The role of ^{18}F-fluorodeoxyglucose positron emission tomography in the monitoring of inflammatory activity in Crohn's disease

Evangelos A Russo[1,2] PhD, Sameer Khan[3] BSc, Ryan Janisch[4], Roger N Gunn[1,4] PhD, Eugenii A Rabiner[4,5] MD, Stuart A Taylor [6] MD, Paul M Matthews[*7] DPhil, Timothy R Orchard[*2,8] MD.

*Joint senior authors

1. Division of Experimental Medicine, Department of Medicine, Imperial College London

2. Department of Gastroenterology, Imperial College Healthcare NHS Trust

3. Department of Imaging, Imperial College Healthcare NHS Trust

4. Imanova, Centre for Imaging Sciences, London

5. Centre for Neuroimaging Sciences, King's College, London

6. Centre for Medical Imaging, University College London

7. Division of Brain Sciences, Department of Medicine, Imperial College London

8. Division of Hepatology and Gastroenterology, Department of Surgery and Cancer, Imperial College Healthcare NHS Trust

Corresponding authors: 1) Dr Evangelos Russo, Department of Gastroenterology, 3rd floor Salton House, St Mary's Hospital, Praed Street, London W2 1NY, Tel: +44(0)203 3121072, email: e.russo@imperial.ac.uk

Conflicts of Interest and Funding Statement

Dr Russo has received a research grant from the Wellcome Trust and GlaxoSmith Kline, and travel grants from Merck, Warner-Chilcott and Actavis. Prof Matthews has consulted or received honoraria for lectures from GlaxoSmithKline, Biogen, IXICO and Novartis. Prof Gunn has consulted for Abbvie, GlaxoSmithKline and UCB. Prof Orchard has received speaking fees and participated in advisory boards for Merck, Abbvie and Napp.

This work was supported by the Imperial College Healthcare Trust Biomedical Research Centre (ICHT BRC) and an Imperial College Wellcome Trust-GlaxoSmithKline Clinical Training Fellowship in Translational Medicine and Therapeutics (Evangelos Russo- Grant number WMET P30902). GlaxoSmtihKline generously provided scanning time and associated staff support as an additional "in kind' contribution.

Prof S Taylor is a National Institute for Health Research (NIHR) senior investigator.

ABSTRACT AND KEY WORDS

Background: ^{18}Fluorine-fluorodeoxyglucose positron emission tomography (18F-FDG PET) has recently attracted interest for the measurement of disease activity in Crohn's disease (CD). The aim of this study was to assess the utility of FDG-PET as a marker of progression of inflammatory activity and its response to treatment in patients with CD.

Methods: 22 patients with active CD were recruited prospectively to undergo FDG-PET scanning at two time points. All 22 index scans were used to assess sensitivity and specificity against a reference standard MRI measure. Correlations with clinico-pathological markers of severity (Harvey-Bradshaw Index (HBI), C-reactive protein (CRP) and calprotectin) were also performed. 17/22 participated in the longitudinal component and underwent scanning before and 12 weeks after initiation of anti-TNFα therapy. Patients were sub-categorised on the basis of a clinically significant response, and responsiveness of the PET-measures was assessed using previously described indices. 5/22 patients took part in the test-retest component of the study and underwent scanning twice within a target interval of 1 week, to assess the reproducibility of the PET measures.

Results: The sensitivity and specificity of ^{18}F-FDG PET were 88% and 70%, respectively. SUV-related PET measures correlated significantly both with CRP and HBI in cross-sectional and longitudinal analyses. (G)SUV_{MAX} and (G)SUV_{MEAN} demonstrated favourable responsiveness and reliability characteristics (RRG >0.80 and %VAR <20%) compared to volume-dependent FDG-PET measures. A proportion of the FDG signal (10-30%) was found to originate from the lumen of diseased segments.

Conclusions: ^{18}F-FDG PET may be useful for longitudinal monitoring of inflammatory activity in CD.

Key words: Crohn's, biomarker, FDG-PET, monitoring

INTRODUCTION

Crohn's disease (CD) is a chronic idiopathic disorder affecting the gastro–intestinal (GI) tract. Its characteristic features are those of segmental inflammation affecting any part of the gut, with lesions extending across the full thickness of the bowel wall[1]. At presentation, a majority of lesions are of the inflammatory sub-type. These lesions are dynamic; over time, an increasing proportion develop a stricturing or penetrating character[2], which may require hospitalization and surgery. Crohn's lesions also have the potential to regress partially or completely. These patterns of progression are not reflected reliably in patients' symptoms; many patients progress to structuring or fistulising disease despite minor or relapsing and remitting symptoms[3,4]. While monitoring the natural history of mucosal disease is facilitated by endoscopy and histology, the progression or regression of transmural lesions remains difficult to characterize *in vivo*.

Objective markers of disease activity are vital both for clinical decision making (e.g., for timely escalation of therapy) and for new drug development. Biomarkers currently in use in clinical practice (e.g., C-reactive protein [CRP] and faecal calprotectin) lack sensitivity and specificity[5,4]. Recently, imaging biomarkers, such as the Magnetic Resonance Index of Activity (MaRIA) have been validated against colonoscopy as a measure of mucosal and transmural disease load in cross sectional studies [6–8]. However a significant unmet need remains for non-invasive longitudinal biomarkers reflecting disease activity.

Recently, there has been an interest into the potential of fluorine-18 fluorodeoxyglucose positron emission tomography (^{18}F-FDG PET) for non-invasively detecting and monitoring lesions of CD[9]. Early studies show that FDG-PET can identify and quantify individual inflammatory foci, as well as global inflammatory load, and that the technique has sensitivity and specificity for lesion identification comparable to conventional

markers of activity (e.g., Crohn's disease endoscopic index of severity (CDEIS), clinical scoring, blood and faecal inflammatory markers)[10]. The majority of these early studies characterized lesions using maximal Standardised Uptake Values (SUV) in involved gut segments. More recent reports also have explored composite measures that additionally reflect signal volume[11,12]. The potential of FDG-PET to demonstrate changes in lesion activity over time was first suggested in a report which included 3 CD patients[13]. However, there have been no systematic prospective studies with defined scanning intervals to further investigate this. Several well-designed experiments in appropriate murine models[14–16], as well as human studies in other chronic inflammatory conditions[17,18] have produced encouraging results, but a longitudinal and test-retest reliability data are lacking.

The mechanism for increased FDG uptake along inflamed intestinal segments has not been evaluated directly. Murine models suggest that the most FDG-avid cellular populations in inflammatory foci are neutrophils, macrophages and lymphocytes [14,15], leading to an inference that increased FDG uptake in CD reflects the aggregation of inflammatory cells in tissue with active disease. However other elements in the intestinal lumen (e.g., microbiota) also accumulate FDG and could contribute to the signal[19].

The aim of this study was to assess the potential of ~~^{18}F-FDG PET to act~~ intestinal FDG uptake as a biomarker of disease activity in patients with CD. We explored the modality's content validity, that is, the ability of the test to measure the parameter intended, in this case, inflammatory activity. This was achieved by testing for correlations between PET and other validated measures of the disease process, including the Harvey-Bradshaw Index (HBI) of activity, the C-reactive protein (CRP), and faecal calprotectin (FC) concentrations.

We also investigated the responsiveness of FDG-PET measures in the detection of longitudinal changes in bowel inflammation following anti-TNFα treatment, defined as their ability to detect a concurrent clinically meaningful change. The test-re-test reliability (i.e., the stability of these endpoints when assessed under conditions in which stable disease can be assumed) was also explored to enable a proper assessment of the utility of FDG-PET for longitudinal monitoring of clinical status[20].

This study tested two primary hypotheses. We hypothesised that active CD lesions would show a stable FDG signal over short periods of time, in patients without changes in their symptoms or treatment regime. In addition, we hypothesised that clinical responses to anti-TNFα would be associated with decreases in FDG signal.

MATERIALS AND METHODS

Patient recruitment

Patients with active CD were prospectively recruited from the Inflammatory Bowel Disease clinics at four participating hospitals in (BLIND). Recruitment took place between November 2012 and May 2014. Recruited patients were aged over 18 with active Crohn's disease, determined on the basis of a Harvey Bradshaw Index (HBI) of >4 and either a positive CRP (>5mg/L) or faecal calprotectin (>100μg/g). For the primary study, patients also were limited to those who were due to commence biologic therapy with anti-tumour necrosis factor alpha antibody (anti-TNF). Anti-TNF naive, and patients commencing a second-line agent were included. For the test-retest reliability sub-study, patients with active CD and no change in therapy over a 4- week period were recruited.

Participants with alternative GI pathologies which could produce additional foci of PET signal (e.g. known polyps, malignancy, GI infection, coeliac disease) were excluded. We also excluded patients with diabetes and those with physical disabilities precluding comfortable positioning on the PET scanner, pregnant women, those with a history of cancer, either personal or in a first degree relative at an age <55 years, and also patients who had participated in research which involved ionising radiation in the previous 3 years.

Consecutive eligible participants from the 4 centres were approached with a study information sheet.

Study design

For the treatment responsiveness analysis, two PET scans were performed under standardised conditions, one at baseline before and one at follow-up after anti-TNF

therapy given at standard doses. The target interval between the start of therapy and the follow-up scan was 3 months. Reference investigations performed on the day of both scans included HBI scoring, CRP and calprotectin (Calpro AS, Norway), as well as Magnetic Resonance Imaging. Three participants in this arm also participated in a sub-study exploring mechanisms determining the FDG signal (see below).

To assess test-retest reproducibility, the target interval between the two scans was 1 week, with a maximum permitted interval of 2 weeks. The HBI score, CRP and a MaRIA score were assessed on both visits, while a calprotectin assay was performed on one of the two visits.

The study was designed as an exploratory pilot and not formally powered.

PET scanning

Scanning was performed at Imanova Centre for Imaging Sciences, London UK. All participants were asked to avoid strenuous exercise for 24 hours and were fasted for 6 hours prior to each scan. On arrival in the imaging suite, a 20 gauge intravenous cannula was placed and capillary blood glucose was checked (Accu-check, Roche, UK). A target dose of 185MBq of ^{18}F-FDG in 10ml of normal saline was administered intravenously, followed by a saline flush. The patients then were asked to rest semi-recumbently in a quiet room, and, in keeping with similar protocols in the literature [10,11,21–24] , they were given between 800-1200ml of an oral bowel distension, agent. 2.5% mannitol was used according to local protocol in our institution. The volume of mannitol offered to each participant prior to the follow-up scan was the same as that consumed before the index scan. 50 minutes following FDG administration, patients were asked to void their bladders, and were positioned on the PET/CT scanner (Siemens Biograph 6 Truepoint,

Siemens Healthcare, Erlangen, Germany). A bolus of 20mg of intravenous hyoscine butylbromide (Buscopan®) was administered to reduce gut motility, and a low dose CT scan of the abdomen was obtained (130KV, 30mAs, pitch 1.5, 6 slice x 3mm collimation). Subsequently, PET emission data from the gut were acquired and reconstructed using a 3-dimensional model. A maximum of 3 bed positions were used for a total scanning time duration of 30 minutes (10 minutes per bed position). The estimated total radiation dose for both scans was 11.2mSv.

To evaluate the compartmental distribution of the focally increased PET FDG signal with CD, 3 patients underwent an additional evaluation. After completing the original 30-minute period of PET signal acquisition, the images were reconstructed and qualitatively reviewed by one of the investigators (BLIND) for the presence of focally increased signal in the recto-sigmoid region. Once this was confirmed, the patient was administered a phosphate enema and, following evacuation of bowel contents, asked to return to the scanner. A repeat, limited low-dose CT scan of the pelvis was performed for localisation, using the parameters described above, followed by an additional 10-minute PET acquisition over a single pelvic bed position was obtained for 1 patient. For the latter 2 patients, we measured radioactivity (RA) in the eliminated bowel contents directly and, following decay correction, expressed the bowel content signal as a proportion of the original total PET FDG signal from the recto-sigmoid in the baseline scan.

MRI Scanning

MRI scanning took place either immediately before or after the PET scan. Patients were asked to ingest an additional 400-1000ml of 2.5% mannitol prior to the second scan according to tolerance. Subsequently, patients were positioned supine on the MRI

Scanner (Siemens Verio 3T). Two surface receiver coils were placed over the abdomen and pelvis. Acquisition sequences are summarised in Table S1 (supplemental material).

PET scan measurements

PET/CT images were analysed using Inveon Research Workplace (Siemens Healthcare, USA). Scans were reconstructed with OSEM (Ordered Subset Expectation Maximisation, 2 iterations, 8 subsets) with a 256 x 256 matrix, zoom of 1.3 and 3D Gaussian image filter with 5 mm full-width at half maximum (FWHM).

The attenuation corrected (AC) CT sequence was fused with OSEM PET sequences to enable Volume of Interest (VOI) delineation. All VOI definition and subsequent analysis were carried out independently by a fellow [BLIND], unblinded to the clinical data but blinded to the MRI measurements, and a nuclear medicine expert [BLIND] who was blinded to all clinical information and MRI measurements. In cases of discrepancies in VOI positioning and margins, VOIs were reviewed to reach a consensus.

A VOI of at least 40cm^3 was created inside the liver parenchyma and the mean SUV ($LivSUV_{MEAN}$) was recorded (Figure 1A). In keeping with published literature, the resulting $LivSUV_{MEAN}$ was used as a threshold above which GI signal was classified as abnormal[25,21]. The entire fused 3-D sequence was thus adjusted to exclude signal of intensity lower than $LivSUV_{MEAN}$ (Figure 1B). The bowel region on the image was then separated into seven segments (small bowel, terminal ileum (TI), caecum and ascending colon, transverse colon, descending colon, sigmoid colon and rectum). TI was defined as the most distal small intestinal focus, regardless of its length, provided it was within 5cm of the ileo-caecal valve (all signal foci proximal to this were classified as small bowel). Initial VOIs were defined manually in those segments which contained visible signal foci,

to include all of the area of abnormal signal (Figure 1C). This initial VOI was then re-processed to exclude those voxels within that initial VOI with sub-threshold signal < $LivSUV_{MEAN}$ (Figure 1D).

Figure 1

The Standardised Uptake Value (SUV) is the typical measure of PET signal:

(1) $SUV = r/(\alpha/w)$

where r is the radioactivity concentration in KBq/ml within a VOI, α is the radioactive decay corrected activity of administered FDG and w is the weight of the patient [26]. For each final VOI, 3 measures were made: SUV_{MAX}, reflecting the SUV of the single voxel of highest intensity, SUV_{MEAN}, the average value of SUV across all voxels in the VOI, and the total VOI volume in cm^3. From these 3 values, a Segmental Lesion Glycolysis Index (SLG) was derived, on a 'per segment' basis:

(2) SLG= SUV_{MEAN} x Volume of VOI

In instances where more than one focus of activity was observed in a single bowel segment, these were expressed as a single VOI using the method above.

Additional FDG-PET measures for each subject included the Global SUV_{MAX} ($GSUV_{MAX}$), which was calculated as the average SUV_{MAX} in all gut segments with supra-threshold signal, the Total Volume, which represented the sum of all segmental volumes, the Total Lesion Glycolysis (TLG), and the Global SUV_{MEAN} ($GSUV_{MEAN}$):

(3) $TLG = \sum SLGs$

(4) $GSUV_{MEAN}$= TLG/ Total Volume

Segmental SUV_{MAX} and SUV_{MEAN} were interpreted as reflecting the severity of disease activity local to that segment, while $GSUV_{MAX}$ and $GSUV_{MEAN}$ were explored as measures reflecting global disease activity in a patient. Each final VOI's Volume (V) as well as the Total V were used as measures of disease extent. The derived measures of SLG and TLG were used as composite measures assuming that the signal reflected metabolically active uptake of the radiotracer by the tissue.

There were several instances when a segment only expressed signal above threshold on either the baseline or the follow-up scan. In those instances, PET data were extracted from the scan with the active segment as described above, but when analysing the scan that contained the segment with no activity, an anatomical VOI of equal volume was placed on the 'normal' segment most closely matching the location of the corresponding "abnormal" VOI on the related scan. All PET endpoints were then calculated as described above.

MRI scan measurements

MRI examinations were assessed by a sub-specialised radiologist (BLIND) blinded to participants' other clinical and PET scan details. A MaRIA was scored for each bowel segment (details in supplemental methods) and used as a reference standard for the segmental location of active disease. A segment was considered positive on MRI if it had a MaRIA > 7.0 [6].

Statistical analyses

Data was analysed using Prism 6.0 (Graphpad, San Diego, CA, USA). Correlations between clinico-pathological markers (HBI, CRP and FC) and PET endpoints using the baseline PET

scan for each participant from both cohorts were performed using two-tailed Spearman rank coefficients ($p<0.05$). In this exploratory analysis, p-values were not corrected for multiple comparisons. Sensitivity and specificity were estimated using all individual bowel segments from the baseline PET scan for each participant, against the MRI reference standard. We tested related hypotheses that, in patients with active CD, the clinical score, CRP and FC would be explained by either the global inflammatory activity, or activity within the worst affected bowel segments, so the correlations above were performed using global, as well as segmental PET endpoints, respectively. The most abnormal segment in each patient was determined on the basis of its SUV_{MEAN}. For an analysis of responsiveness of the PET measures to clinical treatment responses, we assessed 13/17 subjects who successfully completed both scans before and after the introduction of anti-TNF therapy. In a parametric analysis, we assessed correlations between the absolute differences (Δ) in the clinico-pathological endpoints and the corresponding (Δ)PET endpoints. For all these correlations, the Spearman rank correlation coefficient was also used. We also performed an ordinal analysis after classifying patients as either responders or non-responders on the basis of the clinical response 3 months after the initiation of therapy (a decrease in HBI of $\geq$3 points)[27]. Baseline and follow-up PET measures were compared for responders and non-responders. This distinction between responders and non-responders was used to calculate the Responsiveness ratio of Guyatt (RRG) and the Standardised Effect Size (SES) [28]:

$$(5)\ RRG = \Delta PET\ endpoint_{RESPONDERS} / SD\ endpoint_{NON\text{-}RESPONDERS}$$

$$(6)\ SES = \Delta PET\ endpoint_{RESPONDERS} / SD\ endpoint_{RESPONDERS}$$

Ratios > 0.80 suggest high responsiveness

To assess test-retest reliability of PET-endpoints, we measured the % variability (%VAR) for each endpoint using all segments with signal > $LivSUV_{MEAN,}$ as previously described[29]:

$$(7)\ \%VAR = 2 \times (Value_{2} - Value_{1}) / (Value_{1} + Value_{2}) \times 100\%$$

ETHICAL CONSIDERATIONS

Local ethics approval was granted for the protocol, and all participants signed informed consent (approval 12/LO/1018).

RESULTS

Demographics

22 patients (13 male) with an average age of 40 years (range 22-59) were recruited (Table 1)[30]. 17/22 participated in the longitudinal monitoring component of the study and 5/22 were used for the test-retest reliability sub-study.

Table 1

PET scans

Baseline cross-sectional analysis

A total of 145 gut segments were assessed by a baseline PET Scan in 22 patients. MaRIA was available on 139. 52/139 segments were positive for disease and 87/139 were negative as per MaRIA criteria. Segmental FDG PET signal ≥$LivSUV_{MEAN}$ identified 46/52 of the MRI-positive segments, resulting in a sensitivity of 88% for the FDG-PET. 61/87 MRI negative segments had signal of intensity less than $LivSUV_{MEAN}$, for a PET specificity of 70%.

Figure ~~2~~ S1

Table ~~2~~ S2

Cross-sectional correlations of baseline PET endpoints with clinico-pathological scores

Spearman rank correlation coefficients were performed to test correlations between each of the clinico-pathological variables and the FDG-PET endpoints (Table ~~3~~ 2). The HBI correlated significantly with $GSUV_{MAX.}$ CRP showed a statistically significant correlation with all PET outcome measures except for SLG. Calprotectin did not correlate with any of the described PET measures.

Table ~~3~~ 2

Longitudinal correlations of ΔPET endpoints with clinico-pathological scores

We assessed the correlations of the absolute change (Δ) of HBI and PET measures. While HBI itself only correlated with $GSUV_{MAX}$, ΔHBI correlated significantly with PET measures other than ΔSLG of the single most inflamed segment. A longitudinal correlation between CRP and PET measures was found for SUV_{MAX}, $GSUV_{MAX}$ and $GSUV_{MEAN}$ only. Similar to the cross-sectional analysis, Δcalprotectin did not correlate with any of the ΔPET outcome measures. Results are outlined in Table ~~4~~ 3.

Table ~~4~~ 3.

Contrasting FDG-PET measures in responders versus non-responders

9/13 patients who completed PET scanning both at baseline and a median [range] of 12 [11-18] weeks after initiation of anti-TNF treatment had favourable clinical responses [27] (median [range] responder ΔHBI = 5 [3-9]; non-responder ΔHBI = 1[-1-2]),(Table 1). Differences at 3 months with treatment were statistically significant for all PET endpoints in the responder relative to the non-responder group (Table ~~5~~ S3 and Figure 2 ~~s 3, 4, 5~~).

Table ~~5~~ S3
Figure ~~3~~ 2
~~Figure 4~~
~~Figure 5~~

Of the 26 segments that were positive on the baseline PET scan in responders, 23/26 showed improvement in SUV_{MAX} and SUV_{MEAN}, and 22/26 in SLG in the follow-up scan. In contrast, of the 14 segments that were positive on the baseline scan in non-responders, reductions in SUV_{MAX}, SUV_{MEAN} and SLG endpoints were seen in 3/14, 4/14 and 5/14 respectively.

Figure ~~6~~ 3 shows characteristic appearances of baseline and follow-up segmental signals in responders and non-responders.

Figure ~~6~~ 3

Responsiveness ratios

The derived measures of responsiveness, the Guyatt responsiveness ratio (RRG) and the standardized effect size, (SES) for each outcome measure are summarised in Table ~~6~~ 4. Both segmental and global SUV measures all had values for both indices that were over the 0.80, which is considered as indicative of good responsiveness for an evaluative instrument [28].

Table ~~6~~ 4

Test-retest repeatability

5 patients (1 female) with active CD were recruited for this arm of the study, and 4 completed the two scans within a median [range] period of 7[4-12] days. The repeatability of the PET measures were estimated in segments with a signal focus > $LivSUV_{MEAN}$ ~~(Table 7)~~. SUV–related endpoints had superior repeatability profiles than volume-dependent measures: Mean ± St. dev %VAR for SUV_{MAX} /$GSUV_{MAX}$ were 20±16.6 and 13±20.7 respectively. For SUV_{MEAN}/$GSUV_{MEAN}$ the values were 10±6 and 2±2.3 respectively whereas for SLG/TLG 66±55.5 and 51±45.3 respectively.

Table ~~7~~ S4

Figure ~~7~~ S2

Exploration of compartmentalization of the increased signal within an inflammatory focus

The total radioactivity (RA) in KBqs in the whole of the recto-sigmoid in the pre-and post-enema scans were compared as described in the methods for 1 subject. The recto-sigmoid radioactivity in that participant was reduced by 18% after the enema. For the remaining 2 participants the eliminated bowel contents were placed in the scanner for 10 minutes and their total radioactivity measured and expressed as a percentage of the recto-sigmoid

activity: Values of percentage activity voided in bowel contents of 10 and 30%, were measured for the two subjects in this way. All measured radioactivity for the post-enema scan and bowel contents were decay-corrected to the time of the pre-enema PET acquisition ~~(~~Table ~~8~~ S5) .

Table ~~8~~ S5

DISCUSSION

^{18}FDG-PET has been proposed as an additional non-invasive method for assessing both segmental and global inflammatory activity in patients with CD. Animal experiments provide evidence supporting the hypothesis that FDG signal is produced by uptake in inflammatory cell infiltrates in gut tissue[10–12] . Previous studies in Crohn's patients suggest that ^{18}F-FDG PET has a high sensitivity and moderately high specificity in demonstrating bowel segments with active Crohn's disease [25,21,31,32]. Moreover good correlations between PET measures and endoscopic and histo-pathological reference standards have been shown [10–12,33] . However, ~~no studies to date examined~~ the utility of FDG-PET to monitor longitudinal changes in bowel inflammation in CD has only been examined in a small study of three patients[13].

In addition, we confirm significant cross-sectional and longitudinal correlations between clinico-pathological markers of disease, most notably CRP, and several PET measures which further contribute in establishing the modality's validity.

While the sensitivity of FDG-PET to identify inflamed bowel segments was high (88%) the specificity was more modest at 70%. This phenomenon was investigated by Louis et al., who suggested that a large proportion of 'false positive' lesions on endoscopy may have

contained sub-endoscopic features of activity either on histology or involving deeper bowel layers[10].

More importantly however, to our knowledge, this is the first prospective study to attempt a comprehensive evaluation of responsiveness and reliability of all FDG-PET measures in assessing therapeutic response to biologic therapy. ~~tracking disease activity over time.~~ Both of these measurement characteristics are important to define in the evaluation of the PET signal as a biomarker inflammatory activity [20].

We have demonstrated that SUV-related outcome measures of the FDG-PET signal are reliable and responsive endpoints in monitoring disease activity over time. SUV_{MAX} showed good measurement characteristics with responsiveness ratios >0.80. In addition to the significant responsiveness figures, the test-retest %variability of 20% is also favourable. Global SUV_{MAX}, which represents the average SUV_{MAX} in all bowel segments with supra-threshold FDG uptake, performed comparably to the segmental measure. Similarly, SUV_{MEAN} and Global SUV_{MEAN} were also shown to be responsive outcome measures with good correlations to CRP. Moreover, their within-patient test-retest %variabilities also were low.

By contrast, SLG and TLG had poor measurement characteristics for monitoring of CD activity. The cross sectional assessment in all 22 patients showed only a modest correlation between TLG and clinico-pathological parameters. Moreover, the test-retest reproducibility of the measures was very poor, which was attributable to the large variability in the volume of the signal between visits. This study used thorough standardization of scanning parameters and bowel distention protocols. This, and the significantly lower test-retest variability of the signal intensity measures (SUV_{MAX} and

SUV_{MEAN}) suggest that the poor test-retest performance of SLG was not due to methodological shortcomings that could be improved in obvious ways. ~~One explanation is that Crohn's lesions are more than originally believed, with their cellular inflammatory populations changing significantly even within days. This could account for the high variability in the signal volume even within the short scanning interval of one week.~~ Until these variations are better understood, given the poor reliability of these two endpoints on test-retest, these results do not support clinical applications of volume-dependent PET endpoints for the characterisation and monitoring of CD activity. Other studies on SLG and TLG in CD also have demonstrated a large inter-subject variability[11] and lack of correlation to regional histological[11] or endoscopic measures[12] of disease activity.

The favourable responsiveness and repeatability profiles of SUV-related PET measures and their significant correlations to clinico-pathological markers of activity suggest that they are potentially useful candidate biomarkers for ~~monitoring of disease activity and~~ assessing response to treatment in CD. The value of these measures in this context has also been demonstrated in other inflammatory disorders such as sarcoidoisis[34] and rheumatoid arthritis[35]. Interestingly, work from the rheumatological literature has gone a step further by demonstrating that early response of SUV within two weeks of treatment initiation can predict clinical response up to 22 weeks into therapy [17,18]. Based on this, and the results presented here we believe that (G)SUVmax and (G)SUVmean should be evaluated as potential markers of clinical response in CD, in preference to systemic or total lesion glycolysis.

Another interesting and novel finding of this study was the relatively high proportion (mean 20%) of the total activity that was intra-luminal in the two patients studied before and after elimination of their bowel contents. Barrier function impairment in diseased

segments can potentially explain the presence of radioactivity within their lumen. It is already known that epithelial tight junctions are disturbed in active inflammatory bowel disease lesions. While this has been studied more in the context of bacterial translocation and immune sensitisation[36], this disruption also could result in FDG escape from the extracellular space to the lumen. However, further work is needed as this is not the only possible mechanism. Alternative mechanisms include, for example, increased intra-luminal shedding of epithelial and inflammatory cells with the enema. Whatever the mechanism, the association of increased PET signal from intraluminal FDG and diseased gut segments suggests that the phenomenon is a consequence of the disease process.

Several methodological issues had to be addressed in developing our study design. First, a suitable gold-standard comparator was not readily available. We chose not to rely on endoscopy because of its inability to inform on disease load in small bowel proximal to the terminal ileum or within tissue planes deeper than the mucosa. A MRI measure such as MaRIA possibly circumvents these issues. Multi-centre validation of longitudinal data on its performance is still lacking. As such, lack of an appropriate reference investigation with a longitudinal track record ~~thus~~ necessitated the distinction between responders and non-responders on the basis of a clinically meaningful response, i.e. an HBI drop of 3 or more between the baseline and follow-up visits [27]. In planning future studies, an alternative strategy could be the use of a consensus panel (including other markers such as CRP and FC) to determine the response status of each individual patient.

~~A source of bias in this 'open-label' study could be an over-reporting of symptomatic improvement due to anti-TNF therapy by patients, leading a larger number to meet HBI criteria and qualify as 'responders'.~~ A source of bias in the study could be due to a placebo response to therapy, leading to reporting of symptomatic improvement when

there is no objective biological improvement. In the original randomized controlled trials a placebo response rate of 20-30% was reported[37,38], but sub-analyses demonstrated that this was smallest, and the response to anti-TNF was more pronounced in those with objective markers of inflammation[39]. Objective markers of activity in addition to a high HBI was one of our inclusion criteria, and, moreover, the improvements in clinical score was paralleled by improvements in inflammatory markers such as CRP and FC (shown in Table 1), making the likelihood of significant placebo effect and therefore misclassification of patients small.

The choice of time intervals between baseline and follow-up scanning was made on the basis that cellular composition, and therefore metabolic activity, was expected to remain largely unchanged during the target interval of 1 week, which was therefore chosen for the test-retest evaluation. For the longitudinal cohort, a period of 3 months was opted for, anticipating that any effect of the anti-TNF at the cellular level would have materialised, Furthermore, this interval provides a reasonable period over which to make a clinical differentiation between responders and non-responders.

Another important issue is that of physiological FDG uptake from the gut. Torihara et al have highlighted this issue as a possible confounder of bowel activity in a study of 61 participants without known bowel disease[40]. The frequency of signal foci with intensity equal to or higher that of the liver was notable (approximately 10% of the examined segments, exclusively in the colon). The use of oral mannitol as a bowel distension agent in our protocol may have also contributed to a diffuse artefactual FDG signal. However, the actual SUV_{MAX} of such foci was significantly lower in comparison with our and other investigators' measurements in foci within diseased segments. Moreover, in the context of a longitudinal scanning, where each participant acts as their own control in either a

test-retest or a pre and prost treatment setting, an element of a minor superimposed physiological FDG signal is probably of little consequence.

The FDG dose of 185MBq was selected as the lowest dose that has provided meaningful data in a CD study [24]. Optimisation of radiation exposure was the reasoning behind the use of a low-dose CT for attenuation correction and segment localization, instead of a full-dose CT enterography protocol. If FDG-PET scanning does find a clinical niche in the assessment of therapeutic response in ~~monitoring of~~ CD patients, then more research will be required to guide further reductions of radiation exposure. PET-MRI scanners provide a promising alternative, especially the newer generation of instruments that provide increased sensitivity as a result of the SiPM detectors and the ability to improve motion correction with simultaneous MR acquisitions [41].

In conclusion, we believe that this study, although small, provides a first comprehensive evaluation of FDG-PET for assessing early pharmacodynamic response to newly initiated anti-TNF therapy ~~tracking inflammatory activity over time~~. Characteristics of the modality such as its limited availability, high cost and appreciable radiation burden render it an unlikely universal, first-line disease quantifying instrument for routine clinical practice. These early findings however suggest the need for larger studies to define specific indications for dual time-point quantification of disease burden, e.g. in the evaluation or even prediction of efficacy of medical treatments or as a robust surrogate marker in early-stage drug development.

ACKNOWLEDGMENTS

We thank the study participants for their patience and collaboration.

We thank all the staff of Imanova, in particular, Will Hallett and Nicholas Keat, Allan Listanco, Yvonne Lewis, Michelle Cunneen, James Anscombe, Mark Tanner, and Ineke de Meer for their excellent support to the study.

Authors' contributions

Evangelos Russo, Tim Orchard and Paul Matthews carried out the literature search. Data collection was performed by Evangelos Russo and Ryan Janisch. Evangelos Russo, Sameer Khan, Stuart Taylor, Tim Orchard, and Paul Matthews performed the data analysis. All authors participated in the study design, data interpretation and writing of the manuscript. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

REFERENCES

1. Boyapati R, Satsangi J, Tzer Ho G. Pathogenesis of Crohn's disease. F1000Prime Rep. 2015;7:44.
2. Cosnes J, Gower-Rousseau C, Seksik P, et al. Epidemiology and natural history of inflammatory bowel diseases. Gastroenterology 2011;140:1785–94.
3. Regueiro M, Kip KE, Schraut W, et al. Crohn's disease activity index does not correlate with endoscopic recurrence one year after ileocolonic resection. Inflamm. Bowel Dis. 2011;17:118–26.
4. Sipponen T, Savilahti E, Kolho K-L, et al. Crohn's disease activity assessed by fecal calprotectin and lactoferrin: correlation with Crohn's disease activity index and endoscopic findings. Inflamm. Bowel Dis. 2008;14:40–6.
5. Magro F, Sousa P, Ministro P. C-reactive protein in Crohn's disease: how informative is it? Expert Rev. Gastroenterol. Hepatol. 2014;8:393–408.
6. Rimola J, Rodriguez S, García-Bosch O, et al. Magnetic resonance for assessment of disease activity and severity in ileocolonic Crohn's disease. Gut 2009;58:1113–20.
7. Rimola J, Ordás I, Rodriguez S, et al. Magnetic resonance imaging for evaluation of Crohn's disease: validation of parameters of severity and quantitative index of activity. Inflamm. Bowel Dis. 2011;17:1759–68.
8. Ordás I, Rimola J, Rodríguez S, et al. Accuracy of magnetic resonance enterography in assessing response to therapy and mucosal healing in patients with Crohn's disease. Gastroenterology 2014;146:374–82.e1.

9. Shyn PB. 18F-FDG positron emission tomography: potential utility in the assessment of Crohn's disease. Abdom. Imaging 2012;37:377–86.

10. Louis E, Ancion G, Colard A, et al. Noninvasive assessment of Crohn's disease intestinal lesions with (18)F-FDG PET/CT. J. Nucl. Med. 2007;48:1053–9.

11. Jacene H a, Ginsburg P, Kwon J, et al. Prediction of the need for surgical intervention in obstructive Crohn's disease by 18F-FDG PET/CT. J. Nucl. Med. 2009;50:1751–9.

12. Saboury B, Salavati A, Brothers A, et al. FDG PET/CT in Crohn's disease: correlation of quantitative FDG PET/CT parameters with clinical and endoscopic surrogate markers of disease activity. Eur. J. Nucl. Med. Mol. Imaging 2014;41:605–14.

13. Spier BJ, Perlman SB, Jaskowiak CJ, et al. PET/CT in the evaluation of inflammatory bowel disease: Studies in patients before and after treatment. Mol. Imaging Biol. 2010;12:85–88.

14. Brewer S, McPherson M, Fujiwara D, et al. Molecular imaging of murine intestinal inflammation with 2-deoxy-2-[18F]fluoro-D-glucose and positron emission tomography. Gastroenterology 2008;135:744–55.

15. Yamato M, Kataoka Y, Mizuma H, et al. PET and macro- and microautoradiographic studies combined with immunohistochemistry for monitoring rat intestinal ulceration and healing processes. J. Nucl. Med. 2009;50:266–73.

16. Bettenworth D, Reuter S, Hermann S, et al. Translational 18F-FDG PET/CT imaging to monitor lesion activity in intestinal inflammation. J. Nucl. Med. 2013;54:748–55.

17. Elzinga EH, Laken CJ van der, Comans EFI, et al. 18F-FDG PET as a tool to predict the clinical outcome of infliximab treatment of rheumatoid arthritis: an explorative study. J. Nucl. Med. 2011;52:77–80.

18. Roivainen A, Hautaniemi S, Möttönen T, et al. Correlation of 18F-FDG PET/CT assessments with disease activity and markers of inflammation in patients with early rheumatoid arthritis following the initiation of combination therapy with triple oral antirheumatic drugs. Eur. J. Nucl. Med. Mol. Imaging 2013;40:403–10.

19. Shreve PD, Anzai Y, Wahl RL. Pitfalls in oncologic diagnosis with FDG PET imaging: physiologic and benign variants. Radiographics 1999;19:61–77; quiz 150–1.

20. Hobart JC, Lamping DL, Thompson AJ. Evaluating neurological outcome measures: the bare essentials. J. Neurol. Neurosurg. Psychiatry 1996;60:127–30.

21. Das CJ, Makharia G, Kumar R, et al. PET-CT enteroclysis: a new technique for evaluation of inflammatory diseases of the intestine. Eur. J. Nucl. Med. Mol. Imaging 2007;34:2106–14.

22. Ahmadi A, Li Q, Muller K, et al. Diagnostic value of noninvasive combined

fluorine-18 labeled fluoro-2-deoxy-D-glucose positron emission tomography and computed tomography enterography in active Crohn's disease. Inflamm. Bowel Dis. 2010;16:974–81.

23. Groshar D, Bernstine H, Stern D, et al. PET/CT enterography in Crohn disease: correlation of disease activity on CT enterography with 18F-FDG uptake. J. Nucl. Med. 2010;51:1009–14.

24. Shyn PB, Mortele KJ, Britz-Cunningham SH, et al. Low-dose 18F-FDG PET/CT enterography: improving on CT enterography assessment of patients with Crohn disease. J. Nucl. Med. 2010;51:1841–8.

25. Meisner RS, Spier BJ, Einarsson S, et al. Pilot study using PET/CT as a novel, noninvasive assessment of disease activity in inflammatory bowel disease. Inflamm. Bowel Dis. 2007;13:993–1000.

26. Kinahan PE, Fletcher JW. Positron emission tomography-computed tomography standardized uptake values in clinical practice and assessing response to therapy. Semin. Ultrasound. CT. MR 2010;31:496–505.

27. Vermeire S, Schreiber S, Sandborn WJ, et al. Correlation between the Crohn's disease activity and Harvey-Bradshaw indices in assessing Crohn's disease severity. Clin. Gastroenterol. Hepatol. 2010;8:357–63.

28. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. Control. Clin. Trials 1991;12:142S–158S.

29. Gunn RN, Murthy V, Catafau AM, et al. Translational characterization of [11C]GSK931145, a PET ligand for the glycine transporter type 1. Synapse 2011;65:1319–32.

30. Silverberg MS, Satsangi J, Ahmad T, et al. Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: report of a Working Party of the 2005 Montreal World Congress of Gastroenterology. Can. J. Gastroenterol. 2005;19 Suppl A:5A–36A.

31. Skehan, Stephen J, Issenman R, Mernagh John, Nahmias, Claude, Jacobson K. F-fluorodeoxyglucose positron tomography in diagnosis of paediatric inflammatory bowel disease. 1999;354:836–837.

32. Neurath MF, Vehling D, Schunk K, et al. Noninvasive Assessment of Crohn ' s Disease Activity : A Comparison of 18 F-Fluorodeoxyglucose Positron Emission Tomography , Hydromagnetic Resonance Imaging , and Granulocyte Scintigraphy With Labeled Antibodies. 2002;97.

33. Bicik I, Bauerfeind P, Breitbach T, et al. Inflammatory bowel disease activity measured by positron-emission tomography. Lancet (London, England) 1997;350:262.

34. Vorselaars ADM, Crommelin HA, Deneer VHM, et al. Effectiveness of infliximab in refractory FDG PET-positive sarcoidosis. Eur. Respir. J. 2015;46:175–85.

35. Okamura K, Yonemoto Y, Arisaka Y, et al. The assessment of biologic treatment in patients with rheumatoid arthritis using FDG-PET/CT.

Rheumatology (Oxford). 2012;51:1484–91.

36. Schulzke JD, Ploeger S, Amasheh M, et al. Epithelial tight junctions in intestinal inflammation. Ann. N. Y. Acad. Sci. 2009;1165:294–300.

37. Targan SR, Hanauer SB, Deventer SJ van, et al. A short-term study of chimeric monoclonal antibody cA2 to tumor necrosis factor alpha for Crohn's disease. Crohn's Disease cA2 Study Group. N. Engl. J. Med. 1997;337:1029–35.

38. Hanauer SB, Sandborn WJ, Rutgeerts P, et al. Human anti-tumor necrosis factor monoclonal antibody (adalimumab) in Crohn's disease: the CLASSIC-I trial. Gastroenterology 2006;130:323–33; quiz 591.

39. Panaccione R, Ghosh S. Optimal use of biologics in the management of Crohn's disease. Therap. Adv. Gastroenterol. 2010;3:179–89.

40. Toriihara A, Yoshida K, Umehara I, et al. Normal variants of bowel FDG uptake in dual-time-point PET/CT imaging. Ann. Nucl. Med. 2011;25:173–8.

41 http://www3.gehealthcare.com/en/products/categories/magnetic_resonance_imaging/signa_pet-mr