

ORIGINAL RESEARCH

Open Access



# Reproducibility of compartmental modelling of $^{18}\text{F}$ -FDG PET/CT to evaluate lung inflammation

Laurence D. Vass<sup>1\*</sup>, Sarah Lee<sup>2</sup>, Frederick J. Wilson<sup>3</sup>, Marie Fisk<sup>1,4</sup>, Joseph Cheriyan<sup>1,3,4</sup> and Ian Wilkinson<sup>1,4</sup>

\*Correspondence:

ldv22@medschl.cam.ac.uk

<sup>1</sup>Experimental Medicine and Immunotherapeutics, Department of Medicine, Addenbrookes Hospital, Cambridge, UK  
Full list of author information is available at the end of the article

## Abstract

**Introduction:** Compartmental modelling is an established method of quantifying  $^{18}\text{F}$ -FDG uptake; however, only recently has it been applied to evaluate pulmonary inflammation. Implementation of compartmental models remains challenging in the lung, partly due to the low signal-to-noise ratio compared to other organs and the lack of standardisation. Good reproducibility is a key requirement of an imaging biomarker which has yet to be demonstrated in pulmonary compartmental models of  $^{18}\text{F}$ -FDG; in this paper, we address this unmet need.

**Methods:** Retrospective subject data were obtained from the EVOLVE observational study: Ten COPD patients (age =  $66 \pm 9$ ; 8M/2F), 10  $\alpha_1$ ATD patients (age =  $63 \pm 8$ ; 7M/3F) and 10 healthy volunteers (age =  $68 \pm 8$ ; 9M/1F) never smokers. PET and CT images were co-registered, and whole lung regions were extracted from CT using an automated algorithm; the descending aorta was defined using a manually drawn region. Subsequent stages of the compartmental analysis were performed by two independent operators using (i) a MIAKAT<sup>TM</sup> based pipeline and (ii) an in-house developed pipeline. We evaluated the metabolic rate constant of  $^{18}\text{F}$ -FDG ( $K_{im}$ ) and the fractional blood volume ( $V_b$ ); Bland-Altman plots were used to compare the results. Further, we adjusted the in-house pipeline to identify the salient features in the analysis which may help improve the standardisation of this technique in the lung.

**Results:** The initial agreement on a subject level was poor: Bland-Altman coefficients of reproducibility for  $K_{im}$  and  $V_b$  were 0.0031 and 0.047 respectively. However, the effect size between the groups (i.e. COPD,  $\alpha_1$ ATD and healthy subjects) was similar using either pipeline. We identified the key drivers of this difference using an incremental approach: ROI methodology, modelling of the IDIF and time delay estimation. Adjustment of these factors led to improved Bland-Altman coefficients of reproducibility of 0.0015 and 0.027 for  $K_{im}$  and  $V_b$  respectively.

**Conclusions:** Despite similar methodology, differences in implementation can lead to disparate results in the outcome parameters. When reporting the outcomes of lung compartmental modelling, we recommend the inclusion of the details of ROI methodology, input function fitting and time delay estimation to improve reproducibility.

**Keywords:** Positron emission tomography computed tomography, Kinetic modelling, Lung inflammation, Fluorodeoxyglucose F18, Reproducibility of results, Pulmonary disease, Chronic obstructive, Biomarkers

## Introduction

Inflammation is a hallmark of many respiratory diseases and is thought to be complicit in the progression of chronic lung diseases such as chronic obstructive pulmonary disease (COPD). The limitations of clinical measurements to quantify pulmonary inflammation are well documented [1].  $^{18}\text{F}$ -FDG PET-CT is a non-invasive imaging technique; quantification of  $^{18}\text{F}$ -FDG uptake has emerged as a promising biomarker to assess lung inflammation [2]. Recruitment of inflammatory cells requires increased glucose utilisation; therefore,  $^{18}\text{F}$ -FDG uptake should be elevated in inflammatory pathologies [3].

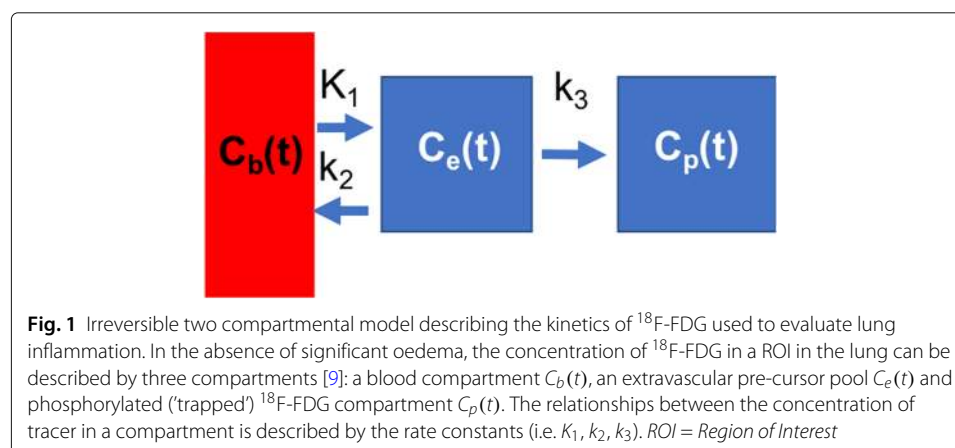
A major challenge in quantifying  $^{18}\text{F}$ -FDG in the lung is the poor signal-to-noise ratio; the low basal uptake of FDG is a consequence of the low density of lung tissue (due to large proportions of air). Interpretation is further confounded by FDG within pulmonary blood, which in the healthy lung is substantially larger (typically 15–20%) than other organs, e.g. the brain (typically 5%). Static measures, such as the standard uptake value (SUV), are likely to be heavily influenced by these factors which has led to the exploration of alternative methods that account for these effects [4].

Kinetic modelling has traditionally been regarded as the gold standard method of quantification in PET studies; well established applications include estimation of cerebral metabolic rate (CMR) and neurological receptor binding (a general framework for kinetic modelling is described in [5]). In the lung, kinetic modelling has recently been used to explicitly account for the effects of air and blood on the rate of  $^{18}\text{F}$ -FDG uptake [6, 7] which may lead to a better estimation of underlying inflammation. Provided there is not significant oedema, the two compartment irreversible model [8] has been widely adopted to model the kinetics of  $^{18}\text{F}$ -FDG in pulmonary inflammation [9] (see Fig. 1).

The concentration of  $^{18}\text{F}$ -FDG measured in a region of interest (ROI) within the lung can be described by:

$$C_m(t) = V_b C_b(t) + (1 - V_a - V_b) C_t(t, K_1, k_2, k_3, V_b) + V_a C_a(t) \quad (1)$$

where  $C_m(t)$  is the concentration of  $^{18}\text{F}$ -FDG in the ROI;  $C_b(t)$  and  $C_t(t)$  are the concentrations of  $^{18}\text{F}$ -FDG in the pulmonary blood vessels and lung cells respectively.  $V_a$  and  $V_b$  are the fractional volumes of air and blood respectively.  $K_1$ ,  $k_2$  and  $k_3$  are the microparameters of the model [5]. The concentration of radioactivity in air,  $C_a(t)$ , is assumed negligible.  $V_a$  can be derived from a CT image as described in [10].



The metabolic rate constant of  $^{18}\text{F}$ -FDG is then given by

$$K_{im} = K_1 k_3 / (k_2 + k_3) \quad (2)$$

The microparameters and  $V_b$  are estimated by minimising the weighted residual sum of squares (WRSS):

$$\text{WRSS} = \sum_{i=1}^N W_i (Y(i) - C_m(i))^2 \quad (3)$$

where  $N$  is the number of time frames of the dynamic scan,  $i$  is the frame number and  $W_i$  is the weighting factor for each frame,  $C_m(i)$  is the estimated concentration of  $^{18}\text{F}$ -FDG fitted from the compartmental model in frame  $i$  (i.e. Eq. 1) and  $Y(i)$  is the measured concentration from the PET scanner.

An essential requirement of this approach is an accurate blood input function ( $C_b(t)$ ) [11]; although arterial blood samples taken during the PET scan are considered the gold standard, obtaining samples from a peripheral vein is less invasive and onerous. Alternatively, a blood TAC can be derived from the dynamic PET images by delineating a ROI within a vessel in the field of view, referred to as an image-derived input function (IDIF). Optimal positioning of the blood vessel ROI has been discussed extensively in application to  $^{18}\text{F}$ -FDG tumour kinetics [12] and cardiac metabolism [13]; IDIFs derived from several different vessels were found to be comparable to arterial samples. In pulmonary  $^{18}\text{F}$ -FDG kinetics, several regions have been explored including the ascending and descending pulmonary aorta, left and right ventricles and aorta [14, 15]. The TAC extracted from the blood pool ROI is then modelled as a continuous curve to reduce noise: models based on an initial linear rise followed by a sum of exponentials have been proposed in tumour kinetics [16]. In the brain, extensive effort has been made to improve input function modelling including those based on reference regions [17] and methods using carotid or other blood vessels with one or more manual samples [18–20] with voxel-based approaches also feasible [21]. In the lung, there remains a need for optimisation of input function modelling.

Early studies of kinetic models of cerebral blood flow (CBF) demonstrated the importance of correcting the blood input function for time delay [22] (i.e. the time taken for the tracer to travel between the blood sampling site and the tissue of interest). Methods to estimate the delay between blood sample point and the tissue ROI are largely based on these earlier observations in neurological PET [23, 24]. More recently, inaccuracies in time delay were shown to cause significant deviations in the microparameters of a CMR kinetic model in rodents [25]. This applies equally to pulmonary compartmental models of  $^{18}\text{F}$ -FDG: incorporating a regional lung time delay has been shown to improve the fit to the experimental data compared to no delay in acute lung injury (ALI) [26]. Further corrections such as accounting for the partial volume effect (PVE) and spill-over in pulmonary  $^{18}\text{F}$ -FDG scans may be important [15], but the impact of this approach on kinetic parameter estimation in humans has yet to be explored.

The approach taken to model the input function and estimate the time delay are clearly operator dependent with several different feasible methodologies. Further, estimation of the microparameters in Eq. 2 depends on the chosen optimisation algorithm and the applied variance model (i.e. choice of  $W_i$ ) [27]—these too are not yet standardised in pulmonary compartmental models. Experts in quantitative  $^{18}\text{F}$ -FDG lung imaging have

recently highlighted the need to improve the standardisation of the analysis and assess the reproducibility of the outcome parameters (e.g.  $K_{im}$  and  $V_b$ ) [2, 28].

To this end, our aim was to evaluate the reproducibility of pulmonary compartmental models—this has yet to be performed in  $^{18}\text{F}$ -FDG models of lung inflammation and, to our knowledge, reproducibility of kinetic modelling outcomes in neurological PET also has yet to be disclosed. Reproducibility will be influenced by the number of operator dependent steps required to estimate  $K_{im}$  and  $V_b$  (see Fig. 2) and the inherent variability associated with measurements of low signal. We investigate the reproducibility of the analysis with two operators who independently analysed 30 lung scans using different analysis pipelines. During the evaluation, we identified the key parameters in the analyses which could be standardised to help improve reproducibility in pulmonary compartmental modelling of  $^{18}\text{F}$ -FDG.

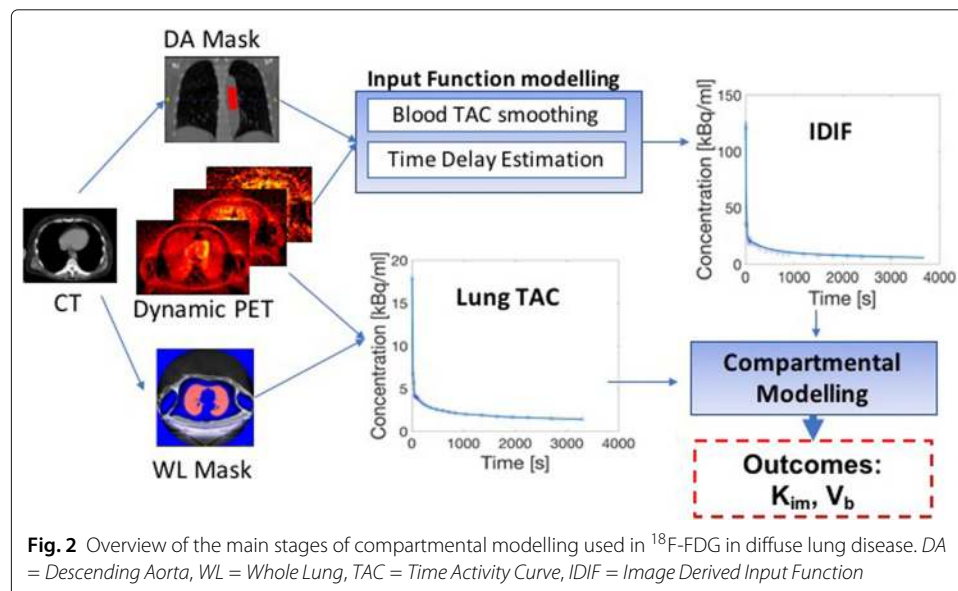
## Methods

### Study data

Thirty age and gender matched patients from the EVOLVE study were included in this evaluation: ten patients with COPD (age=  $66 \pm 9$ ; 8M/2F), ten patients with  $\alpha_1$ ATD (age=  $63 \pm 8$ ; 7M/3F) and ten healthy never smokers (age=  $68 \pm 8$ ; 9M/1F). The EVOLVE study was a cross-sectional multi-centre study to investigate vascular and pulmonary inflammation (REC 13/EE/0165, UK CRN ID 1513); the primary outcomes of this study are reported in [29]. Patients were clinically diagnosed with COPD stratified with fibrinogen  $\leq 2.8\text{g/L}$ . Healthy volunteers (HV) were recruited if they had no regular smoking history and normal predicted spirometry (further protocol details are available [30]).

### Imaging protocol

A single bolus of approximately 240MBq  $^{18}\text{F}$ -FDG was injected into the antecubital vein and images were acquired for 60 min under list mode acquisition (further details are available in the Additional file 1). A low dose attenuation correction CT (CTAC) was acquired under free breathing prior to PET scanning. Three subjects (1 COPD, 2 HV)



were removed from subsequent analysis after failing initial quality control procedures (see Additional file 1). The blood input function was calculated using an IDIF (as described in the following section); discrete venous blood samples were obtained as a means to calculate a plasma-over-blood (POB) ratio.

### Imaging analyses

For the initial analysis, subject data were analysed by operator A (as described in [7]) using a Molecular Imaging and Kinetic Analysis Toolbox (MIAKAT<sup>TM</sup> [31]; Version no: 4.2.6) based pipeline—herein referred to as pipeline A. Operator B independently analysed the same dataset using an in-house developed pipeline using MATLAB [32]—herein referred to as pipeline B.

The pipelines shared common procedures for pre-processing: the segmentation of the whole lung and blood vessels were performed using ITK-SNAP (www.itksnap.org [33]), and blood TACs were obtained from an ROI manually delineated within the descending aorta (DA), drawn in the centre of the vessel to minimise the partial volume effect. Our rationale for using the DA is based on previous work which revealed it to be preferable in estimating kinetic rate constants [15] and based on its previous use [7]. The principle differences between the pipelines are outlined in Table 1.

In pipeline A, subsequent analysis was performed using a MIAKAT-based pipeline: MIAKAT software was modified for lung <sup>18</sup>F-FDG kinetics by operator A. Importantly, both pipelines used the same compartmental model as the basis to estimate the metabolic rate of <sup>18</sup>F-FDG ( $K_{im}$ ) and the fraction blood volume ( $V_b$ )—the main outcome parameters for this study.

In order to understand the drivers of any differences in the results, operator B investigated the salient parts of the analysis methodology, which led to the differences and adjusted elements of pipeline B (described in the Further investigation section) to improve agreement.

**Table 1** Comparison of two analysis pipelines used to estimate metabolic rate of <sup>18</sup>F-FDG to assess pulmonary inflammation

Parameter	Pipeline A	Pipeline B: initial	Pipeline B: final
Blood ROI size	Circular, 5 pixel diameter, aortic arch to variable	Circular, 8 pixel diameter, 25 slices beginning aortic arch	As pipeline A
Lung ROI closing and erosion operation	5 pixel diameter disc	3 pixel diameter disc	As pipeline B initial
Input function model	Exponential basis functions	Tri- or biexponential fits	As pipeline A
Time delay estimation	Inside compartmental model optimisation	Outside compartmental model optimisation	As pipeline A
Time delay fitting	Delays spanning −50 to 50 s using 1-compartmental model fitted for first 5 min—lowest residual sum of squares	Additional parameter within estimation of rate constants	As pipeline A
Optimisation	Local optimum	Global optimum	As pipeline A
Start point of optimisation	$K_1 = 0.5, k_2 = 0.2, k_3 = 0.3, V_b = 0.1$	Multiple start points generated finds best guess (lowest objective function value)	As pipeline A

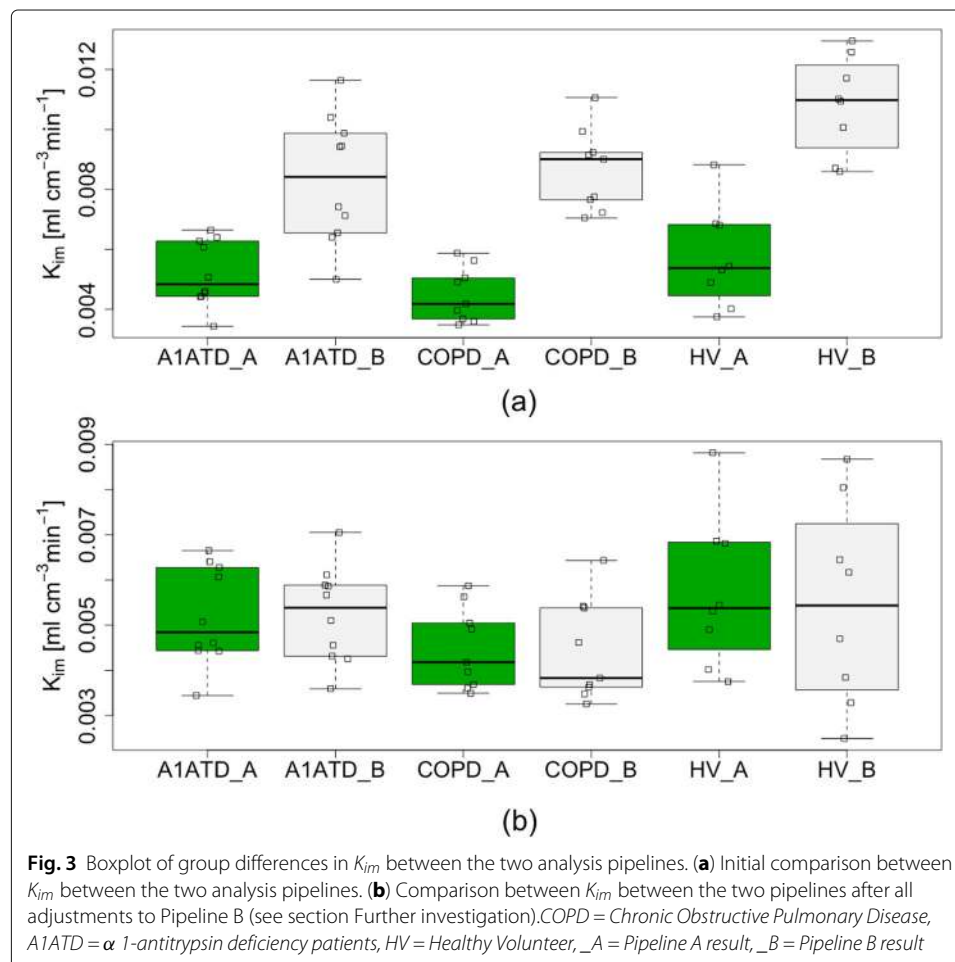
The columns “Pipeline B: initial” and “Pipeline B: final” describes the parameters which were used in the initial evaluation and the final settings used following adjustments to pipeline B respectively. The table highlights the key differences between the implementations of the compartmental model. Parameters not included below were identical between the analysis pipelines. *WL* = whole lung, *DA* = descending aorta, *ROI* = region of interest

### Statistics

Bland-Altman plots were used to compare the different outcome variables; medians, inter-quartile ranges and correlations were also used. The Bland-Altman coefficient of reproducibility is given by  $2 \times SD$ ; we expect 95% of the difference to be less than this value. The differences in  $K_{im}$  and  $V_b$  between the two pipelines were plotted as a histogram to ensure that a normal distribution was observed. To investigate group differences the Hedge's  $g$  effect was used, as a further complementary measure we used the unpaired  $t$  test. Unless otherwise stated, significance is considered when  $P < 0.05$ . The intra-observer repeatability was assessed using the coefficient of variation (COV) by operator B using pipeline B.

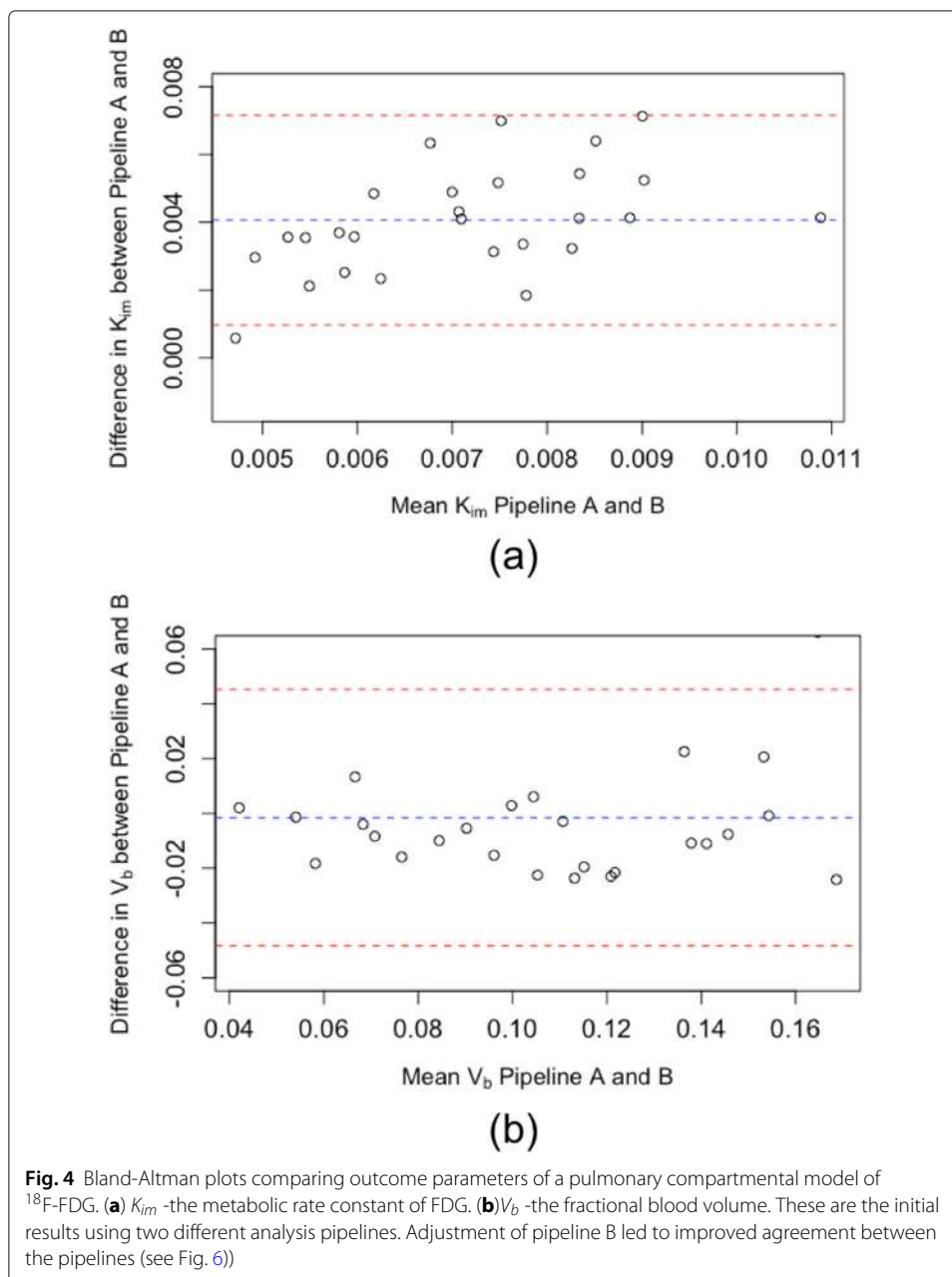
### Results

Initial results are shown in Fig. 3a; boxplots show the differences in  $K_{im}$  between COPD,  $\alpha_1$ ATD and controls using the two pipelines. Although a systemic offset was observed between the values obtained between the pipelines, this did not alter the overall group-level conclusions: the Hedge's  $g$  factor for the difference between COPD and HV was  $-0.89$  for pipeline A and  $-0.57$  for pipeline B. Further, for pipelines A and B, no significant difference was found between these groups using the two sample  $t$  test ( $p = 0.088$  and  $p = 0.26$ , respectively). The variance in pipeline B was greater than pipeline A in the  $\alpha_1$ ATD



**Fig. 3** Boxplot of group differences in  $K_{im}$  between the two analysis pipelines. **(a)** Initial comparison between  $K_{im}$  between the two analysis pipelines. **(b)** Comparison between  $K_{im}$  between the two pipelines after all adjustments to Pipeline B (see section Further investigation). COPD = Chronic Obstructive Pulmonary Disease, A1ATD =  $\alpha_1$ -antitrypsin deficiency patients, HV = Healthy Volunteer, \_A = Pipeline A result, \_B = Pipeline B result

group ( $IQR = 0.0031$  vs  $0.0018$ , respectively). There was poor agreement in  $K_{im}$  on a subject level between the pipelines: Fig. 4a shows the Bland-Altman plot for  $K_{im}$ . The mean difference in  $K_{im}$  between the two platforms was  $0.0041 \text{ ml} \cdot \text{cm}^{-3} \cdot \text{min}^{-1}$  with upper and lower limits of agreement (uloa and lloa) of  $0.00097$  and  $0.0072 \text{ ml} \cdot \text{cm}^{-3} \cdot \text{min}^{-1}$  respectively; the correlation coefficient was  $0.62$ . Figure 4a indicates a systematic relationship may exist between the difference and the mean values of  $K_{im}$  estimated using the two pipelines. Following the approach suggested in [34] to reduce systematic bias, we log transformed the data, giving a mean  $K_{im}$  of  $1.8$ , lloa of  $1.19$  and uloa of  $2.73$  (after transformation back to the original scale). Although the transformation improved



the situation, the agreement between pipelines is still poor. Figure 4b shows the Bland-Altman plot for  $V_b$ . The mean difference in  $V_b$  was  $-0.0015$  (uloa = 0.045, lloa =  $-0.048$ ); the correlation coefficient was 0.80. The Bland-Altman coefficients of reproducibility were 0.0031 and 0.047 for  $K_{im}$  and  $V_b$  respectively. To assess the repeatability of analysis, ten subjects (from COPD group) were analysed five times with pipeline B (operator B): the within subject SD of  $K_{im}$  was  $8.28 \times 10^{-4} \text{ ml} \cdot \text{cm}^{-3} \cdot \text{min}^{-1}$  and the COV was 6.1%.

### Further investigation

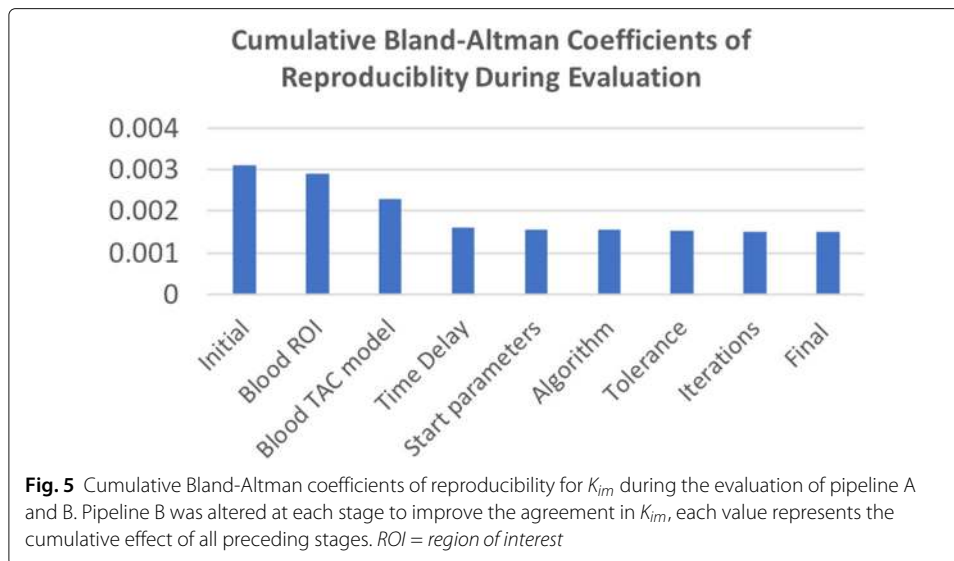
Subsequently, we sought to understand the drivers of the difference described above; this section describes the steps we undertook. Firstly, visual inspection of the lung tissue TACs from pipeline A and B revealed minimal differences. Further, the mean square error (MSE) between lung TACs from pipeline A and pipeline B was  $0.11 \pm 0.040$ ; therefore, we concluded that the lung TACs were not responsible for the differences.

Next, we undertook a visual inspection of the blood TACs, this revealed slight differences due to the ROI methodology (e.g. size and location) used by the two operators. To improve agreement, we applied the same DA ROI methodology: in pipeline B, we reduced the area of each ROI and ensured the same begin and end locations in the axial slice as pipeline A. This led to an improvement in the visual comparison of the blood TACs and a modest improvement in  $K_{im}$  (mean difference between was  $0.0037 \text{ ml} \cdot \text{cm}^{-3} \cdot \text{min}^{-1}$  compared to the original value of  $0.0041 \text{ ml} \cdot \text{cm}^{-3} \cdot \text{min}^{-1}$ ). The modelling of the blood TAC was then compared: we found the fitting of the input functions differed chiefly in the early stage of the scan ( $< 5$  min post-injection)—likely due to the inherent noise due to short frame durations. In pipeline B, we altered the input function modelling to match the approach adopted in pipeline A. Firstly, we applied the same fitting function: here, the blood TAC is modelled by basis functions [35] with a varying number of exponentials; the exponential model that best fits the blood TAC was found using a least squares algorithm. With both pipelines using this approach, the Bland-Altman coefficients of reproducibility were modestly improved (0.0023 for  $K_{im}$  and 0.034 for  $V_b$ ).

The time delay estimation is also highly dependent on the initial time frames of the input function; we found an association between the difference in outcome parameters, particularly  $V_b$ , and the estimated time delay. The mean difference in time delay estimation between pipeline A and B was  $3 \pm 5.3$  s. To improve the agreement between the estimated time delays for each subject, we replicated in pipeline B the method outlined in [7] (used in pipeline A). Namely, a one compartmental model was fitted to the first 5 min of the smoothed blood TAC and lung TAC. Then, a delay of  $\pm 50$  s in 1s increments was introduced and the delay was estimated by finding the minimum value of the residual sum of squares on the model fit. The mean difference in time delay estimation was improved to  $0.76 \pm 1$  s. This improved the Bland-Altman coefficient markedly to 0.016 for  $K_{im}$  and 0.028 for  $V_b$ .

Various factors were adjusted in the optimisation algorithm including the function, number of iterations, tolerance and initial parameter. But these were found to have less influence on outcome parameters. Figure 5 summarises the incremental improvement in agreement during each stage of the evaluation as we altered pipeline B. Following all adjustments to pipeline B, the Bland-Altman plots are shown in Fig. 6; the mean differences were  $K_{im} = 9.0 \times 10^{-4} \text{ ml} \cdot \text{cm}^{-3} \cdot \text{min}^{-1}$  and  $V_b = -0.0014$  with



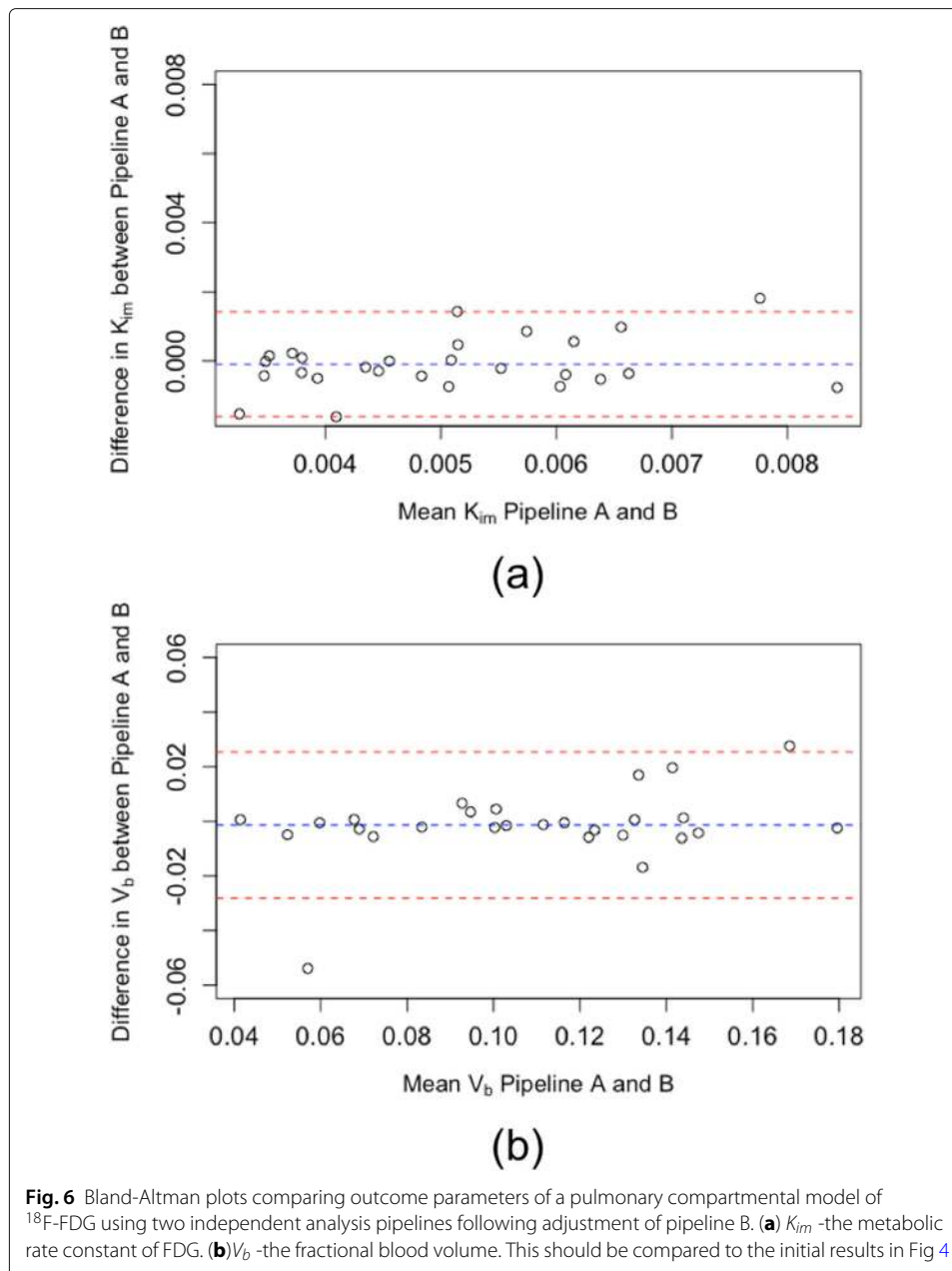


coefficients of reproducibility of 0.0015 and 0.027. The correlation was also improved: for  $K_{im}$ ,  $r = 0.86$  and for  $V_b$ ,  $r = 0.94$ . Figure 3b demonstrates the improvement in agreement on a group level.  $K_{im}$  calculated using the adjusted pipeline B had larger variability than pipeline A in the healthy control group ( $IQR = 0.0031$  vs 0.0022, respectively); we did not observe any notable change in the other groups (see Fig. 3).

## Discussion

We investigated the reproducibility of pulmonary compartmental modelling of inflammation using two independent analysis pipelines;  $K_{im}$  and  $V_b$  were the main outcome parameters of this study,  $K_{im}$  was interpreted as a surrogate for lung inflammation. Two independent operators used reconstructed dynamic PET and CT scans from the same 30 subjects. Initial results showed that the subject-level agreement in  $K_{im}$  and  $V_b$  was poor; despite the application of the same kinetic model (i.e. Eqs. 1 and 2). Interestingly, this did not change the overall interpretation of the group findings; the effect size between groups was comparable using either pipeline. Further, it did not change the outcome of statistical hypothesis testing between groups. Reproducibility of analysis is an important pre-requisite for an imaging biomarker; the results of this evaluation demonstrate the need for standardisation when applying compartmental modelling to assess lung inflammation.

The excellent repeatability of pipeline B ( $COV = 6.1\%$ ) indicates that the poor agreement between the pipelines is likely to be due to the implementation of the compartmental model. To investigate the salient stages responsible for the differences, we used an incremental approach: beginning with the steps involved in generating the TACs and all subsequent stages to estimate  $K_{im}$  and  $V_b$ . Visual comparison of the blood input functions revealed that the largest differences seemed to be in the stages of extracting and modelling the blood input functions. We determined that the salient factors were the blood ROI methodology, input function modelling and time delay estimation. ROI definition led to differences in estimates of concentration due to presence of the PVE -



neither pipeline applied corrections for PVE; it has been suggested this could also help reduce bias in the kinetic parameters estimation [9], but this has yet to be confirmed in human studies. Although adjusting the ROI size did improve the agreement between the pipelines modestly, agreement was improved substantially by the subsequent modelling of the input function.

Modelling the input function provides a means to minimise the noise and spillover problems in PET measurements [36, 37]. Various models exist, common examples include the Feng model [36] and tri-exponential equations, which use regression to estimate the best fit to the blood TAC. In these data, we found the fitting was particularly sensitive to the first few time frames—corresponding to the images with the largest noise content.

There are several proposed methods to calculate the time delay between the blood input function and the tissue of interest. We found large differences in estimates of time delay between pipelines A and B, given its sensitivity to the first few minutes of data collection this is clearly exacerbated if the input function models produce different fits in this region.

By adjusting the blood ROI definitions, the input function fitting method and the time delay calculation in pipeline B, we were able to demonstrate better agreement between the two pipelines (see Figs. 3b and 6). Such parameters are often not reported in the “Methods” section of existing literature; these findings highlight the need for reporting details of analysis methodology to facilitate the reproducibility of outcomes in compartmental models of the lung. Adjustment of pipeline B led to higher variability in  $K_{im}$  in the HV group. It seems likely that this is due to the higher noise content in HV scans, which may have led to more extreme differences in fitting of the IDIF. Future work is needed to determine the importance of this variance in pulmonary compartmental models.

The POB ratio was calculated for each subject using discrete venous blood samples; this allows the conversion of a blood input function to a plasma input function. Although the same factors were used for both pipelines, we recognise that this may cause bias in the outcome variables. We found the POB ratio to be very close to one ( $1.05 \pm 0.03$ ); this agrees with previous findings that  $^{18}\text{F}$ -FDG equilibrates between erythrocytes and plasma nearly instantaneously [38]. Further work should determine whether POB corrections are necessary in quantitative  $^{18}\text{F}$ -FDG lung studies. Previous work has demonstrated that the weighting factors chosen to describe the variance of dynamic PET data affects the outcomes of compartmental modelling [27]; here, the weighting factors were identical in both pipelines and we did not explore the impact of different weighting schemes but acknowledge this could be another key contributor which may improve reproducibility. Respiratory motion leads to both density variations in the lung and inaccuracies in the attenuation correction applied to the PET image; both of these influence quantitative PET and improvements have been suggested by use of an averaged CT scan when available [4]. Since we acquired the CT scans during free breathing, these inaccuracies may have an influence on our quantitative PET data. This study investigated the reproducibility of pulmonary compartmental models of dynamic  $^{18}\text{F}$ -FDG PET/CT; since both operators used the same reconstructed datasets, this does not include within subject biological variability or technical factors (e.g. scanner settings and reconstruction algorithms), which would allow the overall reproducibility of pulmonary  $^{18}\text{F}$ -FDG scans to be evaluated. Nevertheless, with each operator using identical scans, we avoid the systematic variability in  $^{18}\text{F}$ -FDG uptake introduced by differences between hardware, reconstruction and acquisition protocols [39]. We were able to assess the intra-observer agreement for pipeline B (operator B); we found this to have excellent repeatability. We could not explore the inter-observer effect as results from pipeline A, undertaken by operator A, were retrospectively acquired. Yet, we were able to determine which stages of compartmental modelling were responsible for the low concordance observed between the pipelines.

Currently, there is no standard method of assessing pulmonary inflammation using compartmental models of  $^{18}\text{F}$ -FDG and each centre may undertake such analyses using their own bespoke approach. Our findings suggest that despite seemingly similar methodology, individual subject results are sensitive to several factors in the analysis; therefore, care is needed when reporting the exact methods used. In our lung data, we identified that the blood ROI methodology, IDIF modelling and time delay estimation were important

drivers of reproducibility. In light of these findings, we suggest that any forthcoming recommendations for reporting methodology incorporate these key features.

## Conclusions

Reproducibility of pulmonary compartmental modelling of  $^{18}\text{F}$ -FDG PET-CT was evaluated in 30 subjects consisting of three groups: COPD patients,  $\alpha_1$ ATD patients and healthy never smokers. Two operators analysed the imaging data using independent software to estimate the metabolic rate constant of  $^{18}\text{F}$ -FDG and fractional blood volume. Initial comparisons showed good agreement in the overall conclusions at group level; however, subject-level agreement was poor. We identified salient factors in the analysis, which improved the agreement between the two analysis pipelines: blood ROI methodology, input function modelling and time delay estimation. In a field where in-house analyses are commonplace, these findings highlight the need for standardisation in reporting of methods, which will help to improve the reproducibility.

## Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s40658-019-0265-8>.

**Additional file 1:** Online supplement.

## Acknowledgements

We would like to thank all the participants who took part in the original studies and the PET/CT staff at Cambridge University Hospitals NHS Foundation Trust and Invicro, Hammersmith. We would also like to thank Christopher Coello for providing details of Pipeline A and Prof. Roger Gunn for discussions and guidance around the interpretation of the findings.

## Authors' contributions

LV, FJW, SL, IBW and JC contributed to the conception of this work. LV, MF, JC and IBW contributed to the acquisition of data. LV, SL and MF contributed to the data analysis. All authors contributed to the interpretation of the data of the work, drafting and revising the manuscript and have approved the final version for publication. All authors are accountable for all aspects of the work performed.

## Funding

EVOLVE was jointly sponsored by the University of Cambridge and Cambridge University Hospitals NHS Foundation Trust and was funded by a grant from Innovate UK (UK Technology Strategy Board) as a component study of the ERICA (Evaluation of the Role of Inflammation in Chronic Airways Disease) Collaboration, which includes academic and industry partners. GSK, a consortium partner, made contributions towards study management. The views expressed in this publication are those of the authors and not necessarily those of the NHS, The National Institute for Health Research or the Department of Health.

## Availability of data and materials

The datasets used for and generated during the analysis are available on reasonable request from the corresponding author upon completion of a material transfer agreement. MIAKAT<sup>TM</sup> [31] is freely available but requires modification for lung analysis.

## Ethics approval and consent to participate

The studies described in this work were conducted in accordance with Good Clinical Practice standards. The participating sites received a favourable ethical opinion from the local ethics committee and approval from the Administration of Radioactive Substances Advisory Committee. All participants provided informed written consent before enrolment in accordance with the Declaration of Helsinki.

## Competing interests

FJW is an employee and shareholder of GSK and previously a consultant to ECNP R&S, GlaxoSmithKline, IPPEC, King's College London, Lundbeck A/S, Mentis Cura ehf and Pfizer Inc. and has received travel expenses as a guest speaker from Orion Pharma Ltd. IBW is funded by the British Heart Foundation and receives support from the Cambridge NIHR Comprehensive Biomedical Research Centre and Comprehensive Research Network. He has received grant support from GSK. JC is employed by Cambridge University Hospitals NHS Foundation Trust and obligated to spend 50% of his time on GSK clinical trial research via a secondment agreement, representing a significant relationship; however, he receives no benefits or compensation from GSK. JC acknowledges funding support from the Cambridge NIHR Comprehensive Biomedical Research Centre. LV is a PhD student funded by GSK (studentship agreement number: STU100043346/BIDS3000023795), MF declares that she has no competing interests. SL is a consultant to GSK and GE Healthcare.

### Author details

<sup>1</sup>Experimental Medicine and Immunotherapeutics, Department of Medicine, Addenbrookes Hospital, Cambridge, UK. <sup>2</sup>Amallis Consulting LTD, London, UK. <sup>3</sup>GSK R&D, Brentford, UK. <sup>4</sup>Cambridge University Hospitals NHS Trust, Cambridge, UK.

Received: 2 July 2019 Accepted: 25 November 2019

Published online: 16 December 2019

### References

1. Chen D, Kinahan P. Multimodality molecular imaging of the lung. *J Magn Reson Imaging*. 2010;32(6):1409–20. <https://doi.org/10.1002/jmri.22385>. Multimodality.
2. Chen DL, Cheriyan J, Chilvers ER, Choudhury G, Coello C, Connell M, Fisk M, Groves AM, Gunn RN, Holman BF, Hutton BF, Lee S, MacNee W, Mohan D, Parr D, Subramanian D, Tal-Singer R, Thielemans K, van Beek EJR, Vass L, Wellen JW, Wilkinson I, Wilson FJ. Quantification of lung PET images: challenges and opportunities. *J Nuclear Med*. 2017;58(2):201–7. <https://doi.org/10.2967/jnumed.116.184796>.
3. Jones HA, Cadwallader KA, White JF, Uddin M, Peters AM, Chilvers ER. Dissociation between respiratory burst activity and deoxyglucose uptake in human neutrophil granulocytes: implications for interpretation of 18F-FDG PET images. *J Nuclear Med*. 2002;43(5):652–7.
4. Holman BF, Cuplov V, Hutton BF, Groves AM, Thielemans K. The effect of respiratory induced density variations on non-TOF PET quantitation in the lung. *Phys Med Biol*. 2016;61(8):3148–63. <https://doi.org/10.1088/0031-9155/61/8/3148>.
5. Gunn RN, Gunn SR, Cunningham VJ. Positron emission tomography compartmental models. *J Cereb Blood Flow Metab Off J Int Soc Cereb Blood Flow Metab*. 2001;21(6):635–52. <https://doi.org/10.1097/00004647-200106000-00002>.
6. Holman B, Cuplov V, Millner L, Hutton BF, Maher TM, Groves AM, Thielemans K. Improved correction for the tissue fraction effect in lung PET/CT imaging. *Phys Med Biol*. 2015;60(18):7387–402. <https://doi.org/10.1088/0031-9155/60/18/7387>.
7. Coello C, Fisk M, Mohan D, Wilson FJ, Brown AP, Polkey MI, Wilkinson I, Tal-Singer R, Murphy PS, Cheriyan J, Gunn RN. Quantitative analysis of dynamic 18F-FDG PET/CT for measurement of lung inflammation. *EJNMMI Res*. 2017;7(1):47. <https://doi.org/10.1186/s13550-017-0291-2>.
8. Sokoloff L, Reivich M, Kennedy C. The [14C]deoxyglucose method for the measurement of local cerebral glucose utilization: theory, procedure, and normal values in the conscious and anesthetized albino rat. *J Neurochem*. 1977;28(5):897–916.
9. Schroeder T, Vidal Melo MF, Musch G, Harris RS, Venegas JGTW. Modelling pulmonary kinetics of 2-deoxy-2-[18F]fluoro-D-glucose during acute lung injury. *Anat Rec*. 2008;15(6):763–75. <https://doi.org/10.1002/ar.20849.3D>. NIHMS150003.
10. Lambrou T, Groves AM, Erlandsson K, Screation N, Endozo R, Win T, Porter JC, Hutton BF. The importance of correction for tissue fraction effects in lung PET: preliminary findings. *Eur J Nuclear Med Mol Imaging*. 2011;38(12):2238–46. <https://doi.org/10.1007/s00259-011-1906-x>.
11. Carson RE. Precision and accuracy considerations of physiological quantitation in PET. *J Cereb Blood Flow Metab*. 2011;11(1\_suppl):45–50. <https://doi.org/10.1038/jcbfm.1991.36>.
12. De Geus-Oei L-F, Visser EP, Krabbe PFM, van Hoorn BA, Koenders EB, Willemsen AT, Pruim J, Corstens FHM, Oyen WJG. Comparison of image-derived and arterial input functions for estimating the rate of glucose metabolism in therapy-monitoring 18F-FDG PET Studies. *J Nuclear Med*. 2006;47(6):945–9.
13. van der Weerd AP, Klein LJ, Boellaard R, Visser CA, Visser FC, Lammertsma AA. Image-derived input functions for determination of MRGlucose in cardiac (18)F-FDG PET scans. *J Nuclear Medicine Off Publ Soc Nuclear Med*. 2001;42(11):1622–9.
14. Holman B, Hutton B, Thielemans K. Method to determine the voxel-wise blood volume in the lung from dynamic PET data. *J Nuclear Med*. 2017;58(Supplement 1):1310.
15. Schroeder T, Vidal Melo MF, Musch G, Harris RS, Venegas JG, Winkler T. Image-derived input function for assessment of 18F-FDG uptake by the inflamed lung. *J Nuclear Med*. 2007;48(11):1889–96. <https://doi.org/10.2967/jnumed.107.041079>.
16. Vriens D, de Geus-Oei L-F, Oyen WJG, Visser EP. A curve-fitting approach to estimate the arterial plasma input function for the assessment of glucose metabolic rate and response to treatment. *J Nuclear Med*. 2009;50(12):1933–9. <https://doi.org/10.2967/jnumed.109.065243>.
17. Hori Y, Ihara N, Teramoto N, Kunimi M, Honda M, Kato K, Hanakawa T. Noninvasive quantification of cerebral metabolic rate for glucose in rats using 18 F-FDG PET and standard input function. *J Cereb Blood Flow Metab*. 2015;35(10):1664–70. <https://doi.org/10.1038/jcbfm.2015.104>.
18. Guo H, Renaut RA, Chen K. An input function estimation method for FDG-PET human brain studies. *Nuclear Med Biol*. 2007;34(5):483–92. <https://doi.org/10.1016/j.nucmedbio.2007.03.008>. NIHMS150003.
19. Mourik JEM, Lubberink M, Klumpers UMH, Comans EF, Lammertsma AA, Boellaard R. Partial volume corrected image derived input functions for dynamic PET brain studies: Methodology and validation for [11C]flumazenil. *NeuroImage*. 2008;39(3):1041–50. <https://doi.org/10.1016/j.neuroimage.2007.10.022>.
20. Kimura Y, Oda K, Naganawa M, Ishiwata K, Ishii K, Matani A. Extraction of a plasma time-activity curve from dynamic brain PET images based on independent component analysis. *IEEE Trans Biomed Eng*. 2005;52(2):201–10. <https://doi.org/10.1109/tbme.2004.840193>.
21. Zhou S, Chen K, Reiman EM, Li D-m, Shan B. A method for generating image-derived input function in quantitative 18F-FDG PET study based on the monotonicity of the input and output function curve. *Nuclear Med Commun*. 2012;33(4):362–70. <https://doi.org/10.1097/MNM.0b013e32834f262e>.
22. Herscovitch P, Markham J, Raichle ME. Brain blood flow measured with intravenous H<sub>2</sub>(15)O. I. Theory and error analysis. *J Nuclear Med Off Publ Soc Nuclear Med*. 1983;24(9):782–9.

23. Meyer E. Simultaneous correction for tracer arrival delay and dispersion in CBF measurements by the H2150 autoradiographic method and dynamic PET. *J Nuclear Med.* 1989;30(6):1069–78.
24. Van den Hoff J, Burchert W, Müller-Schauenburg W, Meyer GJ, Hundeshagen H. Accurate local blood flow measurements with dynamic PET: fast determination of input function delay and dispersion by multilinear minimization. *J Nuclear Med.* 1993;34(10):1770–7.
25. Alf MF, Martić-Kehl MI, Schibli R, Krämer SD. FDG kinetic modeling in small rodent brain PET: optimization of data acquisition and analysis. *EJNMMI Res.* 2013;3(1):1–14. <https://doi.org/10.1186/2191-219X-3-61>.
26. Wellman TJ, Winkler T, Vidal Melo MF. Modeling of tracer transport delays for improved quantification of regional pulmonary 18F-FDG kinetics, vascular transit times, and perfusion. *Ann Biomed Eng.* 2015;43(11):2722–34. <https://doi.org/10.1007/s10439-015-1327-2>.
27. Yaqub M, Boellaard R, Kroppholler MA, Lammertsma AA. Optimization algorithms and weighting factors for analysis of dynamic PET studies. *Phys Med Biol.* 2006;51(17):4217–32. <https://doi.org/10.1088/0031-9155/51/17/007>.
28. Chen DL, Huang HJ, Byers DE, Shifren A, Belikoff B, Engle JT, Arentson E, Kemp D, Phillips S, Scherrer DE, Fujiwara H, Spayd KJ, Brooks FJ, Pierce RA, Castro M, Isakow W. The peroxisome proliferator-activated receptor agonist pioglitazone and 5-lipoxygenase inhibitor zileuton have no effect on lung inflammation in healthy volunteers by positron emission tomography in a single-blind placebo-controlled cohort study. *PLoS ONE.* 2018;13(2):1–17. <https://doi.org/10.1371/journal.pone.0191783>.
29. Fisk M, Cheriyan J, Mohan D, Forman J, Mäki-Petäjä KM, McEnery CM, Fuld J, Rudd JHF, Hopkinson NS, Lomas DA, Cockcroft JR, Tal-Singer R, Polkey MI, Wilkinson IB. The p38 mitogen activated protein kinase inhibitor losmapimod in chronic obstructive pulmonary disease patients with systemic inflammation, stratified by fibrinogen: a randomised double-blind placebo-controlled trial. *PLoS ONE.* 2018;13(3):1–17. <https://doi.org/10.1371/journal.pone.0194197>.
30. Fisk M, Mohan D, Cheriyan J, Yang L, Fuld J, McEnery CM, Tal-Singer R, Polkey MI, Wilkinson IB. Evaluation of losmapimod in patients with chronic obstructive pulmonary disease (COPD) with systemic inflammation stratified using fibrinogen ('EVOLUTION'): RATIONALE and protocol. *Artery Res.* 2014;8(1):24–34. <https://doi.org/10.1016/j.artres.2013.10.380>.
31. MIAKAT. Version 4.2.6. Hammersmith Hospital: Invicro Ltd; 2010.
32. MATLAB. Version 7.10.0 (R2010a). Natick, Massachusetts: The MathWorks Inc.; 2010.
33. Yushkevich PA, Piven J, Cody Hazlett H, Gimpel Smith R, Ho S, Gee JC, Gerig G. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage.* 2006;31(3):1116–28.
34. Martin Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;327(8476):307–10. [https://doi.org/10.1016/S0140-6736\(86\)90837-8](https://doi.org/10.1016/S0140-6736(86)90837-8).
35. Cunningham VJ, Gunn SR, Gunn RN, Aston JAD, Turkheimer FE. Positron emission tomography compartmental models: a basis pursuit strategy for kinetic modeling. *J Cereb Blood Flow Metab.* 2015;22(12):1425–39. <https://doi.org/10.1097/01.wcb.0000045042.03034.42>.
36. Feng D, Wong K-P, Wu C-M, Siu W-C. A technique for extracting physiological parameters and the required input function simultaneously from PET image measurements: theory and simulation study. *IEEE Trans Inf Technol Biomed.* 2002;1(4):243–54. <https://doi.org/10.1109/4233.681168>.
37. Sari H, Erlandsson K, Barnes A, Atkinson D, Arridge S, Ourselin S, Hutton B. Modelling the impact of injection time on the bolus shapes in PET-MRI AIF Conversion. 2014;1(Suppl 1):4–6. <https://doi.org/10.1186/2197-7364-1-s1-a54>.
38. Nahmias C, Wahl LM, Amano S, Asselin M-c, Chirakal R. Equilibration of 6-[18F]Fluoro-L-m-Tyrosine Between Plasma and Erythrocytes. *J Nuclear Med.* 2000;41(10):1636–41.
39. Doot RK, Scheuermann JS, Christian PE, Karp JS, Kinahan PE. Instrumentation factors affecting variance and bias of quantifying tracer uptake with PET/CT. *Med Phys.* 2010;37(11):6035–46. <https://doi.org/10.1118/1.3499298>. NIHMS150003.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

---