Original article

Comparison and validation of FDG-PET/CT scores for polymyalgia rheumatica

Kornelis S. M. van der Geest [ORCID] [1,*], Yannick van Sleen[1,*], Pieter Nienhuis[2], Maria Sandovici[1], Nynke Westerdijk[1], Andor W. J. M. Glaudemans[2], Elisabeth Brouwer[1,†] and Riemer H. J. A. Slart[2,3,†]

Abstract

Objectives. To compare and validate the diagnostic accuracy of fluorodeoxyglucose (FDG)-PET/CT scores for PMR; and to explore their association with clinical factors.

Methods. This retrospective study included 39 consecutive patients diagnosed with PMR and 19 PMR comparators. The final clinical diagnosis was established after 6 months follow-up. Patients underwent FDG-PET/CT prior to glucocorticoid treatment. Visual grading of FDG uptake was performed at 30 anatomic sites. Three FDG-PET/CT scores (the Leuven Score, two Besançon Scores) and two algorithms (the Saint-Etienne and Heidelberg Algorithms) were investigated. Receiver operating characteristic (ROC) analysis with area under the curve (AUC) was performed. Diagnostic accuracy was assessed at predefined cut-off points.

Results. All three FDG-PET/CT scores showed high diagnostic accuracy for a clinical diagnosis of PMR in the ROC analysis (AUC 0.889–0.914). The Leuven Score provided a sensitivity of 89.7% and specificity of 84.2% at its predefined cut-off point. A simplified Leuven Score showed similar diagnostic accuracy to that of the original score. The Besançon Scores showed limited specificity at their predefined cut-off points (i.e. 47.4% and 63.2%), while ROC analysis suggested that substantially higher cut-off points are needed for these scores. The Heidelberg and Saint-Etienne Algorithms demonstrated high sensitivity, but lower specificity (i.e. 78.9% and 42.1%, respectively) for PMR. Female sex and presence of large-vessel vasculitis were associated with lower FDG-PET/CT scores in patients with PMR.

Conclusion. The Leuven Score showed the highest diagnostic utility for PMR. A modified, concise version of the Leuven Score provided similar diagnostic accuracy to that of the original score.

Key words: polymyalgia rheumatica, imaging, [18F]FDG, PET/CT, PET scoring system

> **Rheumatology key messages**
>
> - The Leuven Score accurately identifies patients with PMR on FDG-PET/CT at its predefined cut-off point.
> - A concise adaptation of the Leuven Score has similar diagnostic accuracy to that of the original score.
> - Female sex and the presence of large-vessel vasculitis were associated with lower FDG-PET/CT scores.

Introduction

PMR is the rheumatic inflammatory disease with the highest incidence above the age of 50 [1]. Patients with PMR develop debilitating pain and stiffness of the shoulder and hip girdle related to inflammation of articular and peri-articular structures. Most patients demonstrate an acute-phase response in the blood [2]. PMR is frequently associated with GCA, an autoimmune vasculitis affecting large and medium-sized arteries [3, 4].

[1]Department of Rheumatology and Clinical Immunology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands, [2]Medical Imaging Center, Department of Nuclear Medicine and Molecular Imaging, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands and [3]Department of Biomedical Photonic Imaging, Faculty of Science and Technology, University of Twente, Enschede, The Netherlands

Submitted 10 March 2021; accepted 28 May 2021

Correspondence to: Kornelis S.M. van der Geest, Department of Rheumatology and Clinical Immunology, University of Groningen,

University Medical Center Groningen, Hanzeplein 1, 9700RB, The Netherlands. E-mail: k.s.m.van.der.geest@umcg.nl

*Kornelis S. M. van der Geest and Yannick van Sleen contributed equally to this study.

†Elisabeth Brouwer and Riemer H. J. A. Slart contributed equally to this study.

The diagnosis of PMR can be challenging, since none of the symptoms and laboratory tests are entirely specific for the disease. For instance, elderly patients with late-onset SpA or OA may also present with predominant shoulder and hip complaints. An acute-phase response can also be observed in patients with RA, malignancies or para-infectious muscle pain. Current classification criteria for PMR (e.g. the 2012 EULAR/ACR criteria and Chuang criteria) have been instrumental for the conduct of research, but are not intended as diagnostic criteria [5, 6].

Accumulating evidence indicates that imaging aids the diagnosis of PMR. Subacromial–subdeltoid bursitis on US provides a sensitivity of 80% and specificity of 68% for PMR [7]. The incorporation of combined US of shoulders and hips improved the specificity of the 2012 EULAR/ACR criteria for PMR from 78–82% towards 81–91% [2, 8]. Findings on contrast-enhanced, pelvic girdle MRI showed a sensitivity and specificity of >95% for PMR [9], whereas whole-body MRI provided a sensitivity of 64% and specificity of 94% [10]. However, the exact specificity of these imaging modalities remains to be further established among a group of PMR comparators: i.e. patients with conditions that closely resemble PMR.

Another promising imaging tool for PMR is 2-deoxy-2-[^{18}F]fluoro-D-glucose PET combined with low-dose CT [fluorodeoxyglucose (FDG)-PET/CT]. FDG-PET/CT is a whole-body imaging method that allows evaluation of all key sites of inflammation in PMR. A recent meta-analysis summarized the diagnostic value of visually graded FDG uptake at defined anatomic sites for a diagnosis of PMR, but it remains unclear when a FDG-PET/CT should be considered as being positive for PMR [11]. Diagnostic algorithms that involve the assessment of 2 or 3 anatomic sites (i.e. PMR-PET algorithms) [12, 13] and more extensive composite FDG-PET/CT scores (i.e. PMR-PET scores) have been developed for this purpose [14–19]. To date, the diagnostic value of the various PMR-PET scores and algorithms has not been compared and validated.

In the current study, we performed a comprehensive analysis of FDG uptake at 30 anatomic sites in patients with PMR and PMR comparators. Subsequently, we assessed the diagnostic value of three PMR-PET scores [18, 19] and two PET algorithms [12, 13] for a clinical diagnosis of PMR. We also investigated the relationship of PMR-PET scores with clinical and laboratory features in patients with PMR.

Methods

Patients and data collection

This is a retrospective, case–control study of patients who underwent FDG-PET/CT during evaluation of suspected PMR at the Department of Rheumatology and Clinical Immunology of the University Medical Center Groningen. Consecutive patients with newly diagnosed PMR were recruited into an ongoing, observational cohort study between December 2010 and May 2020. The control group consisted of consecutive patients who received an alternative diagnosis upon evaluation for suspected PMR between December 2018 and May 2020. The decision to perform FDG-PET/CT was made by the treating physician. Patients were not yet treated with glucocorticoids or other immunosuppressants at the time of the FDG-PET/CT scan. Since 2010, a standard set of clinical and laboratory data is collected for every patient with suspected PMR in our clinic. This includes data on demographics, symptoms, physical signs, and laboratory tests. The study was conducted in accordance with the Declaration of Helsinki and approved by the ethics committee of the University Medical Center Groningen (METc 2020/220). No informed consent was required due to the retrospective nature of the study.

Reference standard

The reference standard for PMR was the final clinical diagnosis after 6 months of follow-up. The clinical diagnosis was independently established by two clinical experts (K.G. and E.B.). If no consensus was obtained, a third expert (M.S.) made the final diagnosis. The clinical diagnosis incorporated the complete history, physical findings, laboratory tests and other imaging tests (e.g. X-rays, US). The clinical experts were aware of the routine FDG-PET/CT report, but not of the outcomes of the PMR-PET scores and the PMR-PET algorithms investigated in the current study.

FDG-PET/CT scanning

Blood glucose was examined immediately before the FDG-PET/CT scan. Scans performed with glucose levels of ≥7 mmol/l were excluded from the main study analysis due to concerns about potential lowering of FDG uptake [20]. All scans were performed using an integrated PET/CT system (Biograph mCT 40 or 64-slice PET/CT; Siemens, Knoxville, TN, USA) with 3 min per bed position according to the European Association of Nuclear Medicine Guidelines [21]. Patients were scanned from the vertex of the skull up to the knees. Patients fasted for a minimum of 6 h before 3 MBq i.v. FDG/kg body weight was administered. When there was also a clinical suspicion of infective endocarditis, patients were prepared with a high-fat, low-carbohydrate diet for at least 24 h. PET/CT imaging was performed ∼60 min after i.v. FDG administration. Low-dose CT was performed for attenuation correction and anatomic mapping with 100 kV and 30 mAs.

FDG-PET/CT scoring

FDG-PET/CT scans were evaluated by a single, experienced nuclear medicine specialist (R.S.), who was unaware of the clinical diagnosis. Visual uptake of FDG was scored at 30 anatomic sites: bilateral shoulders, acromioclavicular joints, sternoclavicular joints, ischial tuberosities, hips, greater trochanters, iliopectineal

bursae, symphysis pubis entheses, S.I. joints, knees, wrists/hands, elbows, the cervical/thoracic/lumbar interspinous bursae and the cervical/thoracic/lumbar facet joints. Elbows were frequently outside the scanning window of the low-dose CT. Therefore, elbows were mostly assessed on the FDG-PET images only. Visual assessment of FDG uptake was performed at the sites of interest and the liver. Visual grading of FDG uptake was performed in accordance with current recommendations [20]: 0, no uptake; 1, uptake lower than liver; 2 uptake equal to liver; 3, uptake higher than liver. The following PMR-PET scores and algorithm were calculated: the Leuven Score according to Henckaerts *et al.* [18], the two Besançon Scores according to Sondag *et al.* [19], the Saint-Etienne Algorithm according to Flaus *et al.* [13] and the Heidelberg Algorithm according to Owen *et al.* [12]. Details on these scores and algorithm are provided in Supplementary Table S1, available at *Rheumatology* online. Other PMR-PET scores were considered but not included in the current study due to: (a) uncertainty about the handling of asymmetric FDG uptake at bilateral sites [14, 16], (b) requirement to assess particular linear/circular *vs* diffuse FDG uptake patterns at the shoulders [15], or (c) requirement to evaluate FDG uptake at the interspinous ligament in addition to the interspinous bursae [17].

Statistics

Comparison of continuous variables between two independent groups was performed by the Mann–Whitney *U* test and paired analysis of continuous variables by the Wilcoxon signed-rank test. Fisher's exact test was used for comparison of categorical variables. Correlations were evaluated by Spearman's rank test. Receiver operating characteristic (ROC) analysis and evaluation of the area under the curve (AUC) were performed. Diagnostic accuracy was evaluated at reported cut-off points (Supplementary Table S1, available at *Rheumatology* online), and at cut-off points obtained in the current study by the Youden Index. Diagnostic accuracy parameters including sensitivity, specificity, diagnostic odds ratio, positive likelihood ratio and negative likelihood ratio were evaluated. Statistical analyses were performed in Graphpad Prism 5 and 8, IBM SPSS Statistics 25 and MetaDisc 1.4. *P* values <0.05 were considered statistically significant.

Results

Patients' characteristics

A total of 69 patients were included in the study. Eleven patients, including 9 patients with a final clinical diagnosis of PMR, were excluded from the main study analysis due to a glucose level of ≥7.0 mmol/l at the time of the FDG-PET/CT scan [20]. Among the remaining 58 patients, 39 patients received a final clinical diagnosis of PMR that was confirmed after 6 months follow-up (Table 1). The majority of patients with PMR fulfilled the 2012

EULAR/ACR criteria and Chuang criteria for PMR [5, 6]. Ten patients with PMR (26%) were diagnosed with concomitant large-vessel GCA. Shoulder and hip complaints were highly prevalent among patients with PMR and non-PMR patients. Alternative diagnoses in non-PMR patients are provided in Supplementary Table S2, available at *Rheumatology* online. Six out of 19 patients were diagnosed with another rheumatic inflammatory condition (late-onset SpA, *n* = 3; RA, *n* = 1; undifferentiated oligo-arthritis, *n* = 1; arthritis induced by an immune checkpoint inhibitor, *n* = 1).

Visual FDG uptake at 30 anatomic sites

Prominent FDG uptake was observed at the bursae, tendon entheses and joints of the shoulder girdle, hip girdle and spinal column in patients with PMR (Fig. 1, and Supplementary Fig. S1 and Supplementary Video S1, both available at *Rheumatology* online). FDG uptake was frequently observed at the knees of patients with PMR, and to some extent at the wrists/hands and elbows, although these sites were not always scanned in every patient. The majority of patients with PMR showed FDG uptake at the lumbar facet joints. FDG uptake occurred symmetrically in patients with PMR, although slight asymmetry was observed for the acromioclavicular joints (i.e. uptake higher on the right side). ROC analysis indicated fair diagnostic accuracy (i.e. AUC ≥ 0.7) of visual FDG uptake at the shoulders, cervical interspinous bursae, greater trochanters, iliopectineal bursae, symphysis pubic entheses, lumbar facet joints and knees (Table 2). Good diagnostic accuracy (i.e. AUC ≥ 0.8) was observed for FDG uptake at the ischial tuberosities, hips, lumbar interspinous bursae and sternoclavicular joints. The optimal cut-off point of visually graded FDG uptake differed substantially between these anatomic sites (i.e. range 1–3). The remaining anatomic sites showed little diagnostic value for PMR (Supplementary Table S3, available at *Rheumatology* online). FDG uptake at the shoulders and hips appeared to be determined by peri-articular FDG uptake rather than articular FDG uptake (Supplementary Fig. S2, available at *Rheumatology* online), as previously shown by Owen *et al.* [12].

Diagnostic value of PMR-PET scores and algorithm

The Leuven Score was significantly higher in patients with a clinical diagnosis of PMR (median 20, range 13–24) compared with non-PMR patients (median 8, range 1–24), as shown in Supplementary Fig. S3A, available at *Rheumatology* online. The Besançon Score (sum) was also higher in patients with PMR (median 11, range 4–17) than in non-PMR patients (median 2, range 0–16). The same was true for the Besançon Score (mean), with a median score of 1.94 (range 1.00–2.88) in patients with PMR *vs* 0.59 (range 0.06–2.76) in non-PMR patients.

The ability to discriminate between patients with and without PMR was excellent for all three PMR-PET

TABLE 1 Patients' characteristics

	PMR (*n* = 39)	Non-PMR (*n* = 19)
Sex, no. of females (%)	26 (67%)	13 (68%)
Age, median (range)	71 (54–82)	59 (42–81)
Fulfilling EULAR/ACR criteria for PMR, no. of patients (%)[a]	34 (87%)	4 (21%)
Fulfilling Chuang criteria for PMR, no. of patients (%)	30 (77%)	0 (0%)
Fulfilling EULAR/ACR or Chuang criteria for PMR, no. of patients (%)	36 (92%)	4 (21%)
Concomitant large-vessel GCA present, no. of patients (%)	10 (26%)	0 (0%)
Neck pain present, no. of patients (%)	22 (56%)	8 (42%)
Bilateral shoulder pain present, no. of patients (%)	38 (97%)	14 (74%)
Hip pain or stiffness present, no. of patients (%)	34 (87%)	16 (84%)
Morning stiffness > 45 min present, no. of patients (%)	32 (82%)	8 (42%)
Haemoglobin, mmol/L, median (range)	7.5 (5.6–9.3)	8.7 (6.1–11.0)
CRP, mg/L, median (range)	35.0 (3.2–127.0)	4.9 (0.3–35.0)
ESR, mm/h, median (range)	57 (7–116)	14 (2–91)
Platelet count, 10^9/L, median (range)	334 (170–552)	276 (140–402)

Data are shown for the 58 patients, of which 39 patients received a final clinical diagnosis of PMR after 6 months follow-up.
[a]Rheumatoid factor and anti-CCP were not tested in 5 patients (presumed negative for classification).

FIG. 1 Visual FDG grades at 30 anatomic sites
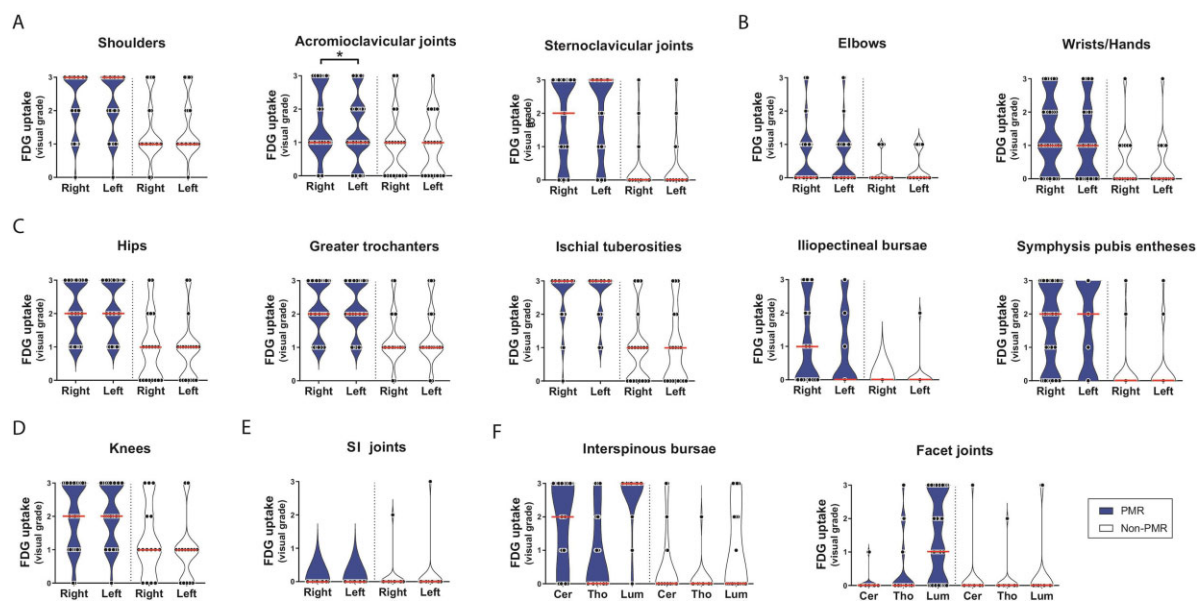
Violin plots indicating FDG grading at distinct anatomic sites. Data are shown for patients with PMR (*n* = 39) and non-PMR patients (*n* = 19) unless stated otherwise. Data on elbows were obtained from 38 PMR patients and 18 non-PMR patients; data on wrists/hands from 34 PMR patients and 15 non-PMR patients; data on knees from 28 PMR patients and 15 non-PMR patients. Visual grading of FDG uptake was performed as in Reference [20]: 0, no uptake; 1, uptake lower than liver; 2, uptake equal to liver; 3, uptake higher than liver. The red bars indicate the median values. Data are shown for (A) anatomic sites at the shoulder girdle, (B) the upper limb, (C) the pelvic girdle, (D) the knees, (E) the SI joints and (F) anatomic sites at the spinal column. Cer = cervical. Lum = lumbar. Tho = thoracic. FDG grades at bilateral sites were compared by the Wilcoxon signed-rank test. $^*P < 0.05$.

scores, as indicated by an AUC close to 0.9 in the ROC analysis (Table 3 and Supplementary Fig. S3B, available at *Rheumatology* online). The Leuven Score provided a sensitivity of 89.7% and specificity of 84.2% at the previously reported cut-off value of 16 [18]. ROC analysis based on data in the current study suggested a comparable, optimal cut-off point of 15 for the Leuven Score. High sensitivity but limited specificity were obtained for the two Besançon Scores at their predefined cut-off values [19, 22]. ROC analyses suggested that the optimal

TABLE 2 Diagnostic accuracy of FDG grades at key anatomic sites for a diagnosis of PMR

Anatomic site	AUC in ROC analysis	Visual FDG grade	Sensitivity (95% CI)	Specificity (95% CI)
Ischial tuberosities	0.865 (0.749–0.981)	≥1	97.4 (86.5, 99.9)	52.6 (28.9, 75.6)
		≥2[b]	89.7 (75.8, 97.1)	78.9 (54.4, 93.9)
		≥3	71.8 (55.1, 85.0)	84.2 (60.4, 96.6)
Hips	0.836 (0.715–0.957)	≥1	100 (91.0, 100.0)	42.1 (20.3, 66.5)
		≥2[b]	71.8 (55.1, 85.0)	84.2 (60.4, 96.6)
		≥3	43.6 (27.8, 60.4)	89.5 (66.9, 98.7)
Lumbar interspinous bursa	0.836 (0.713–0.959)	≥1	92.3 (79.1, 98.4)	57.9 (33.5, 79.7)
		≥2	89.7 (75.8, 97.1)	63.2 (38.4, 83.7)
		≥3[b]	87.2 (72.6, 95.7)	78.9 (54.4, 93.9)
Sternoclavicular joints	0.800 (0.683–0.918)	≥1[b]	71.8 (55.1, 85.0)	84.2 (60.4, 96.6)
		≥2	51.3 (34.8, 67.6)	94.7 (74.0, 99.9)
		≥3	46.2 (30.1, 62.8)	94.7 (74.0, 99.9)
Symphysis pubis entheses	0.785 (0.659–0.910)	≥1[b]	69.2 (52.4, 83.0)	89.5 (66.9, 98.7)
		≥2	48.7 (32.4, 65.2)	89.5 (66.9, 98.7)
		≥3	28.2 (15.0, 44.9)	94.7 (74.0, 99.9)
Greater trochanters	0.781 (0.650–0.912)	≥1	100 (91.0, 100)	10.5 (1.3, 33.1)
		≥2[b]	74.4 (57.9, 87.0)	78.9 (54.4, 93.9)
		≥3	38.5 (23.4, 55.4)	85 (62.1, 96.8)
Lumbar facet joints	0.775 (0.642–0.907)	≥1[b]	69.2 (52.4, 83.0)	89.5 (66.9, 98.7)
		≥2	48.7 (32.4, 65.2)	89.5 (66.9, 98.7)
		≥3	33.3 (19.1, 50.2)	89.5 (66.9, 98.7)
Cervical interspinous bursa	0.769 (0.640–0.898)	≥1[b]	74.4 (57.9, 87.0)	73.7 (48.8, 90.9)
		≥2	61.5 (44.6, 76.6)	84.2 (60.4, 96.6)
		≥3	48.7 (32.4, 65.2)	89.5 (66.9, 98.7)
Knees[a]	0.746 (0.586–0.907)	≥1	89.3 (71.8, 97.7)	33.3 (11.8, 61.6)
		≥2[b]	64.3 (44.1, 81.4)	86.7 (59.5, 98.3)
		≥3	32.1 (15.9, 52.4)	86.7 (59.5, 98.3)
Shoulders	0.719 (0.577–0.862)	≥1	97.4 (86.5, 99.9)	10.5 (1.3, 33.1)
		≥2[b]	71.8 (55.1–85.0)	68.4 (60.4, 87.4)
		≥3	43.6 (27.8, 60.4)	84.2 (28.9, 96.6)
Iliopectineal bursae	0.718 (0.590–0.846)	≥1[b]	43.6 (27.8, 60.4)	100 (82.4, 100.0)
		≥2	28.2 (15.0, 44.9)	100 (82.4, 100.0)
		≥3	12.8 (4.3, 27.4)	100 (82.4, 100.0)

Data are shown for anatomic sites with an area under the curve (AUC) ≥ 0.700 in the receiver operating characteristic (ROC) analysis. Data were obtained from 39 PMR patients and 19 non-PMR patients unless stated otherwise. For bilateral sites with discordant results, the lowest visual score was used. For the interspinous bursae and facet joints, the highest score was used. [a]Data on knees were obtained from 28 PMR patients and 15 non-PMR patients. [b]Optimal cut-off point according to Youden Index.

cut-off points for both Besançon Scores should be substantially higher. The Saint-Etienne and Heidelberg Algorithms also demonstrated high sensitivity for PMR, but their specificities were 42.1% and 78.9%, respectively.

A concise adaptation of the Leuven score

We next determined whether a simplified version of the Leuven Score would perform equally well as the original Leuven Score. We adapted the Leuven Score by only including anatomic sites with an AUC ≥ 0.8 in the ROC analysis (Table 2): the sternoclavicular joints, hips, ischial tuberosities and lumbar interspinous bursa (Fig. 2A). This modified score, hence termed Leuven/Groningen Score, could range from 0 to 14 points. The Leuven/Groningen Score provided an AUC of 0.926 (95% CI 0.84, 100.0) in the ROC analysis; with a sensitivity of 89.7% (95% CI 75.8, 97.1) and specificity of 84.2%

(95% CI 60.4, 96.6) at the optimal cut-off point of 8 (Fig. 2B and Supplementary Table S4, available at *Rheumatology* online). Thus, our concise adaptation of the Leuven Score might provide comparable diagnostic accuracy as the original Leuven Score. Alternative cut-off points providing either a sensitivity or specificity of ≥95% could be determined for the Leuven/Groningen Score (Fig. 2B), as well as the original Leuven Score (Supplementary Fig. S4, available at *Rheumatology* online).

Relationship of PMR-PET scores with clinical factors

Next, we investigated whether PMR-PET scores are influenced by clinical factors in patients with PMR. The Leuven Score and Besançon Score (mean) tended to increase slightly with age among patients with PMR (Fig. 3A). Females showed lower Leuven Scores than males, and a similar trend was observed for the

TABLE 3 Diagnostic accuracy of PMR-PET scores and algorithms

	AUC (95% CI)	Cut-off values		Sensitivity (95% CI)	Specificity (95% CI)	DOR (95% CI)	LR+ (95% CI)	LR– (95% CI)
Leuven Score	0.914 (0.815, 100.0)	Predefined by Henckaerts et al.	16	89.7 (75.8, 97.1)	84.2 (60.4, 96.6)	46.67 (9.33, 233.36)	5.68 (2.00, 16.14)	0.12 (0.05, 0.31)
		Youden Index current study	15	92.3 (79.1, 98.4)	84.2 (60.4, 96.6)	64.00 (11.63, 352.17)	5.85 (2.06, 16.58)	0.09 (0.03, 0.28)
Besançon Score (sum)	0.889 (0.778, 1.000)	Predefined by Sondag et al.	3	100.0 (91.0, 100.0)	63.2 (38.4, 83.7)	131.67 (7.01, 2472.10)	2.63 (1.49, 4.64)	0.02 (0.00, 0.32)
		Youden Index current study	7	89.7 (75.8, 97.1)	84.2 (60.4, 96.6)	46.67 (9.33, 233.36)	5.68 (2.00, 16.14)	0.12 (0.05, 0.31)
Besançon Score (mean)	0.901 (0.791, 1.000)	Predefined by Sondag et al.	0.53	100.0 (91.0, 100.0)	47.4 (24.4, 71.1)	71.48 (3.84, 1330.70)	1.88 (1.24, 2.86)	0.03 (0.00, 0.43)
		Predefined by Amat et al.	0.765	100.0 (91.0, 100.0)	63.2 (38.4, 83.7)	131.67 (7.01, 2472.10)	2.63 (1.49, 4.64)	0.02 (0.00, 0.32)
		Youden Index current study	1.324	87.2 (72.6, 95.7)	89.5 (66.9, 98.7)	57.80 (10.14, 329.34)	8.28 (2.22, 30.89)	0.14 (0.06, 0.33)
Saint-Etienne Algorithm	NA	NA		100.0 (91.0, 100.0)	42.1 (20.3, 66.5)	58.39 (3.13, 1090.00)	1.72 (1.18, 2.51)	0.03 (0.00, 0.48)
Heidelberg Algorithm	NA	NA		89.7 (75.8, 97.1)	78.9 (54.4, 93.9)	32.81 (7.23, 148.85)	4.26 (1.77, 10.25)	0.13 (0.05, 0.34)

Diagnostic accuracy was determined for all three PMR-PET scores at the reported cut-off values. In addition, the diagnostic accuracy was evaluated at optimal cut-off values obtained by receiver operating characteristics (ROC) analysis and the Youden Index in the current study. Area under the curve (AUC) in the ROC analysis, sensitivity, specificity, diagnostic odds ratio (DOR), positive likelihood ratio (LR+) and negative likelihood ratio (LR–) are provided along with their 95% CI. NA = not applicable.

Besançon Scores (Fig. 3B). The Leuven Score and Besançon Scores were lower if large-vessel GCA was present (Fig. 3C). Nevertheless, the sensitivity of these PMR-PET Scores was equally high in patients with and without large vessel GCA (Supplementary Table S5, available at *Rheumatology* online). Thirteen out of 39 (33%) patients with PMR used NSAIDs during the FDG-PET/CT scan. Patients taking NSAIDs tended to have higher PMR-PET scores than those who were not taking NSAIDs, although this was not statistically significant (Fig. 3D).

Relationship between PMR-PET scores and laboratory tests

We further evaluated the association between PMR-PET scores and inflammation markers in the blood. No correlations were found between PMR-PET scores and any of the following laboratory tests in patients with PMR: CRP, ESR, haemoglobin and platelet count (Supplementary Fig. S5, available at *Rheumatology* online). We finally evaluated the relationship between hyperglycaemia and PMR-PET scores in patients with PMR. This analysis also comprised the 9 PMR patients who were excluded from the main study analysis due to hyperglycaemia (glucose levels ranging from 7.1 to 10.8 mmol/l). Overall, PMR-PET scores were quite comparable in patients with and without hyperglycaemia (Supplementary Fig. S6, available at *Rheumatology* online). Nevertheless, the sensitivity of a positive scan as defined by the PMR-PET scores was 8–14% lower in patients with hyperglycaemia than in patients with normal glucose levels, although this difference was not statistically significant (Supplementary Table S6, available at *Rheumatology* online).

Discussion

This is the first study comparing and validating the diagnostic value of PMR-PET scores and algorithms. Our findings confirm the excellent diagnostic accuracy of the Leuven Score [18], while providing evidence that a concise Leuven/Groningen Score might perform equally well. The overall diagnostic accuracy of the Besançon Scores was also good, but previously reported cut-off values for these scores could not be validated [19, 22]. The Heidelberg Algorithm showed excellent sensitivity, but its specificity was slightly lower than that of the PMR-PET scores [12]. The Saint-Etienne Algorithm provided limited diagnostic accuracy in our study [13].

The Leuven Score aids standardized interpretation of FDG-PET/CT in patients with suspected PMR by providing a clear definition of a positive scan. Our study validates the diagnostic accuracy of the Leuven Score at its predefined cut-off point [18]. Henckaerts et al. developed their Leuven Score in a prospective study including consecutive patients with suspected PMR who all underwent FDG-PET/CT [18]. The authors reported that their PMR-PET score had an optimal cut-off point
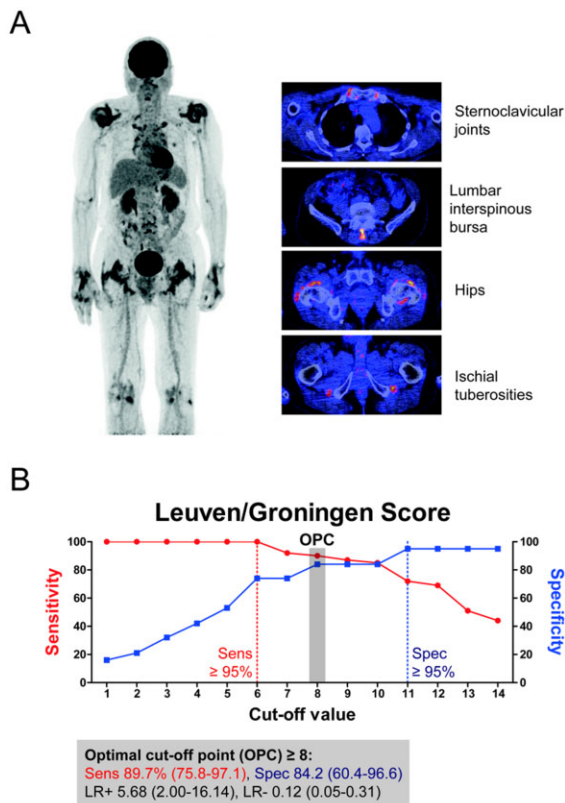
The original Leuven Score [18] was simplified by restricting the analysis to anatomic sites with an area under the curve of ≥0.8 in the receiver operating characteristics (ROC) analysis of FDG grades as shown in Table 2. (A) The Leuven/Groningen Score contained visual FDG grades (0–2) of the sternoclavicular joints, hips, ischial tuberosities and lumbar interspinous bursa. The Leuven/Groningen Score could range from 0 to 14 points. (B) The diagnostic accuracy of the Leuven/Groningen Score was evaluated at the optimal cut-off point (OPC) obtained by ROC analysis and the Youden Index. Sensitivity (Sens), specificity (Spec), positive likelihood ratio (LR+) and negative likelihood ratio (LR−) are provided along with the 95% CI. Alternative cut-off values providing either ≥95% sensitivity or specificity are also depicted. Data are shown for patients with PMR ($n = 39$) and non-PMR patients ($n = 19$).

of 16, at which a sensitivity of 85.1% and specificity of 87.5% were observed for a clinical diagnosis of PMR. Our retrospective study including patients with PMR and PMR comparators yielded comparable results: a sensitivity of 89.7% and a specificity of 84.2% at the same cut-off point. The Leuven Score, which is based on evaluation of 12 anatomic sites, might potentially be simplified. We here report on a concise adaptation of the Leuven Score, which only requires evaluation of 7 anatomic sites while showing a comparable diagnostic

accuracy as the full Leuven Score. Although this concise Leuven/Groningen Score needs to be validated, this adaptation might improve the feasibility and implementation of the PMR-PET score in daily clinical practice. The original Leuven Score and the Leuven/Groningen Score also provided alternative cut-off points with ≥95% sensitivity or specificity, which might be applied depending on the need to rule out or rule in the disease.

We validated the diagnostic potential of the Besançon Scores, but the optimal cut-off points for these scores remain uncertain. The incorporation of more anatomic sites in the Besançon Scores (i.e. 17 sites) did not improve their diagnostic accuracy when compared with that of the Leuven Score. Prior studies suggested that the Besançon Score (sum) has a sensitivity of 74–86% and a specificity of 79–85.5% for a diagnosis of PMR at a cut-off value of 3 [19, 22]. The Besançon Score (mean) previously demonstrated a sensitivity of 80% and specificity of 77% for PMR at a cut-off of 0.53 in one study [19], and a sensitivity of 82.5% and specificity of 75.8% at a cut-off of 0.765 in another study [22]. In our study, the two Besançon Scores provided limited specificity (47 and 63%) at these cut-off points. We found substantially higher optimal cut-off values for both Besançon Scores. This discrepancy might be explained by differences in patient inclusion. The prior studies contained PMR patients who had already received treatment, as well as control patients who were not suspected of having PMR (e.g. oncologic patients) [19, 22].

Additional PMR-PET scores and algorithms have been developed [11]. We tested the diagnostic accuracy of the Saint-Etienne Algorithm, but this algorithm showed substantially lower specificity for PMR (i.e. 42.1%) than reported in the original study (i.e. 80.1%) [13]. The Heidelberg Algorithm includes assessment of the ischial tuberosities, peri-articular shoulder structures and the interspinous bursae, and yielded a sensitivity of 90.9% and specificity of 92.4% for PMR in the original study [12]. Our main assessment of the shoulders did not include evaluation of peri-articular *vs* articular FDG uptake. Nevertheless, a sub-analysis indicated that FDG uptake at the shoulders was determined by peri-articular rather than articular uptake, which is in good agreement with Owen *et al.* [12]. Therefore, we also explored the diagnostic accuracy of the Heidelberg Algorithm. The Heidelberg Algorithm showed high sensitivity in our study, but its specificity was somewhat lower (i.e. 78.9%) than that reported in the original study. This difference might be explained by the inclusion of a control group with patients not suspected of having PMR in the latter study [12, 23]. Additional PMR-PET scores were not evaluated in the current study, since the FDG grading system at particular anatomic sites differed, or was uncertain [14–17].

Mild hyperglycaemia and NSAID treatment showed limited or no effect on FDG-PET/CT findings, respectively. It is currently recommended that glucose levels are <7.0 mmol/l during the scan due to potential lowering of

Fɪɢ. 3 Relationship between PMR-PET scores and clinical characteristics
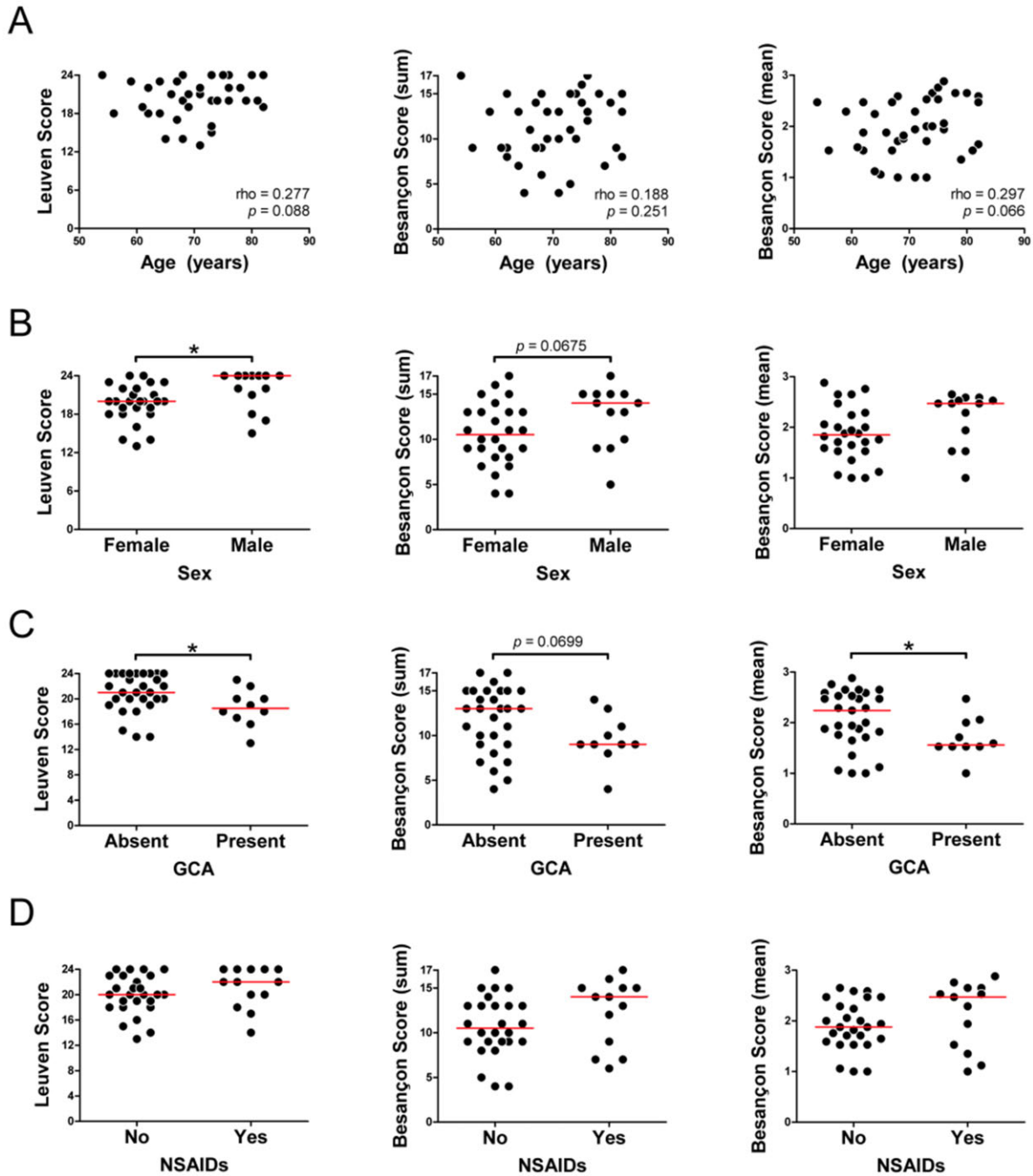
The association between PMR-PET scores and (A) age, (B) sex, (C) presence of concomitant large-vessel GCA and (D) use of NSAIDs during the scan was evaluated in patients with PMR ($n = 39$). Correlations were evaluated by Spearman's rank correlation coefficient, and the various groups were compared by the Mann–Whitney U test. *$P < 0.05$.

FDG uptake by inflammatory lesions [20]. The sensitivity of FDG-PET/CT, as determined by positive PMR-PET scores, tended to be 8–14% lower in PMR patients with mild hyperglycaemia (glucose level up to 10.8 mmol/l) than in those with normal glucose levels. Nevertheless, the FDG-PET/CT remained positive in the majority of PMR patients with mild hyperglycaemia. NSAIDs are sometimes withheld prior to imaging of patients with inflammatory rheumatic conditions, but NSAID use did not seem to lower FDG uptake in patients with PMR.

PMR-PET scores actually tended to be higher in patients on NSAID treatment, which likely reflects the severity of symptoms in those patients. Although further confirmation is needed, our findings suggest that FDG-PET/CT might be performed in patients with mild hyperglycaemia and that NSAID treatment should not be withheld prior to the scan.

Female sex and the presence of large-vessel GCA were associated with lower PMR-PET scores in patients with PMR. We are not aware of prior studies comparing FDG-PET/CT scans of male and female patients with PMR. Two prior studies also reported that (peri)articular FDG uptake is less pronounced in PMR patients with concomitant vasculitis [14, 17]. The explanation and clinical relevance of these associations remain to be elucidated. Importantly, the sensitivity of FDG-PET/CT for PMR remained high in patients with concomitant large-vessel inflammation. PMR-PET scores showed no relationship with inflammation markers in the blood. This is in agreement with the majority of studies suggesting that (peri)articular and systemic inflammation are not tightly linked processes in PMR [24–28].

FDG-PET/CT could be a valuable tool in the diagnostic work-up of PMR. FDG-PET/CT not only allows for the assessment of all key structures involved in PMR, but also aids the detection of co-existing large-vessel vasculitis and important alternative diagnoses, such as occult malignancies [29, 30]. Keeping in mind the costs and radiation burden of FDG-PET/CT, this imaging modality could be applied in selected cases: for instance, patients with atypical presentation, or patients with strong clinical suspicion but a negative US or MRI scan. Recently, it was suggested that FDG-PET/CT may also aid the diagnosis of the PMR-like syndrome developing upon immune checkpoint inhibitor therapy for cancer [31].

Strengths of our study are the performance of FDG-PET/CT scans prior to the start of glucocorticoid treatment and inclusion of a control group of PMR comparators consecutively seen in daily clinical practice. In order to minimize selection bias, we aimed to include the full spectrum of patients with PMR, including those that are more difficult to diagnose [32]. Thus, a small minority of patients did not fulfil the classification criteria for PMR [5, 6]. Importantly, the clinical diagnosis of PMR was established after 6 months of follow-up by expert rheumatologists. Our study may also have limitations. The retrospective, case–control design could have introduced selection bias. However, our study obtained very similar findings to those reported in the prospective study by Henckaerts *et al.* [18]. Evaluation of peri-articular and articular FDG uptake at the shoulders and hips was only performed in a sub-analysis of PMR patients and PMR comparators. The interrater reliability was not evaluated, but the FDG-PET/CT scans were evaluated by an experienced nuclear medicine specialist in accordance with recommendations for the use of FDG-PET/CT for PMR [20]. The clinicians establishing the diagnosis of PMR were aware of the FDG-PET/CT report, since it was performed as part of routine clinical care in our centre.

However, the clinicians were not aware of the PMR-PET scores. Potential bias due to the reference standard was further limited by careful clinical evaluation and rigorous follow-up for 6 months. Consequently, some patients with high PMR-PET scores were diagnosed with a condition other than PMR. Our study was relatively small, and larger studies might improve the certainty of the diagnostic accuracy estimates.

In conclusion, PMR-PET scores provide a clear definition for a positive FDG-PET/CT scan in patients with suspected PMR. This study validates the excellent diagnostic accuracy of the Leuven Score. We propose a concise adaption of the Leuven Score, which could be used more easily in daily clinical practice while retaining the diagnostic value of the original Leuven Score. Female sex and concomitant large-vessel vasculitis were associated with lower PMR-PET scores. A large, prospective, multicentre study is required to further confirm our findings.

Data availability statement

All data relevant to the study are included in the article or uploaded as supplementary information.

Supplementary data

Supplementary data are available at *Rheumatology* online.

References

1 Crowson CS, Matteson EL, Myasoedova E *et al.* The lifetime risk of adult-onset rheumatoid arthritis and other inflammatory autoimmune rheumatic diseases. Arthritis Rheum 2011;63:633–9.

2 Dasgupta B, Cimmino MA, Kremers HM *et al.* 2012 provisional classification criteria for polymyalgia rheumatica: a European League Against Rheumatism/ American College of Rheumatology collaborative initiative. Arthritis Rheum 2012;64:943–54.

3 Dejaco C, Brouwer E, Mason JC *et al.* Giant cell arteritis and polymyalgia rheumatica: current challenges and opportunities. Nat Rev Rheumatol 2017;13:578–92.

4 van Sleen Y, Boots AMH, Abdulahad WH *et al.* High angiopoietin-2 levels associate with arterial inflammation and long-term glucocorticoid requirement in polymyalgia rheumatica. Rheumatology (Oxford) 2020; 58:176–84.

5 Dasgupta B, Cimmino MA, Maradit-Kremers H *et al.* 2012 provisional classification criteria for polymyalgia rheumatica: a European League Against Rheumatism/American College of Rheumatology collaborative initiative. Ann Rheum Dis 2012;71:484–92.

6 Chuang TY, Hunder GG, Ilstrup DM, Kurland LT. Polymyalgia rheumatica: a 10-year epidemiologic and clinical study. Ann Intern Med 1982;97:672–80.

7 Mackie SL, Koduri G, Hill CL *et al.* Accuracy of musculoskeletal imaging for the diagnosis of polymyalgia rheumatica: systematic review. RMD Open 2015;1: e000100. eCollection 2015.

8 Macchioni P, Boiardi L, Catanoso M, Pazzola G, Salvarani C. Performance of the new 2012 EULAR/ACR classification criteria for polymyalgia rheumatica: comparison with the previous criteria in a single-centre study. Ann Rheum Dis 2014;73:1190–3.

9 Fruth M, Seggewiss A, Kozik J *et al.* Diagnostic capability of contrast-enhanced pelvic girdle magnetic resonance imaging in polymyalgia rheumatica. Rheumatology (Oxford) 2020;59:2864–71.

10 Mackie SL, Pease CT, Fukuba E *et al.* Whole-body MRI of patients with polymyalgia rheumatica identifies a distinct subset with complete patient-reported response to glucocorticoids. Ann Rheum Dis 2015;74: 2188–92.

11 van der Geest KSM, Treglia G, Glaudemans AWJM *et al.* Diagnostic value of [^{18}F]FDG-PET/CT in polymyalgia rheumatica: a systematic review and meta-analysis. Eur J Nucl Med Mol Imaging 2020;48:1876–89.

12 Owen CE, Poon AMT, Yang V *et al.* Abnormalities at three musculoskeletal sites on whole-body positron emission tomography/computed tomography can diagnose polymyalgia rheumatica with high sensitivity and specificity. Eur J Nucl Med Mol Imaging 2020;47:2461–8.

13 Flaus A, Amat J, Prevot N *et al.* Decision tree with only two musculoskeletal sites to diagnose polymyalgia rheumatica using [^{18}F]FDG PET-CT. Front Med 2021;8:646974.

14 Yamashita H, Kubota K, Takahashi Y *et al.* Whole-body fluorodeoxyglucose positron emission tomography/computed tomography in patients with active polymyalgia rheumatica: evidence for distinctive bursitis and large-vessel vasculitis. Mod Rheumatol 2012;22:705–11.

15 Takahashi H, Yamashita H, Kubota K *et al.* Differences in fluorodeoxyglucose positron emission tomography/computed tomography findings between elderly onset rheumatoid arthritis and polymyalgia rheumatica. Mod Rheumatol 2015;25:546–51.

16 Wakura D, Kotani T, Takeuchi T *et al.* Differentiation between polymyalgia rheumatica (PMR) and elderly-onset rheumatoid arthritis using 18F-fluorodeoxyglucose positron emission tomography/computed tomography: is enthesitis a new pathological lesion in PMR? PLoS One 2016;11:e0158509.

17 Emamifar A, Ellingsen T, Hess S *et al.* The utility of 18F-FDG PET/CT in patients with clinical suspicion of polymyalgia rheumatica and giant cell arteritis: a prospective, observational, and cross-sectional study. ACR Open Rheumatol 2020;2:478–90.

18 Henckaerts L, Gheysens O, Vanderschueren S, Goffin K, Blockmans D. Use of 18F-fluorodeoxyglucose positron emission tomography in the diagnosis of polymyalgia rheumatica—a prospective study of 99 patients. Rheumatology (Oxford) 2018;57:1908–16.

19 Sondag M, Guillot X, Verhoeven F *et al.* Utility of 18F-fluoro-dexoxyglucose positron emission tomography for the diagnosis of polymyalgia rheumatica: a controlled study. Rheumatology (Oxford) 2016;55:1452–7.

20 Slart RHJA, Writing group, Reviewer group *et al.*; EANM Committee Coordinator. FDG-PET/CT(A) imaging in large vessel vasculitis and polymyalgia rheumatica: joint procedural recommendation of the EANM, SNMMI, and the PET Interest Group (PIG), and endorsed by the ASNC. Eur J Nucl Med Mol Imaging 2018;45:1250–69.

21 Jamar F, Buscombe J, Chiti A *et al.* EANM/SNMMI guideline for 18F-FDG use in inflammation and infection. J Nucl Med 2013;54:647–58.

22 Amat J, Chanchou M, Olagne L *et al.* Utility of ^{18}F-fluorodeoxyglucose positron emission tomography in inflammatory rheumatism, particularly polymyalgia rheumatica: a retrospective study of 222 PET/CT. Front Med (Lausanne) 2020;7:394.

23 Rutjes AW, Reitsma JB, Di Nisio M *et al.* Evidence of bias and variation in diagnostic accuracy studies. CMAJ 2006;174:469–76.

24 Cimmino MA, Camellino D, Paparo F *et al.* High frequency of capsular knee involvement in polymyalgia rheumatica/giant cell arteritis patients studied by positron emission tomography. Rheumatology (Oxford) 2013;52:1865–72.

25 Camellino D, Paparo F, Morbelli S *et al.* Interspinous bursitis is common in polymyalgia rheumatica, but is not associated with spinal pain. Arthritis Res Ther 2014;16:492.

26 Lund-Petersen A, Voss A, Laustrup H. PET-CT findings in patients with polymyalgia rheumatica without symptoms of cranial ischaemia. Dan Med J 2017;64:A5410.

27 Palard-Novello X, Querellou S, Gouillou M *et al.* Value of ^{18}F-FDG PET/CT for therapeutic assessment of patients with polymyalgia rheumatica receiving tocilizumab as first-line treatment. Eur J Nucl Med Mol Imaging 2016;43:773–9.

28 Wendling D, Sondag M, Giraud N *et al.* Muscle involvement on 18F-FDG PET-CT in polymyalgia rheumatica. A controlled retrospective study of 101 patients. Joint Bone Spine 2020;87:225–8.

29 Hop H, Mulder DJ, Sandovici M *et al.* Diagnostic value of axillary artery ultrasound in patients with suspected giant cell arteritis. Rheumatology (Oxford) 2020;59: 3676–84.

30 Emamifar A, Hess S, Ellingsen T *et al.* Prevalence of newly diagnosed malignancies in patients with polymyalgia rheumatica and giant cell arteritis, comparison of 18F-FDG PET/CT scan with chest X-ray

and abdominal ultrasound: data from a 40 week prospective, exploratory, single centre study. J Clin Med 2020;9:3940.

31 van der Geest KSM, Sandovici M, Rutgers A *et al.* Imaging in immune checkpoint inhibitor-induced

polymyalgia rheumatica. Ann Rheum Dis 2020; RA Access published 2 April 2020, doi: 10.1136/annrheumdis-2020-217381

32 Sackett DL, Haynes RB. The architecture of diagnostic research. BMJ 2002;324:539–41.