Perspectives and unmet needs in polymyalgia rheumatica. Providing the fundamental framework for the development of new treatment regimes in polymyalgia rheumatica

B.D. Nielsen¹, B. Dasgupta²

¹Department of Rheumatology and Department of Clinical Medicine, Aarhus University Hospital, Denmark; ²Rheumatology Department, Southend University Hospital, Westcliff-on-sea, United Kingdom

SUMMARY
Polymyalgia rheumatica is effectively treated with glucocorticoids. However, glucocorticoid treatment can cause numerous and potentially serious side effects. Therefore, lowest effective dose and shortest duration to control disease is aimed for and glucocorticoid-sparing treatments are needed. Nevertheless, development of treatment regimens in PMR has been hampered by the lack of reliable classification criteria and evidence-based outcome measures. In this editorial, we discuss the need for valid classification criteria in PMR, the strengths and limitations of the ACR/EULAR 2012 provisional classification criteria for PMR and the need of validation and possible refining of the criteria.

Key words: Polymyalgia rheumatica; Classification; Diagnosis.

INTRODUCTION
Within recent years, major progress has been made considering international collaboration, strategizing and achieving common scientific goals within the field of polymyalgia rheumatica (PMR). Immense efforts have been devoted to provide the best of knowledge for both clinical decision making and future studies considering classification (1), outcome measures (2-5), therapeutic interventions and prognosis (6). Recommendations for the management of PMR have been developed, and serve to standardise clinical practice and improve patient care (7). After several years of stagnation, clinical trials on glucocorticoid (GC) sparing treatment arises. However, there are still fundamental unmet needs within the field of PMR that deserves to be handled to raise the quality of our scientific work. In this editorial, we expand on the need for valid classification criteria for PMR. We also suggest a framework that could form the basis for clinical trials, outcome measures and remission-relapse criteria in PMR.

THE NEED OF AN OPTIMAL FRAMEWORK FOR THE DEVELOPMENT OF EFFECTIVE AND VALID TREATMENT REGIMENS IN POLYMYALGIA RHEUMATICA

In several years, GC treatment has been considered highly effective and the standard of care in PMR treatment. A moderate dose should be efficient, prompt effect could be expected and treatment would be tapered and subsequently discontinued within 1-2 years of treatment (8). However, clinical trials on initial GC treatment dose, tapering regime and treatment duration are sparse and to a large extent studies are low scientific quality (6). Therefore, current GC regimens are mainly based on clinical expert experience rather than the results of
randomized, controlled clinical trials. In many PMR cases, the abovementioned strategy is successful. However, the PMR population is heterogeneous which can challenge a standardized treatment strategy. Studies have highlighted the fact that treatment duration exceeds 3 years in more than 40% of patients (9) and in only about half of the population GCs are discontinued after 2 years of treatment (10). The myth of full resolution of symptoms with steroid treatment is challenged by patients’ perspectives (11). The prevailing clinical opinion of rapid and complete steroid responsiveness is disputed by observational studies showing that complete response, assessed according to study-defined improvement in PMR visual assessment scale (VAS), improvement in morning stiffness and normalization of inflammatory markers, is only obtained in a minority (8%) by week one, less than 50% by week 3 and in a little more than half of the patients by week 4 (5, 12).

Lowest effective dose and shortest treatment duration is aimed for, due to the numerous and potentially serious side effects of GC treatment (13, 14). This is especially true since the PMR population is older and comorbidities therefore more frequent, making these patients even more vulnerable to the potential serious side effects such as osteoporosis, cardiovascular disease, diabetes and infections (15, 16).

From both a patient, doctor and a health economic perspective the need for GC sparing treatment is substantial. Besides being effective, new treatment regimens must also be relatively inexpensive, safe and well tolerated in elderly in order to make treatment applicable to the majority. Prognostic tools for treatment stratification are crucial. Methotrexate (MTX) has been studied in PMR but evidence only allows individualized use in patients at risk of relapse, exposed to prolonged GC therapy, and experiencing GC-related adverse events (17). However, the effect of MTX is modest and RCTs of other synthetic DMARDs such as leflunomide are urgently needed. The IL-6 blocking biologic agent, tocilizumab, has shown promising results in GC resistant or intolerant PMR patients and as monotherapy in new onset PMR (18, 19).

However, development of new treatment regimens in PMR has been hampered by the lack of reliable classification criteria and evidence-based outcome measures. In their absence, clinical trials lack trustworthiness and the opportunities for meta-analyses and comparing results across studies are minimized.

THE ACR/EULAR 2012 PROVISIONAL CLASSIFICATION CRITERIA FOR POLYMYALGIA RHEUMATICA

The main purpose of PMR classification criteria is to obtain a safe, specific and applicable approach to classification of the disease in order to provide a basic tool for clinical trials of novel therapy in PMR and developing outcome measures and remission-relapse criteria in PMR.

Since the phenotypical presentations of PMR is heterogeneous and disease presentation may mimic various diseases such as e.g., degenerative shoulder conditions, inflammatory arthritides, giant cell arteritis (GCA), endocrine disorders and malignancies (20, 21), criteria showing a high specificity is preferred at the expense of high sensitivity. Moreover, the criteria need to be applicable prior to commencement of therapy in order to be useful in clinical trials, that is, specific requirements for GC response should not be incorporated.

The ACR/EULAR 2012 provisional classification criteria for PMR were established according to the ACR guidelines for the development of classification criteria for rheumatic diseases and published by the ACR-EULAR Work Group for PMR and GCA (1). Candidate criteria items were agreed by an international consensus of rheumatologists, general practitioners and allied specialists and imaging criteria items involving ultrasound (US) were included. These candidate criteria items were then studied in a unique case-control cohort, prospective, international 6 months study.
design with a single eligibility criterion, that of patients above 50 years presenting with new onset bilateral shoulder pain. Blinded re-verification of included cases and comparators was done and a scoring algorithm that performed best in distinguishing PMR from non-PMR was obtained (Table I).

Musculoskeletal US has previously been demonstrated valuable in PMR diagnosis (22) and factor analysis in the classification cohort revealed a potential of US of hip and shoulders in classifying PMR patients. By including US in the classification criteria scoring algorithm, a score of ≥5 provided an improved c-statistic, with sensitivity of 66% and specificity of 81% for discriminating all comparison subjects from PMR, a specificity of 89% discriminating shoulder conditions from PMR and of 70% discriminating rheumatoid arthritis (RA) from PMR. These provisional classification criteria hence exceeded the threshold of 80% that is conventionally considered to be useful in clinical decision-making. Yet, the discrimination between RA and PMR was only modest. This may be due to a genuine overlap between the conditions. Therefore, the provisional 2012 classification criteria advocate a stepwise approach with exclusion of PMR mimics, with the essential criterion of new-onset bilateral shoulder pain not better explained by an alternative diagnosis.

### VALIDATION OF THE 2012 PROVISIONAL CLASSIFICATION CRITERIA FOR POLYMYALGIA RHEUMATICA

The 2012 classification criteria for PMR are considered provisional since they were only validated in the same cohort as the candidate criteria they had emerged from. So far, evaluation of the provisional classification criteria has been conducted in few studies. Özen et al. performed a prospective longitudinal observational study in which 133 treatment-naïve patients older than 50 years of age, with new-onset bilateral shoulder pain and elevated acute-phase reactants were included (23). A one-year follow-up to confirm diagnosis was completed. The provisional classification criteria were found to have a high sensitivity and low specificity (89.5% and 57.7%, respectively) and specificity even decreased by applying the US criteria in the scoring algorithm (specificity with US was 52%). Specificity increased when tested against the sub-population of controls with RA, but only to a minor extent when US was applied (66.7% and 53.8% with and without US in the scoring algorithm, respectively). However, US was only performed in less than half of the patients. Moreover, even though shoulder pain was described as new-onset, inclusion criteria allowed

<table>
<thead>
<tr>
<th>Table I - 2012 Provisional PMR classification criteria-scoring algorithm.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Required criteria: age 50 years or older, bilateral shoulder aching and abnormal CRP and/or ESR</strong></td>
</tr>
<tr>
<td><strong>Points without US (0-6)</strong></td>
</tr>
<tr>
<td>Morning stiffness duration &gt;45 min</td>
</tr>
<tr>
<td>Hip pain or limited range of motion</td>
</tr>
<tr>
<td>Absence of RF or ACPA</td>
</tr>
<tr>
<td>Absence of other joint involvement</td>
</tr>
<tr>
<td>At least one shoulder with subdeltoid bursitis and/or biceps tenosynovitis and/or glenohumeral synovitis (either posterior or axillary) and at least one hip with synovitis and/or trochanteric bursitis</td>
</tr>
<tr>
<td>Both shoulders with subdeltoid bursitis, biceps tenosynovitis or glenohumeral synovitis</td>
</tr>
</tbody>
</table>

*Bilateral shoulder aching should not be better explained by another condition. A score of 4 or more is categorized as PMR in the algorithm without US and a score of 5 or more is categorized as PMR in the algorithm with US; *Optional ultrasound criteria. ACPA, anticitrullinated protein antibody; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; PMR, polymyalgia rheumatica; RF, rheumatoid factor; US, ultrasound.
24-weeks duration as opposed to the specified 12-weeks duration by the provisional classification criteria and most essentially, it was not specified whether the requirement of exclusion of PMR mimics was implemented, namely that bilateral shoulder pain should not be better explained by another disease. Not adhering to this central conditional requirement may have significantly influenced classification specificity. The provisional classification criteria were also evaluated in two retrospective studies. In 54 patients with a clinical diagnosis of recent-onset PMR, Weigand et al. found a high sensitivity of the classification criteria but without any improvement by adding US to the scoring algorithm (sensitivity 81.5% and 85.2% with and without US, respectively) (24). Macchioni et al. did a case-control study of 136 PMR-suspected patients with confirmed diagnosis by 12-months follow-up and 149 non-PMR patients from an early arthritis clinic (thereof 94 RA). In all cases, US was performed unblinded by the same rheumatologist and musculoskeletal US expert. A sensitivity of 92.6% and a specificity of 81.5% of the clinical criteria was obtained. As in the original classification cohort, the specificity in the Macchioni study increased by adding US (91.3%) but remained high against the RA controls when the scoring algorithm with US was applied (89.9%) (25). The increase in specificity of US in this study possibly relates to a single, highly skilled operator; highlighting the need for greater US training in clinical practice and as a prerequisite for clinical trials.

A robust validation of the 2012 provisional classification criteria requires a prospective longitudinal observation study of inflammatory shoulder conditions mimicking PMR in GC-naïve patients more than 50 years of age with elevated inflammatory markers. The diagnostic reference standard should be a clinical expert diagnosis in which a stepwise diagnostic approach has been applied as recently recommended (7). The control cohort must reflect the non-PMR polymyalgic patients that clinicians need to differentiate from patients with PMR. This, of course depends on the setting, that is, whether the diagnosis is established by a GP, other specialist, or by a less or more experienced rheumatologist. However, with the purpose of the classification criteria in mind, the setting may most often be in secondary care as one of the latter two. Also, built into the provisional classification criteria is the stepwise approach and the requirement that bilateral shoulder pain, the central required criteria, may not be better explained by another disease. Shoulder conditions definitely mimic PMR, particularly in a primary care setting, but may also be the mimicking condition easiest to recognize early in the diagnostic process as opposed to the inflammatory conditions (inflammatory arthritis, spondyloarthropathies, vasculitides or connective tissue diseases) in the secondary care setting. This highlights the need for long-term validation of the provisional classification criteria in a population of patients who initially presented with an inflammatory polymyalgic syndrome.

A long follow-up time that exceeds the median time of GC treatment in order to ascertain PMR diagnosis will be essential, since mimicking conditions, such as inflammatory joint diseases and GCA, may be masked by GC treatment, and therefore not become apparent until GC dose is lowered or discontinued. All variables included in the scoring algorithm need assessment including US performed by a trained sonographer blinded to the clinical features and management. Such US assessment should incorporate the advances in sonographic equipment and techniques (such as both MSK and vascular assessments).

REFINING OF THE 2012 PROVISIONAL CLASSIFICATION CRITERIA FOR POLYMYALGIA RHEUMATICA

The salient question is whether the 2012 provisional classification criteria for PMR could also be refined to become even more specific. The main challenge of a scoring algorithm is to discriminate PMR from other inflam-
Inflammatory arthritis’ especially late-onset seronegative RA (26), and from GCA which is reported to occur in 20-30% of PMR cases and may be subclinical (27-30). Even though considered different entities of the same disease, the diagnosis of GCA should overrule PMR diagnosis, since the presence of arteritis dictates additional treatment (31). Ideally, an updated classification criteria should be able to discriminate the two (PMR versus LV-GCA) at presentation. Vascular US of temporal arteries has been studied in several cohorts and has proven highly sensitive and specific in the diagnosis of GCA provided high-end US equipment, vascular pre-settings and a trained sonographer (32, 33) are available. US has also been proven non-inferior to temporal artery biopsy (34) and in expert hands can replace biopsy in patients with high clinical probability of GCA and a positive US or in patients with low clinical probability of GCA and a negative US (35). The assessment of signs of vasculitis in the axillary artery has shown to significantly improve the sensitivity of US (36) (Figure 1). Recently, OMERACT definitions of normal and vasculitic lesions in temporal and axillary arteries have been validated in both web based (37) and patient based (38) reliability exercises.

Considering the scientific evidence, US examination of large and medium-sized vessels may be a specific approach with which subclinical GCA may be diagnosed. The low cost, high availability and no risk, makes it necessary to further explore the added value of vascular US in the classification criteria. Added consideration should be given to the role of other imaging modalities such as fluorine-18-fluorodeoxyglucose (FDG) PET/CT, CTA or MRA in polymyalgia situations (such as constitutional symptoms, relapsing disease, persistent raised inflammatory markers) which may indicate underlying large vessel-GCA (39).

In both RA and PMR, synovitis, bursitis and tenosynovitis are common findings. In PMR mainly shoulders, hips and tendons of hands are involved whereas in RA the arthritis is polyarticular and symmetric and usually involves peripheral small joints (40). However, in late-onset seronegative RA involvement of proximal joints is more common, causing a polymyalgic onset challenging the diagnostic interpretation (41-43).

In the scoring algorithm two criteria are particularly aimed at distinguishing PMR from RA, namely the absence of other joint involvement and negative tests for rheumatoid factor (RF) and anti-citrullinated protein antibody (ACPA). The scoring algorithm also requires a specific US pattern recognition, namely either the composite finding of pathology in one shoulder and one hip or in both shoulders. Nevertheless, the inclusion of US of hips and shoulders gained minor

![Figure 1](image-url) - Ultrasound in the diagnosis of GCA: vascular ultrasound examination of large and medium-sized vessels can reveal a homogenous hypoechoic wall swelling in A) the superficial temporal artery and/or B) the axillary artery due to subclinical GCA in PMR patients.
improvement in specificity. This may not be surprising since only dichotomous categories e.g. the presence or absence of synovitis, bursitis, tenosynovitis according to EU-LAR definitions, findings that are also common in RA, were assessed. Other studies on imaging modalities in PMR diagnosis have led to similar results regarding the diagnostic specificity of synovial inflammation in PMR patients compared to other inflammatory arthritis’ (44-47).

On the other hand, detailed qualitative US studies of such lesions may help to discriminate PMR from non-PMR inflammatory disease. Studies suggest that the PMR synovial inflammation may be exudative as opposed to a more proliferative bursitis, synovitis seen in RA (48, 49). Moreover, the increasing spatial resolution and emergence of functional imaging technologies also highlighted potential discriminating findings (50). In MRI studies, the presence of extracapsular, i.e. the soft tissues outside synovial cavities, inflammatory changes has been described in both shoulders, pelvic girdle and MCP joints in PMR, whereas it seems to be seldom in RA (20, 21, 23, 25).

The spatial resolution of fluorine-18-fluorodeoxyglucose (FDG) PET/CT studies and the short timespan between the PET and CT imaging does not always allow rigorous co-localization of FDG uptake and anatomical structures. However, periarticular uptake patterns in shoulders and hips of PMR patients (Figure 2), different from uptake patterns in RA patients, has been described and could reflect the extracapsular pattern found in MR studies (47, 51, 52). The pelvic girdle also shows localized FDG uptake adjacent to greater trochanter, ischial tuberosity, in front of hip joint/pectineus muscle and at ventral edges of the pubic bone reflecting either bursitis or enthesis (47, 52). Another finding from FDG PET/CT studies is inflammation along the spine; most frequently in the cervical but also seen in the lumbar region (53). The increased FDG uptake is most pronounced in the spinous processes and is considered a interspinous bursitis as also described in a MR study of PMR patients with neck pain (54). The combination of pelvic girdle involvement and inter-spinous bursitis discriminates PMR from RA (47, 52) but not from spondylo-arthritis (55).

Nevertheless, the idea of a specific topographic PMR disease pattern potentially provides the basis for a composite PMR-PET score that may discriminate PMR from non-arthritis or RA controls, respectively. However, existing studies are small and the concept needs validation in a large polymyalgic onset cohort.

So far, diagnostic biomarkers have not been found to be useful in PMR diagnosis. A recent study showed serum CXCL9, CXCL10, B-cell activating factor and IL-6 to accurately distinguish new-onset GCA and PMR patients from healthy controls. Even more interesting, these markers were raised in patients with ESR below classification criteria threshold. Several studies found IL-6 to be significantly higher in untreated GCA and PMR patients than healthy controls. However, IL-6 levels cannot distinguish PMR from EORA (56-58).

Figure 2 - The topographic distribution of lesions and the uptake intensity are parameters potentially discriminating PMR patients from PMR mimics on FDG PET/CT. FDG uptake in shoulder and hips due to degenerative joint disease (A) can cause FDG uptake in the same anatomical structures as PMR lesions but with only low uptake intensity. In elderly-onset RA, articular uptake in proximal joints is high intensity (B) and differs from the uptake pattern in PMR, which is more periarticular (C). A combination of several sites of localized FDG uptake in the pelvic girdle (adjacent to greater trochanter, ischial tuberosity, in front of hip joint/pectineus muscle, at ventral edges of the pubic bone) reflecting either bursitis or enthesis and the spine (interspinous bursitis) also discriminates PMR (C) from RA (A).
Whether the serum biomarkers can differentiate PMR from patients other inflammatory conditions, malignancies and infections remains unknown. In that respect, a recent study found ferritin autoantibodies to have high sensitivity in untreated PMR with very low false positive rate in other inflammatory rheumatic diseases, including EORA, and healthy controls (59). Validation of such findings in larger cohorts may reveal a potential role of biomarkers in diagnosis and/or classification of PMR. To summarize, we feel that quantitative cut-offs of the MSK US lesions, detailed qualitative assessment of these findings and/or assessment of additional locations (spine, pelvic girdle) previously described in PMR may be beneficial in discriminating PMR from other mimicking diseases and potentially add to classification criteria’s specificity.

PET/CT may not be an attractive modality to incorporate into a scoring algorithm for classification purposes due to its cost, availability and radiation. However, it provides new insight into the topographic distribution of PMR lesions. Large longitudinal polymyalgic cohort studies may clarify the role of imaging modalities (additional to US) to visualize PMR inflammatory lesions based on available techniques, thereby improving the classification criteria.

In conclusion, a thorough validation of the 2012 provisional classification criteria is needed with a prospective longitudinal observation study of inflammatory shoulder conditions mimicking PMR in patients more than 50 years of age, with a long follow-up time that exceeds the median time of GC treatments in order to confirm final diagnoses. The study should include assessment of all variables included in the scoring algorithm and comprehensive US performed blinded by a trained sonographer. The US assessment should include quantitative and qualitative parameters as well as vascular examination that may refine the criteria and add to its specificity. Such a study will clarify the contribution of other imaging modalities such as FDG PET-CT towards verification of PMR inflammatory lesions. Besides from validation of the 2012 provisional classification criteria for PMR, the outlined prospective observational study will also provide the essential information needed to establish the reproducibility, redundancy and validity of previously proposed measures for defining remission and response in PMR (2, 4, 5). Taken together, the framework for high quality clinical trials in PMR will then be available.

Acknowledgements: the authors would like to acknowledge Lars Christian Gormsen, Department of Nuclear Medicine and PET centre, Aarhus University Hospital, Denmark, for providing FDG PET/CT pictures.

REFERENCES

34. Luqmani R, Lee E, Singh S, et al. The role of


