



2012 provisional classification criteria for polymyalgia rheumatica: a European League Against Rheumatism/American College of Rheumatology collaborative initiative

Bhaskar Dasgupta,¹ Marco A Cimmino,² Hilal Maradit-Kremers,³ Wolfgang A Schmidt,⁴ Michael Schirmer,⁵ Carlo Salvarani,⁶ Artur Bachta,⁷ Christian Dejaco,⁸ Christina Duftner,^{5,9} Hanne Slott Jensen,¹⁰ Pierre Duhaut,¹¹ Gyula Poór,¹² Novák Pál Kaposi,¹³ Peter Mandl,¹⁴ Peter V Balint,¹⁴ Zsuzsa Schmidt,¹² Annamaria Iagnocco,¹⁵ Carlotta Nannini,¹⁶ Fabrizio Cantini,¹⁶ Pierluigi Macchioni,⁶ Nicolò Pipitone,⁶ Montserrat Del Amo,¹⁷ Georgina Espígol-Frigolé,¹⁸ Maria C Cid,¹⁸ Víctor M Martínez-Taboada,¹⁹ Elisabeth Nordborg,²⁰ Haner Direskeneli,²¹ Sibel Zehra Aydin,²¹ Khalid Ahmed,²² Brian Hazleman,²³ Barbara Silverman,²³ Colin Pease,²⁴ Richard J Wakefield,²⁴ Raashid Luqmani,²⁵ Andy Abril,²⁶ Clement J Michet,²⁷ Ralph Marcus,²⁸ Neil J Gonter,²⁸ Mehrdad Maz,²⁹ Rickey E Carter,³ Cynthia S Crowson,^{3,27} Eric L Matteson^{3,27}

► Additional supplementary tables are published online only. To view the files please visit the journal online (<http://ard.bmj.com>)

For numbered affiliations see end of article

Correspondence to

Eric L Matteson, Department of Internal Medicine, Division of Rheumatology, Mayo Clinic, 200 1st Street SW, Rochester, Minnesota, USA; matteson.eric@mayo.edu

Received 21 June 2011

Accepted 31 October 2011

This article is published simultaneously in the April 2012 issue of *Arthritis & Rheumatism*.

ABSTRACT

The objective of this study was to develop EULAR/ACR classification criteria for polymyalgia rheumatica (PMR). Candidate criteria were evaluated in a 6-month prospective cohort study of 125 patients with new onset PMR and 169 non-PMR comparison subjects with conditions mimicking PMR. A scoring algorithm was developed based on morning stiffness >45 minutes (2 points), hip pain/limited range of motion (1 point), absence of RF and/or ACPA (2 points), and absence of peripheral joint pain (1 point). A score ≥ 4 had 68% sensitivity and 78% specificity for discriminating all comparison subjects from PMR. The specificity was higher (88%) for discriminating shoulder conditions from PMR and lower (65%) for discriminating RA from PMR. Adding ultrasound, a score ≥ 5 had increased sensitivity to 66% and specificity to 81%. According to these provisional classification criteria, patients ≥ 50 years old presenting with bilateral shoulder pain, not better explained by an alternative pathology, can be classified as having PMR in the presence of morning stiffness >45 minutes, elevated CRP and/or ESR and new hip pain. These criteria are not meant for diagnostic purposes.

Polymyalgia rheumatica (PMR) is a common inflammatory rheumatic disease of older individuals and a common indication for long-term corticosteroid therapy.^{1–3} PMR is also subject to wide variations of clinical practice, due to the considerable uncertainty related to diagnosis, course, and management in primary and secondary care.^{4–7} There is no diagnostic laboratory test, inflammatory markers are not specific, and clinicians often turn to the corticosteroid response as a ‘test of treatment’ to establish the diagnosis.^{1,2,8}

Difficulties in diagnosing and classifying patients with PMR are inherent in its definitions.^{9,10} The proximal pain and stiffness syndrome can occur at presentation in many other rheumatological inflammatory illnesses in older people.^{12,7,9} Approximately half of patients diagnosed with PMR may have

distal manifestations such as peripheral arthritis, hand swelling with pitting oedema and carpal tunnel syndrome.^{1,2,7,11} Polymyalgic presentation is common in late-onset rheumatoid arthritis (RA) and spondyloarthritis and is also associated with giant cell arteritis in 10–30% of cases.^{1,2} Heterogeneity in the disease course, uncertainty regarding disease assessment parameters, and evolution of alternative diagnoses on follow-up complicate the management of PMR.^{9,12–15} For the above reasons a safe and specific approach preferring a relative underdiagnosis to an overdiagnosis is needed in PMR.^{9,11}

Uniform responsiveness to low doses of corticosteroids has been assumed to be a cardinal feature of PMR. However, there is little hard evidence to substantiate this assertion. A previous report showed that 3 weeks after starting prednisolone 15 mg a day only 55% showed a complete response to therapy.¹⁰ This also emphasises that clinical trials of novel effective agents are needed in PMR.

The lack of standardised classification criteria has been a major factor hampering the development of rational therapeutic approaches^{12,16,17} and causing difficulties in evaluating patients in clinical studies. In response to a European League Against Rheumatism (EULAR)/American College of Rheumatology (ACR) initiative, a criteria development work group convened in 2005.

A systematic literature review, a three-phase hybrid consensus process and a wider survey were undertaken to identify candidate criteria items.¹⁸ The first stage consisted of a meeting of 27 international experts who anonymously rated 68 potential criteria identified through literature review. In the second round the experts were provided with the results of the first round and re-rated the criteria items. In the third round, the wider acceptance of the chosen criteria (>50% support) was evaluated using a survey of 111 rheumatologists and 53 non-rheumatologists in North America and western Europe.



This paper is freely available online under the BMJ Journals unlocked scheme, see <http://ard.bmj.com/info/unlocked.dtl>

In round three, over 70% of respondents agreed on the importance of seven core criteria (all achieving 100% support in round two). These were aged 50 years or older, symptom duration 2 weeks or longer, bilateral shoulder and/or pelvic girdle aching, duration of morning stiffness more than 45 min, elevated erythrocyte sedimentation rate (ESR), elevated C-reactive protein (CRP) and rapid corticosteroid response. More than 70% agreed on assessing pain and limitation of shoulder (84%) and/or hip (76%) on motion, but agreement was low for peripheral signs (eg, carpal tunnel syndrome, tenosynovitis, peripheral arthritis).

The group reached consensus on the need for a prospective cohort study to evaluate the disease course from presentation in patients included on the basis of proximal pain and stiffness with evaluations over a 6-month period while receiving a standardised corticosteroid treatment regimen.^{7 18–25} The group also agreed to assess musculoskeletal ultrasound as part of the PMR classification criteria. In this paper, we present the results from this prospective study and propose new classification criteria for PMR.

STUDY DESIGN AND METHODS

Consensus decisions about study design

A priori, the work group decided that a specific approach should be adopted for classifying newly presenting patients with bilateral shoulder pain as PMR.¹⁸ The group agreed on the following:

1. Patients presenting with polymyalgic syndrome should have stepwise evaluation on the basis of inclusion and exclusion criteria, response to a standardised corticosteroid challenge and follow-up confirmation. The criteria items would be agreed upon clinical features and laboratory investigations.
2. The need to standardise response to corticosteroid therapy in PMR. Because the goal of classification criteria is to identify patients for enrollment into clinical trials before any treatment, the response to corticosteroid therapy should be used in verifying the classification of a patient as having PMR, although it should not be used as a classification criterion. Because scientific evidence is poor as to what constitutes such a response it was agreed (>75% agreement) this response would be defined as greater than 75% global response in clinical and laboratory parameters within 7 days to corticosteroid challenge with oral 15 mg prednisone or prednisolone and subsequent resolution of inflammatory indices.¹⁸
3. That a prospective study would be needed to evaluate the disease course from presentation in patients included on the basis of the mandatory 'core' criteria of proximal pain and stiffness. New-onset bilateral shoulder pain was selected as the main eligibility criterion as the percentage of PMR presenting with hip pain without shoulder pain was very small (<5%), and that as hip girdle pain is due to a wide range of conditions it would require the enrollment of an impracticably large number of comparator patient groups.¹⁸ The study would evaluate at prespecified intervals symptoms, examination, investigations and their evolution with standardised corticosteroid treatment in a prospective cohort of patients with new-onset bilateral shoulder pain (comparing the PMR case cohort with the comparator cohort of mimicking conditions) over a 6-month period.
4. Musculoskeletal ultrasound should be evaluated in a substudy as a feasible mode of investigation of possible PMR. A secondary objective of this substudy would be the evaluation of

clinical and patient-based outcomes in PMR over a 6-month period.^{26 27}

Study population

The study was a prospective cohort study that included a cohort of patients with new-onset PMR and a comparison cohort of non-PMR patients with various conditions mimicking PMR. Study subjects were recruited from 21 community-based and academic rheumatology clinics in 10 European countries and the USA. Inclusion criteria for PMR patients were age 50 years or older, new-onset bilateral shoulder pain and no corticosteroid treatment (for any condition) within 12 weeks before study entry, fulfilling all the inclusion and exclusion criteria defined by our previous report and in accordance with the judgement of the participating investigator that the patient had PMR.¹⁸ Every effort was made to choose patients across the spectrum of disease severity. Corticosteroid treatment for PMR patients was initiated according to a predefined treatment protocol starting with 15 mg a day oral prednisone for weeks 1 and 2, 12.5 mg a day for weeks 3 and 4, 10 mg a day for weeks 6–11, 10 mg/7.5 mg every other day for weeks 12–15, 7.5 mg a day for weeks 16–25 and tapering according to treatment response from week 26 onwards. The gold standard for the pre-steroid diagnosis of PMR was established as above at presentation and when the diagnosis was maintained without an alternative diagnosis at week 26 of follow-up.

The non-PMR comparison cohort included conditions representative of the types that need to be distinguished from PMR, in both primary and secondary care. Inclusion criteria for the non-PMR comparison cohort were age 50 years or older, new-onset bilateral shoulder pain and a diagnosis of either inflammatory or non-inflammatory conditions, including new-onset RA, connective tissue diseases, various shoulder conditions (eg, bilateral rotator cuff syndrome and/or adhesive capsulitis, rotator cuff tear, glenohumeral osteoarthritis), fibromyalgia, generalised osteoarthritis and others. Patients known to have the condition for more than 12 weeks before the baseline evaluation (except fibromyalgia and chronic pain) were not eligible for inclusion. PMR patients with clinical suspicion of giant cell arteritis were included as part of the comparison cohort because these patients required different corticosteroid doses. Patients in the comparison cohort were included on the basis of clinician diagnosis and not on formal criteria. No guidelines were provided for treatment of the conditions in the comparison cohort.

Ethics board approval was obtained at all participating institutions before initiation of the study, and all participants gave written informed consent before enrollment.

Follow-up and data collection

PMR patients were evaluated at baseline, and at 1, 4, 12 and 26 weeks. At each follow-up visit, clinical evaluation included response to corticosteroid therapy and opinion on the emergence of alternative diagnoses. Patients not considered as having PMR at any time were evaluated and treated according to accepted clinical practice. They were excluded from the PMR cohort and included in the non-PMR comparison cohort. Patients in the comparison cohort were evaluated at baseline and at 26 weeks.

Data were collected using standardised data collection forms and questionnaires translated into national languages. Data collection included the candidate inclusion/exclusion criteria items for classification of PMR, physical examination and assessment

of corticosteroid response. Criteria items were age 50 years or older, symptom duration 2 weeks or more, bilateral shoulder and/or pelvic girdle aching, recent weight loss greater than 2 kg, duration of morning stiffness more than 45 min, elevated ESR, elevated CRP and rapid response of symptoms to corticosteroids (>75% global response within 1 week to prednisolone/prednisone 15–20 mg a day). Pain was assessed using a horizontal 100 mm visual analogue scale (VAS) in four separate locations (shoulder, pelvic, neck and overall) with zero indicating no pain and 100 indicating worst pain. Morning stiffness (in the past 24 h) was assessed in minutes; functional status and quality of life were assessed by the modified health assessment questionnaire (MHAQ) and short form 36. A 100-mm VAS was also used for recording global wellbeing measures (patient and physician global) and fatigue. Physical examination included the presence or absence of tenderness, pain on movement and limitation of the shoulders and hips. Aspects of corticosteroid therapy including dose, therapeutic response and change in dose and therapy discontinuation were documented.

Data regarding laboratory measures (including ESR, CRP, rheumatoid factor (RF) and anticitrullinated protein antibody (ACPA)) were obtained from clinically ordered tests performed at each study centre. As the laboratory assays used at each centre varied, test results were classified as normal/abnormal using the reference ranges from each centre (see supplementary appendix table S1, available online only). Both PMR and non-PMR subjects underwent ultrasound evaluation of shoulders and hips at baseline and at 26 weeks. Evaluations were made according to EULAR guidelines²⁸ to assess for features previously reported to be associated with PMR, including bicipital tenosynovitis, subacromial and subdeltoid bursitis, trochanteric bursitis and glenohumeral and hip effusion. A rheumatologist or radiologist experienced in musculoskeletal ultrasound performed the ultrasound examination using linear probes with the frequency range 6–10 MHz for shoulders and linear or curved array probes with the frequency range 5–8 MHz for hips.

Statistical analysis

Descriptive statistics (means, percentages, etc.) were used to summarise the candidate criteria data. Demographics (age and gender) and candidate criteria were compared between PMR and comparison subjects using χ^2 and rank sum tests. Several statistical approaches were considered in order to develop a scoring algorithm for PMR and to assess the proposed classification criteria in patients judged by expert clinician investigators to have PMR 26 weeks after enrolment.²⁹

First, logistic regression models to distinguish PMR patients from all comparison subjects and each subset of comparison subjects were examined. The c-statistic, a measure of concordance analogous to the area under the receiver operating characteristic curve, was used to assess the ability of each individual criterion to distinguish between PMR and comparison subjects. The c-statistic ranges from 0.5 to 1, with 0.5 indicating a criterion that provides no information. The sensitivity, specificity and the positive and negative likelihood ratios were also examined.

Second, exploratory factor analysis was used to examine the interdependencies between the candidate criteria.³⁰ Maximum likelihood factor analysis with varimax rotation was used. Maximum likelihood tests were used to examine goodness of fit (eg, to determine the number of factors). This method is thought to be superior to the eigenvalue greater than 1 or Catell's scree plot method for selecting the number of factors.³⁰ For each factor, variables with factor loadings greater than 0.5 were

examined and found to represent a similar domain. This technique allowed a reduction of the number of variables, as one variable from each factor was examined in multivariable logistic regression models. In addition, to avoid discarding a relevant domain as identified by expert consensus, a few of the variables with factor loading between 0.4 and 0.5 were also considered in the multivariable models.

Classification trees including the variables determined by the factor analysis were also considered, but were not found to be optimal for distinguishing PMR from comparison subjects.^{31,32} An integer scoring algorithm was defined based on the odds ratios in the final multivariable logistic regression model. Performance characteristics (sensitivity, specificity, etc.) of this scoring algorithm were assessed. In addition, the utility of ultrasound assessments for classifying PMR patients was examined using factor analysis and adding potential ultrasound criteria to the scoring algorithm. Odds ratios for clinical criteria varied somewhat in the scoring algorithm that included ultrasound criteria. Scoring weights based on both models were considered and found to perform similarly, so a common set of scoring weights was used for the clinical items in both algorithms in order to ease comparison and application of the criteria.

Finally, gradient boosting regression tree models, which are a machine learning technique, were examined to determine whether a better prediction could be achieved using a more complex algorithm.³³

RESULTS

At baseline, 128 patients were recruited into the PMR cohort and 184 patients were recruited into the non-PMR comparison cohort. During follow-up, 10 PMR patients were reclassified as not having PMR and moved into the non-PMR cohort. Similarly, eight non-PMR comparison cohort patients were reclassified as having PMR and moved into the PMR cohort. In addition, seven patients (one PMR and six non-PMR) were excluded due to missing information, and two non-PMR subjects with age less than 50 years and nine non-PMR subjects with no shoulder pain were also excluded. Therefore, the final analysis was based on 125 PMR and 169 non-PMR subjects.

The diagnoses of the 169 non-PMR subjects (see supplementary appendix table S2, available online only) were new-onset RA (49, 29%), new-onset other seronegative arthritis (20, 12%), new-onset connective tissue diseases or vasculitis (9, 5%), shoulder conditions (52, 31%), chronic pain (26, 15%), infection (5, 3%), previously undiagnosed malignancy (4, 2%), and two each of endocrinopathy and neurological disorders.

The distribution of the candidate classification criteria for the 125 PMR patients and 169 non-PMR comparison subjects (all, RA only and shoulder condition only) is displayed in table 1. Criteria items present in over 80% of the PMR subjects were 2 weeks or greater duration of symptoms, bilateral shoulder pain and elevated CRP and/or ESR. Relevant clinical features that best discriminated RA from PMR were peripheral synovitis, the presence of RF and/or ACPA and hip pain/limited range of motion. Features best discriminating shoulder conditions from PMR were hip pain/limited range of motion, morning stiffness and elevated CRP and/or ESR.

Development of a scoring algorithm

Table 2 shows the results of univariate logistic regression models to distinguish PMR from all comparison subjects, RA only and shoulder conditions only. Criteria items related to hip involvement (pain, tenderness, limited movement) had

Table 1 Distribution of candidate criteria for 125 PMR patients and 169 comparison subjects*

| | PMR (N=125) | Comparison subjects (N=169) | p Value | RA (N=49) | p Value | Shoulder conditions† (N=52) | p Value |
|---|---------------|-----------------------------|---------|--------------|---------|-----------------------------|---------|
| Age, years, mean±SD | 71.5±8.4 | 67.0±9.6 | <0.001 | 68.6±10.0 | 0.06 | 67.0±10.6 | 0.017 |
| Sex (female) | 72 (58) | 112 (66) | 0.13 | 28 (57) | 0.96 | 35 (67) | 0.23 |
| Duration of symptoms ≥2 weeks | 121 (97) | 163 (96) | 0.87 | 47 (96) | 0.77 | 51 (98) | 0.64 |
| Duration of symptoms, weeks, mean±SD | 10.0±7.2 | 14.4±19.1 | 0.17 | 13.8±16.1 | 0.62 | 12.9±11.8 | 0.21 |
| Bilateral shoulder aching | 124 (99) | 162 (96) | 0.08 | 46 (94) | 0.035 | 50 (96) | 0.15 |
| Bilateral pelvic girdle (hip) aching | 91 (73) | 90 (54) | 0.001 | 28 (57) | 0.046 | 21 (40) | <0.001 |
| Neck aching | 71 (57) | 92 (54) | 0.69 | 29 (59) | 0.78 | 27 (52) | 0.55 |
| Morning stiffness duration >45 min‡ | 92 (77) | 68 (43) | <0.001 | 33 (69) | 0.25 | 10 (20) | <0.001 |
| Morning stiffness, min, median (IQR) | 120 (60, 240) | 30 (5, 120) | <0.001 | 60 (30, 180) | 0.11 | 10 (0,30) | <0.001 |
| Weight loss of >2 kg | 45 (36) | 40 (24) | 0.021 | 16 (33) | 0.68 | 4 (8) | <0.001 |
| Shoulder pain or limited range of motion | 121 (97) | 158 (93) | 0.20 | 47 (96) | 0.77 | 49 (94) | 0.42 |
| Hip pain or limited range of motion | 71 (57) | 59 (35) | <0.001 | 15 (31) | 0.002 | 12 (23) | <0.001 |
| Shoulder tenderness | 96 (77) | 126 (75) | 0.66 | 40 (82) | 0.49 | 34 (65) | 0.12 |
| Hip tenderness | 59 (47) | 47 (28) | 0.001 | 12 (24) | 0.006 | 9 (17) | <0.001 |
| Carpal tunnel syndrome | 19 (15) | 27 (16) | 0.90 | 11 (22) | 0.28 | 8 (15) | 0.99 |
| Peripheral synovitis (distal swelling, tenosynovitis or arthritis)‡ | 48 (39) | 78 (46) | 0.20 | 41 (84) | <0.001 | 11 (21) | 0.024 |
| Other joint pain‡ | 63 (51) | 109 (66) | 0.011 | 40 (85) | <0.001 | 29 (57) | 0.50 |
| Abnormal CRP and/or ESR‡ | 116 (96) | 99 (63) | <0.001 | 41 (85) | 0.017 | 18 (41) | <0.001 |
| Presence of RF and/or ACPA‡ | 11 (10) | 37 (25) | 0.004 | 19 (41) | <0.001 | 5 (12) | 0.79 |
| Abnormal serum protein electrophoresis‡ | 43 (52) | 32 (35) | 0.027 | 9 (36) | 0.17 | 9 (35) | 0.13 |
| MHAQ, mean±SD | 1.2±0.6 | 0.8±0.6 | <0.001 | 1.1±0.7 | 0.32 | 0.5±0.6 | <0.001 |

*Values are number (percentage) unless specified. Percentages for laboratory results are the number of abnormal tests results divided by the number of patients tested.

ACPA, anticitrullinated protein antibody; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; MHAQ, modified health assessment questionnaire; PMR, polymyalgia rheumatica; RF, rheumatoid factor.

†Bilateral rotator cuff syndrome and/or adhesive capsulitis, rotator cuff tear, glenohumeral osteoarthritis.

‡Data not available on all subject.

Table 2 Univariate logistic regression models to distinguish subjects with PMR from comparison subjects

| Variable | PMR vs all comparison subjects | | PMR vs RA | | PMR vs shoulder conditions* | |
|--|--------------------------------|------|---------------------|------|-----------------------------|------|
| | Odds ratio (95% CI) | c | Odds ratio (95% CI) | c | Odds ratio (95% CI) | c |
| Duration of symptoms ≥2 weeks | 1.1 (0.3 to 4.0) | 0.50 | 1.3 (0.2 to 7.3) | 0.50 | 0.6 (0.1 to 5.4) | 0.51 |
| Shoulder pain or limited range of motion | 2.1 (0.7 to 6.8) | 0.52 | 1.3 (0.2 to 7.3) | 0.50 | 1.9 (0.4 to 8.6) | 0.51 |
| Shoulder tenderness | 1.1 (0.7 to 1.9) | 0.51 | 0.7 (0.3 to 1.7) | 0.52 | 1.8 (0.9 to 3.6) | 0.56 |
| Hip pain or limited range of motion | 2.5 (1.5 to 3.9) | 0.61 | 3.0 (1.5 to 6.0) | 0.63 | 4.4 (2.1 to 9.1) | 0.67 |
| Hip tenderness | 2.3 (1.4 to 3.8) | 0.60 | 2.8 (1.3 to 5.8) | 0.61 | 4.3 (1.9 to 9.5) | 0.65 |
| Neck aching | 1.1 (0.7 to 1.8) | 0.51 | 0.9 (0.5 to 1.8) | 0.51 | 1.2 (0.6 to 2.3) | 0.52 |
| Morning stiffness duration >45 min | 4.5 (2.6 to 7.7) | 0.67 | 1.5 (0.7 to 3.3) | 0.54 | 13.6 (6.0 to 31) | 0.79 |
| Weight loss >2 kg | 1.8 (1.1 to 3.0) | 0.56 | 1.2 (0.6 to 2.4) | 0.52 | 6.8 (2.3 to 19.9) | 0.64 |
| Carpal tunnel syndrome | 1.0 (0.5 to 1.8) | 0.50 | 0.6 (0.3 to 1.5) | 0.54 | – | – |
| Peripheral synovitis (distal swelling, tenosynovitis or arthritis) | 0.7 (0.5 to 1.2) | 0.54 | 0.1 (0.08 to 0.3) | 0.72 | 2.4 (1.1 to 5.0) | 0.59 |
| Other joint pain | 0.5 (0.3 to 0.9) | 0.57 | 0.2 (0.1 to 0.4) | 0.67 | 0.8 (0.4 to 1.5) | 0.53 |
| Abnormal CRP and/or ESR | 13.8 (5.3 to 36) | 0.67 | 4.0 (1.2 to 13) | 0.55 | 33.5 (11 to 98) | 0.78 |
| Presence of RF and/or ACPA | 0.4 (0.2 to 0.7) | 0.57 | 0.2 (0.07 to 0.4) | 0.66 | 0.9 (0.3 to 2.6) | 0.51 |
| Abnormal serum protein electrophoresis | 2.0 (1.1 to 3.6) | 0.58 | 1.9 (0.8 to 4.8) | 0.58 | 2.0 (0.8 to 5.1) | 0.59 |
| MHAQ (per 1 unit increase) | 2.3 (1.6 to 3.4) | 0.66 | 1.3 (0.7 to 2.2) | 0.55 | 6.7 (3.2 to 14) | 0.78 |

ACPA, anticitrullinated protein antibody; c, c-statistic; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; MHAQ, modified health assessment questionnaire; PMR, polymyalgia rheumatica; RA, rheumatoid arthritis; RF, rheumatoid factor.

*Bilateral rotator cuff syndrome and/or adhesive capsulitis, rotator cuff tear, glenohumeral osteoarthritis.

significant ability to discriminate PMR from all comparison subjects, RA and shoulder conditions based on the c-statistic. Early morning stiffness, MHAQ, weight loss and raised laboratory markers of inflammation distinguished PMR from comparison subjects, particularly those with shoulder conditions. The presence of ACPA and/or RF, peripheral synovitis and joint pains had significant ability (with high c-statistic) to distinguish PMR from RA. In addition, the odds ratios for abnormal CRP and/or ESR were particularly high. This resulted because only five PMR patients did not have abnormal CRP and/or ESR, perhaps reflecting that the diagnosis of PMR is less certain in the presence of normal CRP and ESR. Therefore, abnormal CRP

or ESR would be included as a required criterion in the scoring algorithm for PMR. Similarly, as all subjects in the study were required to have shoulder pain, this was also included as a required criterion for PMR.

Factor analysis revealed that four factors were sufficient to represent all the criteria (see supplementary appendix table S3, available online only). These four factors were: hip pain/tenderness, peripheral synovitis or other joint pain, morning stiffness and shoulder tenderness. Note that duration of symptoms, neck aching, carpal tunnel syndrome, weight loss, presence of RF/ACPA and MHAQ, did not play a prominent role in any of the factors.

Table 3 Multivariable logistic regression models

| Criterion | Model based on factors | | Model based on factors without shoulder tenderness plus presence of RF/ACPA | | Model based on factors plus presence of RF/ACPA and MHAQ | |
|--|------------------------|---------|---|---------|--|---------|
| | Odds ratio (95% CI) | p Value | Odds ratio (95% CI) | p Value | Odds ratio (95% CI) | p Value |
| Hip pain or limited hip range of motion | 2.7 (1.5 to 4.8) | 0.001 | 2.1 (1.1 to 4.0) | 0.019 | 1.6 (0.8 to 3.2) | 0.16 |
| Other joint pain | 0.4 (0.2 to 0.6) | <0.001 | 0.4 (0.2 to 0.7) | 0.002 | 0.3 (0.1 to 0.6) | <0.001 |
| Morning stiffness >45 min duration | 5.2 (2.9 to 9.4) | <0.001 | 6.2 (3.2 to 11.8) | <0.001 | 4.8 (2.4 to 9.6) | <0.001 |
| Shoulder tenderness | 0.9 (0.5 to 1.8) | 0.80 | | | | |
| Presence of RF or ACPA | | | 0.3 (0.1 to 0.8) | 0.009 | 0.3 (0.1 to 0.8) | 0.013 |
| MHAQ, per 1 unit increase | | | | | 2.4 (1.4 to 4.2) | 0.002 |
| Likelihood ratio test for additional terms | | | p<0.001 | | p<0.001 | |
| c-Statistic | 79% | | 81% | | 81% | |

ACPA, anticitrullinated protein antibody; MHAQ, modified health assessment questionnaire; RF, rheumatoid factor.

Table 4 Scoring algorithm with and without optional ultrasound criterion—required criteria: age 50 years or greater, bilateral shoulder aching and abnormal CRP and/or ESR

| Criteria | Clinical criteria (without ultrasound)* | | Criteria including ultrasound† | |
|--|---|--------|--------------------------------|--------|
| | Odds ratio (95% CI) | Points | Odds ratio (95% CI) | Points |
| Morning stiffness duration >45 min | 6.2 (3.2 to 11.8) | 2 | 5.0 (2.8 to 9.1) | 2 |
| Hip pain or limited range of motion | 2.1 (1.1 to 4.0) | 1 | 1.4 (0.8 to 2.6) | 1 |
| Absence of RF or ACPA | 3.0 (1.3 to 6.8) | 2 | 5.2 (2.1 to 12.6) | 2 |
| Absence of other joint pain | 2.7 (1.4 to 5.0) | 1 | 2.2 (1.3 to 4.0) | 1 |
| Ultrasound criteria | | | | |
| 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 | | | 2.6 (1.3 to 5.3) | 1§ |
| Both shoulders with subdeltoid bursitis, biceps tenosynovitis or glenohumeral synovitis | | | 2.1 (1.2 to 3.7) | 1¶ |

ACPA, anticitrullinated protein antibody; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; RF, rheumatoid factor; CI, confidence interval.

*The optimal cut point is 4. A patient with a score of 4 or more is categorised as having polymyalgia rheumatica (PMR).

†The optimal cut point is 5. A patient with a score of 5 or more is categorised as having PMR.

§ P = 0.008.

¶ P = 0.009.

Next, multivariable logistic regression models were used to determine the importance of each criterion when assessed simultaneously (table 3). Three models were considered: (1) including the four factors identified in factor analysis; (2) removing shoulder tenderness and adding presence of RF/ACPA; and (3) subsequently adding MHAQ. While the subsequent additions appeared to be significant, they also negatively impacted the hip pain criteria. The utility of the inclusion of MHAQ was questionable, as hip pain was easier to assess than MHAQ in a clinical setting. Therefore, the second model, which included hip pain, other joint pain, morning stiffness and abnormal RF/ACPA was deemed the best multivariable model.

Additional analyses were performed using classification trees and assessing combinations of criteria. Classification trees were examined using three sets of potential variables: (1) the four identified factors; (2) adding presence of RF/ACPA; and (3) subsequently adding MHAQ. The resulting trees were deemed inadequate. For example, the second tree, which was fit using the same five variables that were included in our scoring algorithm, had a sensitivity of 66% and specificity of 66%. This specificity was lower than the scoring algorithm developed from the logistic models. In addition, this tree only included morning stiffness and absence of RF and/or ACPA, so it was inadequate because it excluded other domains deemed necessary for content validity. Content validity requires that the set of criteria identified is comprehensive. This was the case in all three trees. Therefore, classification trees were deemed to have poorer performance and content validity than the logistic regression models.

A scoring algorithm was developed (table 4) based on the multivariable logistic regression model presented in table 3 and

included morning stiffness for more than 45 min (two points), hip pain/limited range of motion (one point), absence of RF and/or ACPA (two points) and the absence of peripheral joint pain (one point). The score was evaluated using all PMR subjects (including the five with normal CRP/ESR) and all comparison subjects. This was done to account properly for the influence of CRP/ESR on the performance of the scoring algorithm. A score of 4 or greater had 68% sensitivity and 78% specificity for discriminating all comparison subjects from PMR. The specificity was higher (88%) for discriminating shoulder conditions from PMR and lower (65%) for discriminating RA from PMR. The c-statistic for the scoring algorithm was 81%. A total of 40 (32%) PMR patients and 38 (22%) comparison subjects was incorrectly classified. The positive predictive value was 69% and the negative predictive value was 77%. Supplementary appendix table S4 (available online only) shows the sensitivity and specificity for all possible cut points of the scoring algorithm.

Finally, gradient boosting regression tree models, which are a model averaging technique, were examined to determine whether a better prediction could be achieved using a more complex algorithm. The resulting c-statistic from the gradient boosting model was 80%, which was quite comparable to the results for our scoring algorithm. This indicated that we will not be able to make better predictions from these data with another modelling approach.

Ultrasound findings

Ultrasound was performed in 120 PMR subjects, 154 of the comparison subjects (including 46 with RA and 47 with shoulder conditions) and 21 additional controls (not included in our study

Table 5 Ultrasound findings in 120 patients with PMR, 154 comparison subjects (including 46 with RA and 47 with shoulder conditions) and 21 subjects without shoulder conditions

| | PMR (N=120) | Comparison subjects (N=154) | RA (N=46) | Shoulder condition (N=47) | Subjects without shoulder conditions (N=21) |
|--|----------------|--------------------------------|-----------|------------------------------|---|
| At least one shoulder with subdeltoid bursitis, biceps tenosynovitis or glenohumeral synovitis | 83% | 70%** | 78% | 62%** | 19%** |
| Both shoulders with subdeltoid bursitis, biceps tenosynovitis or glenohumeral synovitis | 59% | 43%** | 65% | 26%** | 0%** |
| At least one shoulder with subdeltoid bursitis or biceps tenosynovitis | 82% | 63%** | 72% | 53%** | 19%** |
| Both shoulders with subdeltoid bursitis or biceps tenosynovitis | 57% | 35%** | 52% | 21%** | 0%** |
| At least one hip with synovitis or trochanteric bursitis | 38% | 23%* | 30% | 18%* | 0%** |
| Both hips with synovitis or trochanteric bursitis | 19% | 8%** | 9% | 4%* | 0%* |
| At least one shoulder and one hip with findings as above | 33% | 16%** | 17%* | 11%** | 0%** |
| Both shoulder and both hips with findings as above | 12% | 7% | 6% | 2%* | 0% |

*p<0.05 compared with PMR.

**p<0.01 compared with PMR.

PMR, polymyalgia rheumatica; RA, rheumatoid arthritis.

Table 6 PMR classification criteria scoring algorithm—required criteria: age 50 years or older, bilateral shoulder aching and abnormal CRP and/or ESR*

| | Points without US (0–6) | Points with US† (0–8) |
|--|-------------------------|-----------------------|
| Morning stiffness duration >45 min | 2 | 2 |
| Hip pain or limited range of motion | 1 | 1 |
| Absence of RF or ACPA | 2 | 2 |
| Absence of other joint involvement | 1 | 1 |
| 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 | Not applicable | 1 |
| Both shoulders with subdeltoid bursitis, biceps tenosynovitis or glenohumeral synovitis | Not applicable | 1 |

*A score of 4 or more is categorised as PMR in the algorithm without US and a score of 5 or more is categorised 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.

cohorts) who did not have shoulder conditions. Patients with PMR were more likely to have abnormal ultrasound findings in the shoulder (particularly subdeltoid bursitis and biceps tenosynovitis), and somewhat more likely to have abnormal findings in the hips than comparison subjects as a group (table 5). PMR could not be distinguished from RA on the basis of ultrasound, but could be distinguished from non-RA shoulder conditions and subjects without shoulder conditions.

Assessing the utility of ultrasound in classifying PMR

Factor analysis (see supplementary appendix table S5, available online only) revealed several strong factors potentially useful for classifying patients with PMR from the ultrasound data. The inclusion of ultrasound findings in the scoring algorithm resulted in improving the c-statistic to 82%. A score of 5 or greater had 66% sensitivity and 81% specificity for discriminating all comparison subjects from PMR. The specificity was higher (89%) for discriminating shoulder conditions from PMR and lower (70%) for discriminating RA from PMR. A total of 41 (34%) PMR patients and 30 (19%) comparison subjects was incorrectly classified. The positive predictive value was 72% and the negative predictive value was 75%. Therefore, ultrasound findings were useful in discriminating PMR from shoulder conditions,

but less so in discriminating PMR from RA. Table 4 also shows the scoring algorithm including the ultrasound criteria. Supplementary appendix table S4 (available online only) shows the sensitivity and specificity for all possible cut points of the scoring algorithm.

Response to corticosteroids in PMR

Complete corticosteroid response at 4 weeks was seen in 71% of patients and the response was sustained in 78% of responders at 26 weeks. As expected by the plan for tapering of the corticosteroids, the median prednisone dose decreased from 15 mg at baseline to 5 mg at 26 weeks. Response to treatment (percentage improvement in global pain VAS at weeks 4 and 26) was highly correlated with percentage improvement in other VAS measures (correlation >0.5 and p<0.001 at weeks 4 and 26), but was not correlated with percentage change in corticosteroid dose (p=0.20 at week 4 and p=0.47 at week 26). There was no association between the points obtained on either scoring algorithm and the response to corticosteroids at 4 weeks (Spearman correlation coefficient 0.09; p=0.38 for the scoring algorithm including ultrasound) and 26 weeks (data not shown), indicating that corticosteroid response can not be used as part of PMR classification.

We also re-evaluated our risk score model for the scoring algorithm using only the PMR subjects who responded to corticosteroids. When the final risk score model was re-computed using only the subset of subjects with PMR who responded to corticosteroids (and all of the comparison subjects), the odds ratios for each of the criteria remained essentially unchanged. In addition, the specificity was identical and the sensitivity increased by an insignificant amount (0.5%).

Blinded re-evaluation of selected PMR patients and non-PMR controls

The re-evaluation exercise showed that most candidate criteria items performed well in discriminating PMR patients from controls. However, a third of the sample of PMR patients/comparison subjects was difficult to classify. The high c-statistic levels associated with the corticosteroid response and post-treatment CRP and ESR suggested that the uncertainty originated from the pivotal role of corticosteroids in the investigator assessment, in deciding whether a patient does or does not have PMR. It raises questions such as whether PMR always responds adequately to corticosteroids and whether polymyalgic RF-positive disease without peripheral synovitis can occur (see supplementary appendix 2, available online only, for details).

DISCUSSION

This is the first international multicentre prospective study examining consensus-based candidate classification criteria for PMR proposed by an international work group. Findings indicate that patients 50 years of age or older presenting with new bilateral shoulder pain (not better explained by an alternative diagnosis) and elevated CRP and/or ESR can be classified as having PMR in the presence of morning stiffness for more than 45 min, and new hip involvement (pain, tenderness, limited movement). The absence of peripheral synovitis or of positive RA serology increases the likelihood of PMR. While recognising that RF particularly may be present in patients with PMR, the absence of RF serology is useful in distinguishing PMR from RA in older patients for classification purposes.^{18 34} Ultrasound findings of bilateral shoulder abnormalities (subacromial bursitis/bicipital tenosynovitis/glenohumeral effusion) or abnormalities in one shoulder and hip (hip effusion, trochanteric bursitis) may significantly improve the specificity of the clinical criteria. These criteria are not meant for diagnostic purposes and have not been tested as diagnostic criteria.

Newer concepts of PMR are revealed by this and other recent studies—heterogeneity at presentation and course, lack of uniform responsiveness to low-dose steroids and overlap with inflammatory arthritis. However, we feel that at present these classification criteria provide a basic framework for developing clinical trials of novel therapies in PMR.

How should the PMR classification criteria be applied?

The target population will be patients aged 50 years or older presenting with new-onset (<12 weeks) bilateral shoulder pain and abnormal acute phase response. The criteria may only be applied to those patients in whom the symptoms are not better explained by an alternative diagnosis. Mimicking conditions include the inflammatory and non-inflammatory conditions studied as comparators in this report.

Four clinical and laboratory criteria along with optional ultrasound criteria (table 4 and table 6) can be applied to eligible patients to identify patients with PMR suitable for low-dose corticosteroid therapy. The scoring scale is 0–6 (without ultrasound)

and 0–8 (with ultrasound). In the absence of competing diagnoses, a score of 4 or greater (without ultrasound), or 5 or greater (with ultrasound) is indicative of PMR. Patients with a score of less than 4 (based on clinical plus laboratory criteria) cannot be considered to have PMR. Ultrasound improves the specificity of PMR diagnosis, and shows particularly good performance in differentiating PMR from non-inflammatory conditions and thus is a recommended investigation for PMR.

Classification criteria for PMR should be useful for identifying patients appropriate for enrollment into clinical trials of novel medications for the treatment of PMR, and studying long-term outcomes in more homogeneous patient cohorts. Our analyses indicate that even typical PMR patients at presentation may vary in their response to low-dose corticosteroid therapy, indicating that corticosteroid response is not reliable as a classification feature for PMR. This is similar to other rheumatic diseases such as RA, in which phenotypically similar patients may exhibit different responses to disease-modifying antirheumatic drugs.

Strengths and limitations

The strengths of the study relate to harnessing international effort to address a disease area subject to wide variation of practice; to develop agreement on the definition of what may or may not be treated as PMR and what needs further evaluation.

Our study methodology satisfies the ACR guidelines for the development of classification criteria for rheumatic diseases.³⁵ The consensus-based candidate criteria were generated by a multispecialty international group whose views were supported by the results of a wide trans-Atlantic survey. The work group suggested a prospective study designed to separate PMR patients from comparison subjects in patients included on a single eligibility criterion (new bilateral shoulder pain in subjects ≥ 50 years).

This ensured an inception cohort longitudinal observational design wherein the PMR cohort could be compared with the comparison cohort at similar chronological time points of disease. All PMR patients were evaluated before treatment with corticosteroids, were treated with a standard corticosteroid schedule and assessed at predetermined time points. Our study is in keeping with the EULAR/ACR goal of developing rheumatic disease classification criteria as opposed to diagnostic criteria. We focused on subjects with new-onset/incident disease, and a 6-month longitudinal follow-up allowed an accurate evaluation of the disease course and diagnoses. Previous criteria for PMR were developed using cross-sectional comparisons. Only two of the previous criteria were developed through an evaluative process.^{20 36} Neither had a definition of the ‘gold standard’ diagnosis other than the physician considered the patient to have ‘unequivocal’ PMR.

Another strength of our proposed classification criteria is the imaging component. Musculoskeletal ultrasound has promise due to widespread availability, feasibility and good research evidence.¹⁸ The PMR work group standardised the examination of shoulders and hips by ultrasound for the purposes of the current study.³⁷ Our findings indicate that ultrasound evaluation of hips and shoulders adds significantly to the evaluation of the polymyalgic syndrome. The lack of a ‘gold standard’ and the challenge of circularity was addressed through a blinded multi-rater re-evaluation exercise in selected cases and comparison subjects (in supplementary appendix 2, available online only). Most of the candidate criteria items performed well. The misclassification of several subjects reflects the difficulty in discriminating PMR from other inflammatory conditions such as RA. This uncertainty originates from the pivotal role of (and circular

reasoning related to using) the corticosteroid response in deciding whether a patient does or does not have PMR. Although the proposed scoring algorithm has high specificity for identifying PMR patients, it is nevertheless unable to predict the subsequent corticosteroid response, suggesting heterogeneity of disease course and treatment response. This has also been reported previously.^{10,38}

While we were able to scrutinise the basis for PMR diagnosis, no formal criteria were required for diagnosis of the comparison conditions. This is a limitation of the study. However, the study was led in all centres by experienced rheumatologists with major clinical and research interest in PMR and related conditions. We did not include hip pain without shoulder pain as an eligibility criterion for reasons discussed in the 'Methods' section. Funding constraints limited the follow-up duration to only 6 months. We were also limited by lack of funding for the central measurement of laboratory data. Values of ESR, CRP, RF and ACPA were based on local laboratory assays. Our study approach reflects a pragmatic view, which perhaps lends wider applicability to the results of the study.

Our classification algorithm had a c-statistic of 81%, which exceeds the threshold of 80% that is conventionally considered to be useful in clinical decision-making. However, we suggest that the criteria are regarded as provisional at this point awaiting validation in a separate cohort.

In conclusion, patients aged 50 years or older presenting with bilateral shoulder pain and elevated CRP and/or ESR can be classified as having PMR in the presence of morning stiffness for more than 45 min, and new hip pain in the absence of peripheral synovitis or positive RA serology. Using ultrasound, a score of 5 or greater had 66% sensitivity and 81% specificity for discriminating all comparison subjects from PMR. In our view, this approach can now be used to test eligibility for trials with newer therapies in PMR. A number of future research questions are highlighted, including: (1) Should PMR be considered as a part of the spectrum of late-onset inflammatory arthritis? (2) Can polymyalgic disease without peripheral synovitis occur in RF-positive disease? (3) Can we diagnose PMR in patients with normal acute phase response? (4) What is the role of the early introduction of disease-modifying antirheumatic drugs in PMR?

We have collected and stored blood samples from both the case and the comparator groups in the study. We hope to develop research proposals using these biospecimens to test several candidate biomarkers in PMR. These proposals will also examine the acute phase response; whether stratification of the response and differential levels of acute phase reactants and cytokines may function as additional classification criteria items.

Contributors All authors contributed and were involved in the conduct of this study, in study design and/or subject recruitment, analysis and manuscript preparation.

Acknowledgements The investigators would like to acknowledge EULAR, ACR, the Mayo Foundation and the Biobanque de Picardie in Amiens, France, for material and intellectual support of this project. The investigators also wish to recognise the individual and uncompensated efforts of the staff at each participating institution, without which this study could not have been completed.

Funding Funding for this study was provided by EULAR, ACR, the Mayo Foundation, and was supported further by the individual and uncompensated efforts of participating investigators and their staffs and institutions, and the Biobanque de Picardie in Amiens, France. GE-F and MCC were supported by Ministerio de Ciencia y Tecnología (SAF 08/04328).

Competing interests None.

Ethics approval The study was approved by the ethics boards of the participating investigator institutions, which are listed.

Patient consent Obtained.

Provenance and peer review Not commissioned; externally peer reviewed.

Author affiliations ¹Department of Rheumatology, Southend University Hospital, Westcliff-on-Sea, Essex, UK

²Department of Internal Medicine, University of Genova, Genova, Italy

³Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA

⁴Department of Rheumatology, Immanuel Krankenhaus Berlin: Medical Center for Rheumatology Berlin–Buch Berlin, Berlin, Germany

⁵Department of Internal Medicine I, Innsbruck Medical University, Innsbruck, Austria

⁶Department of Rheumatology, Arcispedale S Maria Nuova, Reggio Emilia, Italy

⁷Department of Internal Medicine and Rheumatology, Military Institute of Medicine, Warsaw, Poland

⁸Department of Rheumatology, Medical University Graz, Graz, Austria

⁹Department of Internal Medicine, General Hospital of Kufstein, Kufstein, Austria

¹⁰Gentofte Hospital, Rheumatology Division, Hellerup, Denmark

¹¹Service de Médecine Interne, Amiens, France

¹²National Institute of Rheumatology and Physiotherapy, Budapest, Hungary

¹³Radiology Department, National Institute of Rheumatology and Physiotherapy, Budapest, Hungary

¹⁴General and Pediatric Rheumatology Department, National Institute of Rheumatology and Physiotherapy, Budapest, Hungary

¹⁵Rheumatology Unit, Dipartimento di Medicina Interna e Specialità Mediche, Sapienza Università di Roma, Rome, Italy

¹⁶Rheumatology Unit, Ospedale Misericordia e Dolce, Prato, Italy

¹⁷Center for Diagnosis Imaging, Hospital Clinic, Barcelona, Spain

¹⁸Department of Systemic Autoimmune Diseases, Hospital Clinic, University of Barcelona, Barcelona, Spain

¹⁹Servicio de Reumatología, Hospital Universitario Marqués de Valdecilla, Facultad de Medicina, Universidad de Cantabria, Santander, Spain

²⁰Sahlgren University Hospital, Department of Rheumatology, Göteborg, Sweden

²¹Department of Rheumatology, Marmara University Medical School, Istanbul, Turkey

²²Department of Rheumatology, Princess Alexandra Hospital, Harlow, UK

²³Department of Rheumatology, Addenbrookes Hospital, Cambridge, UK

²⁴Rheumatology and Rehabilitation Research Unit, University of Leeds, Leeds, UK

²⁵Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, Oxford University, Oxford, UK

²⁶Department of Internal Medicine, Division of Rheumatology, Mayo Clinic, Jacksonville, Florida, USA

²⁷Department of Internal Medicine, Division of Rheumatology, Mayo Clinic, Rochester, Minnesota, USA

²⁸Rheumatology Associates of North Jersey, Teaneck, New Jersey, USA

²⁹Department of Internal Medicine, Division of Rheumatology, Mayo Clinic, Scottsdale, Arizona, USA

REFERENCES

1. **Salvarani C**, Cantini F, Boiardi L, *et al*. Polymyalgia rheumatica and giant-cell arteritis. *N Engl J Med* 2002;**347**:261–71.
2. **Weyand CM**, Goronzy JJ. Giant-cell arteritis and polymyalgia rheumatica. *Ann Intern Med* 2003;**139**:505–15.
3. **Doran MF**, Crowson CS, O'Fallon WM, *et al*. Trends in the incidence of polymyalgia rheumatica over a 30 year period in Olmsted County, Minnesota, USA. *J Rheumatol* 2002;**29**:1694–7.
4. **Chakravarty K**, Elgabani SH, Scott DG, *et al*. A district audit on the management of polymyalgia rheumatica and giant cell arteritis. *Br J Rheumatol* 1994;**33**:152–6.
5. **Maradit Kremers H**, Reinalda MS, Crowson CS, *et al*. Use of physician services in a population-based cohort of patients with polymyalgia rheumatica over the course of their disease. *Arthritis Care Res* 2005;**53**:395–403.
6. **Gamez-Nava JI**, Gonzalez-Lopez L, Davis P, *et al*. Referral and diagnosis of common rheumatic diseases by primary care physicians. *Br J Rheumatol* 1998;**37**:1215–19.
7. **González-Gay MA**, García-Porrúa C, Vázquez-Caruncho M, *et al*. The spectrum of polymyalgia rheumatica in northwestern Spain: incidence and analysis of variables associated with relapse in a 10 year study. *J Rheumatol* 1999;**26**:1326–32.
8. **Cantini F**, Salvarani C, Olivieri I. Erythrocyte sedimentation rate and C-reactive protein in the diagnosis of polymyalgia rheumatica. *Ann Intern Med* 1998;**128**:873–4.
9. **Dasgupta B**, Borg FA, Hassan N, *et al*. BSR and BHPR guidelines for the management of polymyalgia rheumatica. *Rheumatology (Oxford)* 2010;**49**:186–90.
10. **Hutchings A**, Hollywood J, Lamping DL, *et al*. Clinical outcomes, quality of life, and diagnostic uncertainty in the first year of polymyalgia rheumatica. *Arthritis Rheum* 2007;**57**:803–9.
11. **Dasgupta B**, Matteson EL, Maradit-Kremers H. Management guidelines and outcome measures in polymyalgia rheumatica (PMR). *Clin Exp Rheumatol* 2007;**25**(6 Suppl 47):130–6.
12. **Matteson EL**. Clinical guidelines: unraveling the tautology of polymyalgia rheumatica. *Nat Rev Rheumatol* 2010;**6**:249–50.

13. **Caporali R**, Montecucco C, Epis O, *et al*. Presenting features of polymyalgia rheumatica (PMR) and rheumatoid arthritis with PMR-like onset: a prospective study. *Ann Rheum Dis* 2001;**60**:1021–4.
14. **Kremers HM**, Reinalda MS, Crowson CS, *et al*. Relapse in a population based cohort of patients with polymyalgia rheumatica. *J Rheumatol* 2005;**32**:65–73.
15. **Delecoeuillerie G**, Joly P, Cohen de Lara A, *et al*. Polymyalgia rheumatica and temporal arteritis: a retrospective analysis of prognostic features and different corticosteroid regimens (11 year survey of 210 patients). *Ann Rheum Dis* 1988;**47**:733–9.
16. **Jones JG**, Hazleman BL. Prognosis and management of polymyalgia rheumatica. *Ann Rheum Dis* 1981;**40**:1–5.
17. **Dasgupta B**, Hutchings A, Matteson EL. Polymyalgia rheumatica: the mess we are now in and what we need to do about it. *Arthritis Rheum* 2006;**55**:518–20.
18. **Dasgupta B**, Salvarani C, Schirmer M, *et al*. Developing classification criteria for polymyalgia rheumatica: comparison of views from an expert panel and wider survey. *J Rheumatol* 2008;**35**:270–7.
19. **Ferraccioli GF**, Di Poi E, Damato R. Steroid sparing therapeutic approaches to polymyalgia rheumatica-giant cell arteritis. State of the art and perspectives. *Clin Exp Rheumatol* 2000;**18**(4 Suppl 20):S58–60.
20. **Bird HA**, Esselinckx W, Dixon AS, *et al*. An evaluation of criteria for polymyalgia rheumatica. *Ann Rheum Dis* 1979;**38**:434–9.
21. **Leeb BF**, Bird HA, Neshor G, *et al*. EULAR response criteria for polymyalgia rheumatica: results of an initiative of the European Collaborating Polymyalgia Rheumatica Group (subcommittee of ESCISIT). *Ann Rheum Dis* 2003;**62**:1189–94.
22. **Kyle V**, Hazleman BL. The clinical and laboratory course of polymyalgia rheumatica/giant cell arteritis after the first two months of treatment. *Ann Rheum Dis* 1993;**52**:847–50.
23. **Dejaco C**, Duftner C, Cimmino MA, *et al*. Definition of remission and relapse in polymyalgia rheumatica: data from a literature search compared with a Delphi-based expert consensus. *Ann Rheum Dis* 2011;**70**:447–53.
24. **Dasgupta B**, Hutchings A, Hollywood J, *et al*. Clinical outcomes, quality of life and diagnostic uncertainty in the first year in polymyalgia rheumatica: a prospective cohort study. *Rheumatology* 2006;**45**(Suppl 1):i170.
25. **Weyand CM**, Fulbright JW, Evans JM, *et al*. Corticosteroid requirements in polymyalgia rheumatica. *Arch Intern Med* 1999;**159**:577–84.
26. **Cantini F**, Salvarani C, Olivieri I, *et al*. Inflamed shoulder structures in polymyalgia rheumatica with normal erythrocyte sedimentation rate. *Arthritis Rheum* 2001;**44**:1155–9.
27. **Cantini F**, Niccoli L, Nannini C, *et al*. Inflammatory changes of hip synovial structures in polymyalgia rheumatica. *Clin Exp Rheumatol* 2005;**23**:462–8.
28. **Backhaus M**, Burmester GR, Gerber T, *et al*. Guidelines for musculoskeletal ultrasound in rheumatology. *Ann Rheum Dis* 2001;**60**:641–9.
29. **Steyerberg EW**. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. Statistics for Biology and Health*. New York: Springer, 2010:57–71.
30. **Fabrigar LR**, Wegener DT, MacCallum RC, *et al*. Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods* 1999;**4**:272–99.
31. **Breiman L**. Random forests. *Machine Learning* 2001;**45**:5–32.
32. **Therneau TM**, Atkinson EJ. *An Introduction to Recursive Partitioning using the RPART Routines*. Department of Health Sciences Research Technical Report Series: no 66. Rochester, Minnesota: Mayo Clinic, 1997:1–52.
33. **Hastie T**, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd edn. New York: Springer Science + Business Media, 2009:337–88.
34. **Lopez-Hoyos M**, Ruiz de Alegria C, Blanco R, *et al*. Clinical utility of anti-CCP antibodies in the differential diagnosis of elderly-onset rheumatoid arthritis and polymyalgia rheumatica. *Rheumatology (Oxford)* 2004;**43**:655–7.
35. Classification and Response Criteria Subcommittee of the American College of Rheumatology Committee on Quality Measures. Development of classification and response criteria for rheumatic diseases. *Arthritis Rheum* 2006;**55**:348–52.
36. **Nobunaga M**, Yoshioka K, Yasuda M, *et al*. Clinical studies of polymyalgia rheumatica. A proposal of diagnostic criteria. *Jpn J Med* 1989;**28**:452–6.
37. **Scheel AK**, Matteson EL, Dasgupta B, *et al*. Reliability exercise for the polymyalgia rheumatica classification criteria study: the Oranjewoud Ultrasound Substudy. *Int J Rheumatol* 2009;**2009**:738931.
38. **Hernández-Rodríguez J**, Cid MC, López-Soto A, *et al*. Treatment of polymyalgia rheumatica: a systematic review. *Arch Intern Med* 2009;**169**:1839–50.

**PMR Classification Criteria Manuscript (Gupta, Matteson)
ACR Quality of Care Committee & Criteria Subcommittee Combined Reviews
April 2011**

QUALITY OF CARE COMMITTEE

Quality of Care Committee Reviewer (1):

Vote: Not certain that this should be approved with such low performance characteristics

Comments: The cardinal issue is whether these criteria have adequate performance characteristics to be considered "valid" criteria.

As they state in their abstract, the main criteria have only low-moderate sensitivity and specificity for PMR; it may be that no criteria can do better for PMR, but this seems to be an important issue.

"A score ≥ 4 had 69% sensitivity and 77% specificity for discriminating all comparison subjects from PMR."

Quality of Care Committee Reviewer (2):

Vote: Approve

Quality of Care Committee Reviewer (3):

Vote: Approve

Quality of Care Committee Reviewer (4):

Vote: Provisional Approval (full approval pending validation in an independent sample).

Quality of Care Committee Reviewer (5):

Vote: Approve

Comments: Very thorough work in a difficult condition. Authors are to be commended.

A few concerns:

Abstract conclusion is not entirely accurate...classification of PMR does not occur "in the absence of periph synovitis or positive RA serologies"--the algorithms simply make the classification of PMR less or more likely when considering individual criterion. This statement is repeated in the end of the discussion and should be correct there too.

If the purported problem that these criteria aim to address is "The lack of standardized classification criteria has been a major factor hampering development of rational therapeutic approaches in management of PMR^{12, 16, 17}. These deficiencies have contributed to difficulties in evaluating patients in clinical studies." then the goal should be to define a cohort that is as uniform as possible, with little variation or question that the subjects do, indeed, have PMR (even if it is only a restricted

**PMR Classification Criteria Manuscript (Gupta, Matteson)
ACR Quality of Care Committee & Criteria Subcommittee Combined Reviews
April 2011**

"sub-type" of PMR). Once trials demonstrate effective treatment/clearer understanding of the clinical characteristics for that homogenous group, further studies can be performed to assess whether clinical findings and treatment responses extend to other "sub-types" of PMR, or whether those conditions are distinct entities. But I think it is critical to avoid a heterogeneous population in this circumstance. Thus, sensitivity should be less emphasized compared to specificity, as the authors mention is their response to the Criterion Committee review--does it make sense to set the cut off higher for each algorithm and sacrifice sensitivity? Inclusion of the full ROC data is VERY helpful.

No formal analysis comparing ultrasound-based algorithm to the non-U/S algorithm was performed. While the specificity in the U/S algorithm is numerically higher, I am not certain whether the incremental benefit of the U/S algorithms is statistically significant. The following statement in the discussion (1st paragraph) is not correct: **Ultrasound findings of bilateral shoulder abnormalities...significantly improve both the sensitivity and specificity of the clinical criteria.**

The sensitivity of the ultrasound algorithm at the recommended cut off of 5 points (65.8) is numerically lower than the non-U/S algorithm at the recommended cut off of 4 points (68). More accurate to say that U/S-based algorithm with a 5 point cut off **may** improve specificity.

Recognizing that a single reference laboratory was not used, were there even vague rules about what constitutes an abnormal CRP/ESR value? Audience will be lost without some guidance. It might be useful to at least include a statement about what reference laboratories used as cut-offs--at least for CRP and ESR, whether age was considered while interpreting ESR, etc. Assuming all labs were certified, the absolute CRP concentration and ESR score should have been roughly comparable across labs. I'm simply trying to determine if most labs used CRP cut offs of 3mg/L, 5mg/L or 10mg/L. If no guidelines were given, then simply saying so would be helpful.

The C in Table 2 and the Table Appendix 2 (should this be Appendix 2 Table?), should be defined in the text under the table or labeled "C statistic", I think

I do think the authors have done about as good a job as they can. Kudos to all involved.

Quality of Care Committee Reviewer (6):

Vote: Approve

Comments: A fairly rigorous study and certainly an addition to the literature.

1. In table 3, the authors should insert values for C-statistic for the different models shown.

**PMR Classification Criteria Manuscript (Gupta, Matteson)
ACR Quality of Care Committee & Criteria Subcommittee Combined Reviews
April 2011**

2. In the methods, it would be very helpful to provide the clinical audience some interpretation of the c-statistic. There are published conventions suggesting that a C-stat ≥ 0.8 is probably of some value where values lower than this are probably only marginally helpful in clinical decision making.

Quality of Care Committee Reviewer (7):

Vote: Approval **based on minor revisions** (see comments below)

Comment: For ESR and/or CRP, the authors state that institution specific values were used to define an elevated level. It would be helpful to know the range of these thresholds and the methods used to measure ESR and CRP. Also, it seems like the authors characterized a person as positive when either ESR or CRP was elevated (if both were obtained). If true, perhaps they should say ESR or CRP rather than 'and/or'.

The paper also needs to be edited by a native English speaker.

Quality of Care Committee Reviewer (8):

Vote: Approve **as provisional criteria**

Comment: I vote for provisional acceptance since the criteria have not been validated in an external independent dataset.

I don't feel that the authors adequately addressed every comment, but overall I think the clarity of the manuscript was improved.

CRITERIA SUBCOMMITTEE

Criteria Subcommittee Reviewer (1):

Vote: Approval **based on minor revisions** (see comments)

Comment: For ESR and/or CRP, the authors state that institution specific values were used to define an elevated level. It would be helpful to know the range of these thresholds and the methods used to measure ESR and CRP. Also, it seems like the authors characterized a person as positive when either ESR or CRP was elevated (if both were obtained). If true, perhaps they should say ESR or CRP rather than 'and/or'.

The paper also needs to be edited by a native English speaker.

Criteria Subcommittee Reviewer (2):

Vote: Approve **as provisional criteria**

Comment: I think they have done a very good job of responding to our points and that these criteria are definitely better than anything that is out there thus far. I vote that we endorse them as the provisional ACR criteria for classification of PMR, as they will need independent validation.

Criteria Subcommittee Reviewer (3):

Vote: Approve **as provisional criteria**

**PMR Classification Criteria Manuscript (Gupta, Matteson)
ACR Quality of Care Committee & Criteria Subcommittee Combined Reviews
April 2011**

Comment: The group has answered the comments appropriately.

Criteria Subcommittee Reviewer (4):

Vote: Approve as provisional criteria

Comment: I was underwhelmed by their accuracy which may be due to lack of sufficient item generation or overzealous item reduction.

Nonetheless, this seems to be the best there is at the present time and there is a clear need to do the upcoming RCTs at least with some kind of standard diagnostic criteria.

Also, I do not see that there is a prospective validation (beyond the development phase) provided as far as I can see. Rather the prospective dataset is used to develop the scoring system. Although the development dataset size is acceptable, if a separate validation dataset was now considered a prerequisite for [full] endorsement, then [full] endorsement should not be provided.

Criteria Subcommittee Reviewer (5):

Vote: Approve

Criteria Subcommittee Reviewer (6):

Vote: Approve

Criteria Subcommittee Reviewer (7):

Vote: Approve as provisional criteria

Comment: For criteria that will be used to select subjects for clinical trials of potential therapies, one would prefer highly specific criteria that exclude patients who do not have the condition being studied. This may require accepting lower sensitivity, i.e., excluding more patients who will have PMR from the clinical trial, but will strengthen the trial design to require fewer subjects to demonstrate benefit for the subjects who actually have PMR. Perhaps the authors could include a small table showing the effects of requiring scores of 4, 5 and 6 (without ultrasound) or 5, 6, 7 and 8 (with ultrasound) on the specificity and sensitivity of the proposed criteria. Then investigators and statisticians could use this information when designing clinical trials, to balance recruitment issues (sensitivity) versus analysis issues (specificity).

Criteria Subcommittee Reviewer (9):

Vote: Disapproval

Comment: Overall I think the paper is remarkably improved and the data is elegantly presented. The authors did a great job of responding to the comments. In the end, I'm still very concerned about the selection of only patients with bilateral shoulder involvement. Moreover, the resulting criteria still have high misclassification rates. I understand that the aim is for specificity given the proposed use in clinical trials but the low sensitivity is concerning. Are these criteria really selecting for a particular subpopulation of PMR, albeit a large subpopulation? I realize there is a trade off but wonder if we could do better.

Criteria Subcommittee Reviewer (10):

**PMR Classification Criteria Manuscript (Gupta, Matteson)
ACR Quality of Care Committee & Criteria Subcommittee Combined Reviews
April 2011**

Vote: Approve as **provisional criteria**

Comment: It seems to me the authors addressed reviewers' comments in a reasonable manner. It also seems to me from reading the critiques that there did seem to be some concern from the group with the approach even though the authors defend it. Given that they are looking for endorsement, perhaps a provisional approval would be warranted. It seems like replication in another cohort would be appropriate.

Criteria Subcommittee Reviewer (11):

Vote: Approve as **provisional criteria**

APPENDIX 1: SUPPLEMENTARY TABLES

Table S1. Reference ranges for laboratory tests according to center

| Center | ESR (mm/hr) | CRP (mg/L) | RF (IU/ml) | ACPA (U/ml) |
|--------|---|---------------|---------------|-----------------------------------|
| 1 | < 25 | < 5 | <40 | < 20 |
| 2 | 2-37 in 1st hr | < 5 | <30 | 0-10 |
| 4 | <10 | < 5 | <40 | <20. |
| 6 | < 8 in 1st hr, <18 in 2nd hr | < 5 | < 14 | < 10 |
| 8 | 0-9 | < 5 | 0-13 | 0-6 |
| 8 | 0-9 | < 8 | 0-14 | 0-10 |
| 13 | FEMALE 2-20 MALE 2-15 | <10 | <16 | < 7 |
| 14 | Male: Yrs divided by 2 Female: yrs+10 divided by 2 | < 5 | 1-20 | <25l |
| 18 | 1-20 | <10 | < 25 | < 25 |
| 20 | < 15 | < 5 | <50 | <50 |
| 21 | < 28 | < 5 | <20 | <26 |
| 22 | < 30 | < 5 | <20 | <10 |
| 23 | <37 | <10 | 16 | < 8 |
| 24 | <37 | <10 | 16 | < 8 |
| 25 | 1-15 | <10 | <40 | <7 |
| 30 | <37 | <10 | 16 | < 8 |
| 31 | 0-29 | < 8 | <10 | <15.6 |
| 32 | Male: 0-22 Female: 0-29 | ≤8 | <10 | <15.6 |
| 33 | 0 - 29 | < 8 | <14 | <20 |
| 34 | <20 | <8 | <14 | <16 |
| 40 | <20 women < 14 men | < 5 | 1-20 | <25l |
| 41 | women – (age+10)/2 men – age/2 | ≤ 8 | ≤ 35 | < 7 – negative > 10 – positive |

Table S2. Diagnoses of the 169 Comparison Subjects

| | |
|---|----------|
| New-onset rheumatoid arthritis (RA) | 49 (29%) |
| New-onset other sero-negative arthritis (e.g. spondyloarthritis, psoriatic arthritis) | 20 (12%) |
| New-onset connective tissue diseases and vasculitis (e.g. ANCA-associated vasculitis, systemic lupus erythematosus, inflammatory myopathy) | 9 (5%) |
| Shoulder conditions (e.g. bilateral rotator cuff syndrome and/or adhesive capsulitis, rotator cuff tear, glenohumeral osteoarthritis) | 52 (31%) |
| Chronic pain | 26 (15%) |
| Endocrinopathy | 2 (1%) |
| Neurological disorder | 2 (1%) |
| Previously undiagnosed malignancy | 4 (2%) |
| Infection | 5 (3%) |

Table S3. Factor Analysis: Factor Loadings for Individual Criteria

| Criterion | Factor 1 | Factor 2 | Factor 3 | Factor 4 |
|---|--------------|--------------|--------------|--------------|
| Duration of symptoms \geq 2 weeks | | | | -0.207 |
| Shoulder tenderness | | | 0.162 | 0.578 |
| Pain or limited hip range of motion | 0.825 | | 0.142 | |
| Bilateral pelvic girdle (hip) aching | 0.778 | 0.173 | | |
| Hip tenderness | 0.641 | | 0.170 | 0.349 |
| Neck aching | | 0.127 | | 0.265 |
| Morning stiffness > 45 minutes duration | | 0.112 | 0.916 | |
| Recent weight loss of >2 kg | 0.104 | 0.243 | 0.221 | -0.134 |
| Carpal Tunnel | | 0.294 | 0.149 | |
| Peripheral synovitis | | 0.677 | | |
| Other joint pain | | 0.757 | | 0.132 |
| Abnormal RF and/or ACPA | -0.265 | | 0.141 | |
| MHAQ | 0.244 | 0.409 | 0.351 | |

Abbreviations: RF = rheumatoid factor; ACPA= anti-citrullinated protein antibody; MHAQ= modified health assessment questionnaire

*factor loadings >0.5 are shown in bold. Factor loadings between -0.1 and 0.1 were not included in the table.

Table S4. Receiver Operating Characteristic (ROC) analysis for both scores

| Clinical score w/o ultrasound | | | Score with ultrasound | | |
|-------------------------------|-------------|-------------|-----------------------|-------------|-------------|
| Score | Sensitivity | Specificity | Score | Sensitivity | Specificity |
| 0 | 92.8 | 41.4 | 0 | 92.5 | 42.2 |
| 1 | 92.8 | 43.8 | 1 | 92.5 | 42.9 |
| 2 | 92.8 | 46.8 | 2 | 92.5 | 44.8 |
| 3 | 87.2 | 61.5 | 3 | 91.7 | 51.3 |
| 4 | 68.0 | 77.5 | 4 | 82.5 | 70.1 |
| 5 | 47.2 | 88.8 | 5 | 65.8 | 80.5 |
| 6 | 19.2 | 98.2 | 6 | 36.7 | 92.2 |
| | | | 7 | 20.0 | 98.0 |
| | | | 8 | 8.3 | 100 |

Table S5. Factor Analysis: Factor Loadings for Ultrasound Criteria

| Criterion | Factor 1 | Factor 2 | Factor 3 | Factor 4 |
|---|--------------|--------------|--------------|--------------|
| at least 1 shoulder with subdeltoid bursitis, biceps tenosynovitis and/or glenohumeral synovitis(either posterior or axillary) | 0.121 | 0.236 | 0.959 | |
| Both shoulders with subdeltoid bursitis, biceps tenosynovitis and/or glenohumeral synovitis(either posterior or axillary) | | 0.803 | 0.351 | 0.141 |
| at least 1 shoulder with subdeltoid bursitis, biceps tenosynovitis and/or glenohumeral synovitis(either posterior or axillary) | 0.112 | 0.373 | 0.775 | |
| Both shoulders with subdeltoid bursitis, tenosynovitis and/or glenohumeral synovitis(either posterior or axillary) | | 0.960 | 0.245 | |
| at least 1 hip with synovitis and/or trochanteric bursitis | 0.906 | | | 0.286 |
| Both hips with synovitis and/or trochanteric bursitis | 0.492 | | 0.104 | 0.720 |
| At least 1 shoulder with subdeltoid bursitis and/or biceps tenosynovitis and/or glenohumeral synovitis(either posterior or axillary) AND at least 1 hip with synovitis and/or trochanteric bursitis | 0.933 | | 0.167 | 0.301 |
| US findings as detailed above in BOTH shoulder and BOTH hips | 0.292 | 0.209 | | 0.930 |

*factor loadings >0.5 are shown in bold

APPENDIX 2: Blinded re-evaluation of selected PMR patients and comparison

subjects

While the prospective study was ongoing, we performed a blinded re-evaluation of a subset of PMR and comparison subjects included in the prospective study. Clearly, the development of classification criteria is inherently exposed to the risk of circularity of reasoning, since the same rheumatologists who judge whether the patient has the disease or not also develop the list of candidate criteria items. Outcome prediction at follow-up can obviate, at least to some extent, the problem of circularity. Therefore, in order to ensure the quality of the data and create a standard for assessment of patients included in the study, we performed a blinded re-evaluation of a subset of patients. This evaluation was intended to ensure uniformity of decision to classify PMR, to constitute a reproducibility assessment of the newly developed criteria and to estimate the reliability of classification based on assessment at the first visit. Every investigator was sent a set of data on 30 patients (10 PMR, 20 comparison subjects). For each disease feature the investigator was asked (on a five point scale) to circle a weight to indicate how much this disease feature influenced their clinical judgment or their decision whether the patient has or does not have PMR (1=strongly influences diagnosis of PMR to 5=strongly influences the diagnosis was not PMR). Each reviewer was then asked to re-evaluate the diagnosis (PMR or not PMR) and indicate on a 5-point scale their degree of confidence in the diagnosis. The reviewer was then asked whether they would treat such a case with corticosteroid, and whether they would enter such a case in a clinical trial for PMR.

To assess the diagnostic accuracy of each candidate criteria, the mean rating across all raters was taken. This composite score was then used to determine the areas under the receiver operating characteristic (ROC) curve (AUC), denoted as the c-statistic. Subjects were categorised into 3 groups based on raters' misclassification rates. Group 1: greater than 50% misclassified; Group 2: 20–50% misclassified, Group 3: less than 20% misclassified.

The assessment of multi-rater discrimination of PMR (10 patients) from comparison subjects (20 subjects) was undertaken by 23 investigators. The results are shown in Table below. Misclassification proportion was high in 10 patients. In Group 1 (n=3, 1 PMR, 2 comparison subjects), the factors that contributed to the misclassification were normal CRP and/or ESR, poor or ill- sustained corticosteroid response, and RF positivity without peripheral synovitis. In Group 2 (n=7; 4 PMR, 3 comparison subjects), misclassification was related to persistent synovitis, lack of complete/sustained corticosteroid response, RF or CCP positivity and low baseline CRP and/or ESR.

The c-statistic suggested that gender, duration of symptoms, systemic symptoms such as weight loss, neck pain, limitation of movement and serum electrophoresis were unhelpful to

the blinded rater, in discriminating PMR from comparison subjects (c-statistic < 0.8 in all). Bilateral hip pain, morning stiffness, CRP and ESR levels (pre- and especially post-corticosteroid), and corticosteroid response were good discriminators of PMR from comparison subjects (c-statistic > 0.8 in all, Table).

Table Appendix 2: Blinded Multi-rater Evaluation of Diagnosis and Candidate Criteria

| Candidate Criteria | PMR subjects Mean (SD) | Comparison subjects Mean (SD) | P value | C statistic (95% CI) |
|--------------------------------|------------------------|-------------------------------|---------|----------------------|
| Bilateral pelvic girdle aching | 1.8 (1.1) | 3.3 (1.1) | 0.01 | 0.80 (0.54, 0.92) |
| Morning stiffness >45min | 1.7 (1.0) | 3.0 (1.4) | <0.01 | 0.87 (0.63, 0.96) |
| Abnormal CRP at baseline | 1.8 (0.9) | 3.2 (1.4) | <0.01 | 0.81 (0.58, 0.92) |
| Abnormal ESR at 26 weeks | 2.3 (1.0) | 3.2 (0.8) | <0.01 | 0.89 (0.64, 0.97) |
| Abnormal CRP at 26 weeks | 2.2 (0.9) | 3.1 (1.1) | <0.01 | 0.85 (0.62, 0.94) |
| Rapid steroid response | 2.6 (1.5) | 4.6 (0.6) | <0.01 | 0.99 (0.90, 1.00) |
| Complete steroid response | 2.2 (1.4) | 4.5 (0.7) | <0.01 | 0.98 (0.84, 1.00) |
| Sustained steroid response | 2.6 (1.4) | 4.3 (0.7) | <0.01 | 0.99 (0.90, 1.00) |

PMR = polymyalgia rheumatica; ESR = erythrocyte sedimentation rate; CRP = C-reactive protein; CI = confidence interval

Furthermore, the clinical scoring algorithm was evaluated in these 30 subjects. The algorithm correctly classified 8 of the 10 PMR subjects and 16 of the 20 comparison subjects. The 6 subjects who were incorrectly classified by our algorithm were also frequently misclassified by the expert raters. The median percentage of raters who incorrectly classified these patients was 33% compared to 4% for the other 24 subjects; $p=0.014$). This indicated our criteria perform well in distinguishing PMR from comparison subjects among patients that are agreed upon by experts. In addition, the subjects misclassified by our algorithm are more likely to be those which experts could not agree upon.

The re-evaluation exercise showed that most candidate criteria items performed well in discriminating PMR from comparison subjects. However, a significant proportion (a third of the sample) of PMR/comparison subjects was difficult to classify, as noted by the high percentage of raters who misclassified these subjects. The high c-statistic levels associated with the corticosteroid response and post-treatment CRP and ESR suggested that the uncertainty originated from the pivotal role of the corticosteroid in the investigator assessment, in deciding whether a patient does or does not have PMR. Questions such as whether PMR may not always adequately respond to corticosteroid and whether polymyalgic RF positive disease without peripheral synovitis can occur clearly require further investigation.

