5% in that of Herder et al. [3]. A meta-analysis by Gould et al [25]comprising 1474 pulmonary lesions found that semiquantitative analysis of FDG uptake provided no additional benefit to the diagnostic accuracy achieved through qualitative visual assessment.

There are several advantages to the use of risk prediction models. Firstly, they enable an estimate of malignancy risk based upon several established clinical and imaging predictors of malignancy rather than sole reliance on a single imaging parameter, i.e. SPN FDG uptake. Secondly, the estimation of risk helps guide management strategies with the lowest risk favouring the least invasive approach, i.e. CT surveillance, and vice versa; such practice is supported by data from cost-effectiveness studies [26,27]. Finally, and perhaps most importantly, it facilitates shared decision making between clinicians and patients by helping interpret SPN malignancy risk in the clinical context of risk of potential complications associated with any proposed invasive procedure, such as image-guided biopsy.

Our study has limitations. It is a single centre, retrospective study, with a large malignant cohort of SPNs. Identification of more benign SPNs was difficult due to a malignant bias in local FDG PET-CT referral patterns, which reflects real world clinical practice. This malignant bias did not however impact our primary aim to investigate the effect of BPL on BTS scale classification and Herder score in SPNs undergoing FDG PET-CT assessment. Inclusion of small nodules, down to 5 mm in size, are at the limits of the spatial resolution of PET for accurate characterisation. Histological validation was not available for all benign SPNs, but this again reflects real world practice, where stability on follow-up imaging obviates the need for histological sampling. We used a freehand ROI to measure SPN FDG uptake for BPL and OSEM images; this could potentially introduce operator error in drawing the ROI around the nodule, but this limitation is mitigated by our use of SUV_{max}, which has lower operator related variability compared with SUV_{mean} [28]. There was heterogeneity in the tracer uptake time across

the cohort, which also has the potential to affect FDG uptake in SPNs, although this effect is difficult to quantify.

5. Conclusion

Classification of FDG uptake in a SPN as either below (BTS score 2) or above that of the MBP (BTS score 3) is the most important PET-related factor when using the Herder model and can be undertaken with very good interobserver agreement for both OSEM and BPL. BPL reconstruction did not affect the overall diagnostic performance of the Herder model in comparison with OSEM, but BPL did change the BTS score in 26% of nodules and significantly increased the Herder score compared to OSEM, particularly in nodules ˜13 mm. Smaller nodules may show increased FDG uptake with BPL due to a better estimation of true uptake but also possibly due to false overestimation due to Gibb's artefact. If using BPL to quantify FDG uptake in a SPN, we recommend having OSEM reconstruction available for simultaneous review. Studies with a larger cohort and a higher proportion of benign SPNs may help determine the optimum Herder score cut-off value to distinguish between benign and malignant SPNs using BPL reconstructed images.

Funding

The authors acknowledge support from the King's College London / University College London Comprehensive Cancer Imaging Centres funded by Cancer Research UK and Engineering and Physical Sciences Research Council in association with the Medical Research Council and the Department of Health (C1519/A16463) and the Wellcome Trust EPSRC Centre for Medical Engineering at King's College London (WT203148/Z/16/Z).

Conflicts of interest

No author declares a conflict of interest.

Ethical approval

Ethical approval was obtained for this retrospective study by the local institutional review board. There were no additional procedures performed in human or animal participants.

Informed consent

Informed consent was waived by the IRB for this retrospective study.

References

[1] M.E.J. Callister, D.R. Baldwin, A.R. Akram, S. Barnard, P. Cane, J. Draffan, et al., British Thoracic Society guidelines for the investigation and management of pulmonary nodules, Thorax 70 (Suppl 2) (2015) ii1–ii54, https://doi.org/10.1136/thoraxjnl-2015-207168.

[2] A. McWilliams, M.C. Tammemagi, J.R. Mayo, H. Roberts, G. Liu, K. Soghrati, et al., Probability of cancer in pulmonary nodules detected on first screening CT, N. Engl. J. Med. 369 (2013) 910–919, https://doi.org/10.1056/NEJMoa1214726.

[3] G.J. Herder, H. van Tinteren, R.P. Golding, P.J. Kostense, E.F. Comans, E.F. Smit, et al., Clinical prediction model to characterize pulmonary nodules: validation and added value of 18F-fluorodeoxyglucose positron emission tomography, Chest 128 (2005) 2490–2496, https://doi.org/10.1378/chest.128.4.2490.

[4] A. Al-Ameri, P. Malhotra, H. Thygesen, P.K. Plant, S. Vaidyanathan, S. Karthik, et al., Risk of malignancy in pulmonary nodules: a validation study of four prediction models, Lung Cancer 89 (2015) 27–30, https://doi.org/10.1016/j.lungcan.2015.03.018.

[5] J.W. Fletcher, S.M. Kymes, M. Gould, N. Alazraki, R.E. Coleman, V.J. Lowe, et al., A comparison of the diagnostic accuracy of 18F-FDG PET and CT in the characterization of solitary pulmonary nodules, J. Nucl. Med. 49 (2008) 179–185, https://doi.org/10.2967/jnumed.107.044990.

[6] L. Evangelista, A. Panunzio, R. Polverosi, F. Pomerri, D. Rubello, Indeterminate lung nodules in cancer patients: pretest probability of malignancy and the role of 18F-FDG PET/CT, AJR Am. J. Roentgenol. 202 (2014) 507–514, https://doi.org/10.2214/AJR.13.11728.

[7] M.C. Adams, T.G. Turkington, J.M. Wilson, T.Z. Wong, A systematic review of the factors affecting accuracy of SUV measurements, AJR Am. J. Roentgenol. 195 (2010) 310–320, https://doi.org/10.2214/AJR.10.4923.

[8] S. Tong, A.M. Alessio, P.E. Kinahan, Image reconstruction for PET/CT scanners: past achievements and future challenges, Imaging Med. 2 (2010) 529–545, https://doi.org/10.2217/iim.10.49.

[9] General Electric (GE), GE Healthcare White Paper: Q.Clear, (2014) Available from:https://www.gehealthcare.com/-/media/739d885baa59485aaef5ac0e0eeb44a4.pdf (Accessed 12 January 2019).

[10] E.J. Teoh, D.R. McGowan, K.M. Bradley, E. Belcher, E. Black, A. Moore, et al., 18F-FDG PET/CT assessment of histopathologically confirmed mediastinal lymph nodes in non-small cell lung cancer using a penalised likelihood reconstruction, Eur. Radiol. 26 (2016) 4098–4106, https://doi.org/10.1007/s00330-016-4253-2.

[11] E.J. Teoh, D.R. McGowan, K.M. Bradley, E. Belcher, E. Black, F.V. Gleeson, Novel penalised likelihood reconstruction of PET in the assessment of histologically verified small pulmonary nodules, Eur. Radiol. 26 (2016) 576–584, https://doi.org/10.1007/s00330-015-3832-y.

[12] N. Parvizi, J.M. Franklin, D.R. McGowan, E.J. Teoh, K.M. Bradley, F.V. Gleeson, Does a novel penalized likelihood reconstruction of 18F-FDG PET-CT improve signal-to-background in colorectal liver metastases? Eur. J. Radiol. 84 (2015) 1873–1878, https://doi.org/10.1016/j.ejrad.2015.06.025.

[13] E.J. Teoh, D.R. McGowan, R.E. Macpherson, K.M. Bradley, F.V. Gleeson, Phantom and clinical evaluation of the Bayesian penalized likelihood reconstruction algorithm Q.CLear on an LYSO PET/CT system, J. Nucl. Med. 56 (2015) 1447–1452, https://doi.org/10.2967/jnumed.115.159301.

[14] P.E. Kinahan, J.W. Fletcher, Positron emission tomography-computed tomography standardized uptake values in clinical practice and assessing response to therapy, Semin. Ultrasound CT MR. 31 (2010) 496–505, https://doi.org/10.1053/j.sult.2010.10.001.

[15] S.F. Barrington, R. Kluge, FDG PET for therapy monitoring in Hodgkin and non-Hodgkin lymphomas, Eur. J. Nucl. Med. Mol. Imaging 44 (2017) 97–110, https://doi.org/10.1007/s00259-017-3690-8.

[16] British Thoracic Society, BTS Pulmonary Nodule Risk Prediction Calculator, (2016) Available from: www.brit-thoracic.org.uk/standards-of-care/guidelines/bts-guidelines-for-the-investigation-and-management-of-pulmonary-nodules/bts-pulmonary-nodule-risk-prediction-calculator/ (Accessed 12 January 2019).

[17] S.J. Swensen, M.D. Silverstein, D.M. Ilstrup, C.D. Schleck, E.S. Edell, The probability of malignancy in solitary pulmonary nodules. Application to small radiologically indeterminate nodules, Arch. Intern. Med. 157 (1997) 849–855.

[18] J.R. Landis, G.G. Koch, The measurement of observer agreement for categorical data, Biometrics 33 (1977) 159–174.

[19] J.V. Carter, J. Pan, S.N. Rai, S. Galandiuk, ROC-ing along: evaluation and interpretation of receiver operating characteristic curves, Surgery 159 (2016) 1638–1645, https://doi.org/10.1016/j.surg.2015.12.029.

[20] E.R. DeLong, D.M. DeLong, D.L. Clarke-Pearson, Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach, Biometrics 44 (1988) 837–845.

[21] B.A. Howard, R. Morgan, M.P. Thorpe, T.G. Turkington, J. Oldan, O.G. James, et al., Comparison of Bayesian penalized likelihood reconstruction versus OS-EM for characterization of small pulmonary nodules in oncologic PET/CT, Ann. Nucl. Med. 31 (2017) 623–628, https://doi.org/10.1007/s12149-017-1192-1.

[22] C.S. van der Vos, D. Koopman, S. Rijnsdorp, A.J. Arends, R. Boellaard, J.A. van Dalen, et al., Quantification, improvement, and harmonization of small lesion detection with state-of-the-art PET, Eur. J. Nucl. Med. Mol. Imaging 44 (2017) 4–16, https://doi.org/10.1007/s00259-017-3727-z.

[23] S. Yamaguchi, K. Wagatsuma, K. Miwa, K. Ishii, K. Inoue, M. Fukushi, Bayesian penalized-likelihood reconstruction algorithm suppresses edge artifacts in PET reconstruction based on point-spread-function, Phys. Med. 47 (2018) 73–79, https://doi.org/10.1016/j.ejmp.2018.02.013.

[24] S.F. Barrington, N.G. Mikhaeel, L. Kostakoglu, M. Meignan, M. Hutchings, S.P. Müeller, et al., Role of imaging in the staging and response assessment of lymphoma: consensus of the international conference on malignant lymphomas imaging working group, J. Clin. Oncol. (2014) 3048–3058, https://doi.org/10.1200/JCO.2013.53.5229.

[25] M.K. Gould, C.C. Maclean, W.G. Kuschner, C.E. Rydzak, D.K. Owens, Accuracy of positron emission tomography for diagnosis of pulmonary nodules and mass lesions: a meta-analysis, JAMA 285 (2001) 914–924.

[26] S.S. Gambhir, J.E. Shepherd, B.D. Shah, E. Hart, C.K. Hoh, P.E. Valk, et al., Analytical decision model for the cost-effective management of solitary pulmonary nodules, J. Clin. Oncol. 16 (1998) 2113–2125, https://doi.org/10.1200/JCO.1998.16.6.2113.

[27] C. Lejeune, K. Al Zahouri, M.-C. Woronoff-Lemsi, P. Arveux, A. Bernard, C. Binquet, et al., Use of a decision analysis model to assess the medicoeconomic implications of FDG PET imaging in diagnosing a solitary pulmonary nodule, Eur. J. Health Econ. 6 (2005) 203–214, https://doi.org/10.1007/s10198-005-0279-0.

[28] Y.-E. Huang, C.-F. Chen, Y.-J. Huang, S.D. Konda, D.E. Appelbaum, Y. Pu, Interobserver variability among measurements of the maximum and mean standardized uptake values on (18)F-FDG PET/CT and measurements of tumor size on diagnostic CT in patients with pulmonary tumors, Acta Radiol. 51 (2010) 782–788, https://doi.org/10.3109/02841851.2010.497772.