A SPECIFIC AIMS

Polymyalgia rheumatica (PMR) is the most common inflammatory rheumatic disease of the elderly and represents the most common indication for long-term glucocorticoid therapy in the community. PMR is the rheumatic disease subject to the widest variations of clinical practice since it is managed both in primary and secondary care setting by general practitioners, rheumatologists and non-rheumatologists. There is considerable uncertainty related to diagnosis and outcomes in patients presenting with the polymyalgic syndrome. Furthermore, the course of the disease is heterogeneous and unpredictable. Glucocorticoid treatment leads to a rapid and dramatic improvement in disease symptoms in the majority of patients but treatment typically lasts for several years and results in serious adverse effects, of which osteoporosis, fractures, diabetes and infections are among the worst. As a result, treatment of PMR remains largely empiric and unsatisfactory. There is an unmet clinical need for the treatment of PMR due to the long-term adverse effects of prolonged use of glucocorticoids.

The foremost factor that has hampered development of rational therapeutic approaches to management of PMR is the lack of standardized diagnostic criteria and reliable, valid and sensitive outcome measures. These deficiencies have led to an inability to accurately distinguish this clinical disease entity from other conditions presenting with the polymyalgic syndrome, and difficulty with evaluation and comparison of the efficacy of different therapeutic approaches, including novel drug therapies. In order to tackle these challenges, an international PMR Consortium has convened and initiated a multinational effort. The long-term goal of the Consortium is to standardize classification criteria and outcome measures in patients with PMR and facilitate the conduct of well-designed clinical studies that incorporate these standardized and clinically relevant outcomes. The objective of this particular application is to develop and quantitatively evaluate the reliability, validity and sensitivity to change of various measures for disease activity, response and remission in patients with PMR in a multinational prospective observational study. To this end, our specific aims are as follows:

Specific Aim 1. To develop a reliable and valid composite disease activity score and remission criteria in PMR

Aim 1a. To evaluate the validity, redundancy and reproducibility of core disease activity domains based on patient-reported (pain, stiffness, function, global assessment, fatigue, QOL), physician-reported (global assessment, tenderness, pain on motion and limitation, therapy), laboratory (ESR, CRP) and ultrasound-based measures

Aim 1b. To derive a preliminary composite disease activity score and remission criteria incorporating these core domains

Aim 1c. To assess the discriminant validity of core domains and composite score in differentiating PMR from conditions mimicking PMR

Specific Aim 2. To develop a preliminary definition of improvement for the evaluation of response to therapy in PMR

Aim 2a. To evaluate the within group sensitivity to change (responsiveness) and discriminant validity of core disease activity domains as measured by patient-based, physician-based, laboratory and ultrasound-based measures

Aim 2b. To derive a preliminary composite definition of improvement based on disease activity (Aim 1b) and change in core disease activity domains

Specific Aim 3. To create an infrastructure for storage of biospecimens in order to facilitate future research on novel disease-specific biomarkers and their therapeutic potential

This will be the first comprehensive effort to develop classification and outcome criteria in PMR. The findings of this project will be used to develop preliminary evidence-based, consensus guidelines for
evaluating outcomes in PMR and form the basis for international guidelines. **The conduct of the proposed project will be optimized by several key elements.** First, this project will be implemented as a multinational effort. A PMR Consortium consisting of highly committed rheumatologists and non-rheumatologists with synergistic experience on a number of fronts (epidemiology, rheumatology, general medicine, health services research, biostatistics) is actively working together to develop the methodology for the prospective validation study. Recent and ongoing work by the PMR Consortium provides important insights into the objectives and the methodologies proposed herein. This international, collaborative and broadly based effort has a high likelihood of resulting in wider acceptability and utilization of the newly developed criteria. Given the uncertainties associated with PMR diagnosis and fast evolution of clinical findings and symptoms, it is of paramount importance to develop classification and outcome measures simultaneously and this is ensured in this proposal. The **infrastructure for the proposed study has already been established with 22 centers in 10 countries** having participated in preliminary studies and agreeing to participate in the current effort. In this context, we demonstrated the feasibility of the major methodological components central to this proposal through our ongoing collaborative work. This effort is also endorsed by the American College of Rheumatology (ACR) and the European League Against Rheumatism (EULAR).

**B BACKGROUND AND SIGNIFICANCE**

**B.1 Problems with PMR diagnosis and management**

PMR is an inflammatory condition of **unknown etiology** characterized by aching and morning stiffness in the proximal regions of the extremities and torso, and elevated markers of inflammation including the erythrocyte sedimentation rate (ESR). Few population-based studies have evaluated the epidemiological aspects of PMR because of the lack of a diagnostic hallmark and universally accepted classification criteria. PMR is extremely rare in persons under 50 years of age. The prevalence among those over 50 is estimated to be around 0.74% \(^1\). The overall age- and sex-adjusted annual incidence per 100,000 population aged ≥50 years in Olmsted County is estimated at 58.7 (95% confidence interval: 52.8, 64.7) \(^2\). The incidence of PMR increases with age, with a peak in those 70–80 years of age. PMR is more frequent in females than in males in all age groups. PMR is probably a polygenic disease in which multiple environmental and genetic factors influence susceptibility and severity. The increased incidence at higher latitudes (i.e. in Scandinavian countries and in USA communities with a strong Scandinavian ethnic background) and occasional familial cases support the etiological role of both environmental and genetic causes \(^3\).

In the past, PMR was considered a manifestation of giant-cell arteritis (GCA) or a variant of elderly-onset rheumatoid arthritis (EORA). This fact can in part explain the lack of efforts to standardize its diagnosis. Therefore, the criteria most frequently used for diagnostic purposes are empirical. They have been defined by clinical experts who had studied the disease extensively. The criteria defined by Hunder at the Mayo Clinic \(^4\) and by Healey at Mason Clinic \(^5\) are similar, the only difference being the inclusion in the Healey criteria of the response to steroids. So far, the only effort to develop diagnostic criteria for PMR was a collaborative study performed by 11 British rheumatology units (Bird criteria) \(^6\). There are several similarities and dissimilarities between these criteria, for example, different cut-offs for age and ESR values and the need to exclude other conditions.

There are 3 diagnostic challenges in PMR. First, PMR is a **heterogeneous** disease and many features of PMR predispose the unwary clinician to diagnostic error \(^7\). The proximal pain and stiffness syndrome - the main symptoms of PMR - can occur in many other illnesses. A third of the patients have systemic symptoms such as fever, anorexia and weight loss. Half the patients may have distal musculoskeletal manifestations such as peripheral arthritis, distal swelling with pitting edema and
carpal tunnel syndrome. PMR is also associated with GCA in 10% to 20% of cases. Although ESR of at least 40 mm/hour is considered an important finding for the diagnosis of PMR, an acute phase response and elevated ESR values can occur in other settings such as neoplasia, rheumatoid arthritis and infection. Furthermore, some studies have reported that a normal ESR at diagnosis accounted for 7–20% of the patients with PMR\(^8\)–\(^{10}\). PMR and GCA are regarded as urgent indications for the start of steroid therapy. Many clinicians and 2 of the published diagnostic criteria use a response to steroids as a main defining feature of these conditions\(^{11, 12}\). This may encourage diagnostic error since steroids are potent anti-inflammatory agents that can mask symptoms from a host of serious conditions ranging from osteoarthritis, rotator cuff problems, RA, cancer, infection, migraine and intra-cranial tumor, especially if used in traditional high doses and for protracted lengths of time.

Second, there is uncertainty about the origin of symptoms that determines duration of treatment. Reports from the Mayo Clinic in the USA originally described a short duration of disease both for PMR (11 months median duration of treatment with 75% of patients off steroids at 2 years) and giant cell arteritis\(^4\). However, reports from Europe describe a more prolonged illness requiring several years of steroid therapy marked by frequent relapses\(^{13-16}\). The fact that the majority of patients with PMR remain on long-term steroids suggests persistent symptoms. Kyle and Hazleman found that only 24% of patients were off steroids at 2 years\(^{13}\). A recent cohort study reported a high incidence of relapses, this was largely irrespective of original steroid responsiveness and up to a third complained of proximal pain at 12 months in the absence of raised inflammatory indices. Results of our Delphi survey of generalists corroborate these observations\(^{17}\).

A third difficulty is the wide variations in clinical practice. In comparison to other rheumatic diseases, PMR is managed both in the primary and secondary care setting by general practitioners, rheumatologists and non-rheumatologists\(^{18-20}\). Rheumatologist referral is mostly for management of steroid complications\(^{21}\). There is a lack of recognition of PMR, particularly in primary care. Bahlas et al reported low accuracy of PMR diagnosis and an over-utilization of diagnostic investigations by family practitioners\(^{18}\). Among the various rheumatologic diseases, the lowest agreement between primary care physicians and rheumatologists is on the subject of diagnosis of PMR\(^{22}\). In another study in the north-eastern USA, 128 physicians were presented with videos of patients presenting with symptoms of PMR and depression\(^{23}\). The study was designed to assess whether patient or physician attributes influenced clinical decision-making. Only 9 of the 128 physicians identified PMR as the most likely diagnosis whereas 84 considered depression most likely\(^{23}\).

**In conclusion, PMR is a heterogeneous disease. There is considerable uncertainty and variations related to diagnosis and outcomes in patients presenting with the polymyalgic syndrome. The heterogeneity of the disease along with wide variations in practice in the management of PMR hinders easy diagnosis and recognition of changes in disease activity status. These difficulties present a challenge when evaluating patients in the clinic and when conducting clinical studies. There is an urgent need for guidelines for safe diagnosis, ongoing monitoring of the disease, vigilance regarding an alternative diagnosis and early referral of patients with atypical features and poor steroid response. Our proposed study will provide information on several of these issues.**

### B.2 Unmet clinical need in PMR

The mainstay of therapy in PMR is oral steroids\(^{24}\). Yet, the true outcome of standard steroid therapy, which may only partially control disease and symptoms, is not known. Prednisone, 15 to 25 mg, usually suppresses inflammation dramatically\(^{25}\). However, up to 60% of patients experience disease relapse during steroid tapering and several studies indicate that steroid treatment can only rarely be discontinued before 2 years\(^{12, 24, 26, 27}\). A recent study identified complete, partial and non-responders...
to initial standard steroid therapy with partial and non-responders having earlier relapses leading to greater cumulative dose. Observational studies indicate that higher doses of steroids are not more effective in reducing the incidence of relapses but this has never been tested in clinical trials.

In a large series of patients taking steroids, the median reported starting dose of prednisone was 20 mg/d. Other studies reported similar figures. Recent studies indicate that oral steroids should be administered for 18 to 36 months. The obvious risk for side effects of this long-term treatment, particularly osteoporosis, hypertension, hyperglycemia, and cataracts, support the need for improved therapeutic options, particularly in patients at high risk for steroid-related toxicity.

Randomized controlled clinical trials of different therapeutic agents are rare in PMR. In one small clinical trial, intramuscular methylprednisolone acetate on a monthly basis was shown to have benefit in reducing cumulative steroid dose and toxicity while maintaining control of disease. A number of studies examined the efficacy of methotrexate with conflicting findings. There is some evidence to suggest that methotrexate use can reduce the duration of steroid therapy and disease flares. Other than these small studies, there are only 1-2 trials with newer biologics. Infliximab was recently found to be ineffective in a small clinical trial.

In conclusion, it is imperative to evaluate the efficacy of disease modifying anti-rheumatic drugs in PMR. Unlike the extensive experience in rheumatoid arthritis, previous therapeutic studies in PMR are based on small patient numbers, have examined only short-term effects and have not evaluated patient-based outcomes such as functional ability and quality of life. Classification criteria for PMR will ensure recruitment of individuals with similar clinical entity, and response criteria will ensure standardized definitions of therapeutic response and outcomes and qualitative comparisons across trials using the same metrics. In addition, this project will provide an infrastructure for storage of biospecimens. Combined with the phenotypic information collected as part of the proposed study, the biospecimen infrastructure will facilitate future research on novel disease-specific biomarkers (e.g. IL-1, IL-6, proteomic and genomics applications) and their therapeutic potential.

B.3 Need for outcomes and classification criteria for PMR

There are few standard definitions for diagnosis, clinical response and remission in PMR. Several different outcome measures have been used in the past, and there is no consensus about the optimal endpoints for evaluating treatment (see Table 1). Moreover, none of these remission definitions have been validated, especially with patient-based outcome measures. These remission definitions are arbitrary in a number of fronts, in particular importance of various criteria components at different stages of the disease, time/duration of the disease activity and remission states, correlation with prognostic markers and failure to account for the continuity of disease activity process. For example, validity and sensitivity to change of different measures can be different early in the disease course versus later (e.g. pain resolves early but laboratory values remain high for longer). There are other unanswered questions, such as, how many patients with proximal pain and stiffness respond to steroids even though they are not thought to have PMR? Can a patient have PMR even if he/she doesn't respond to steroids at a specified dose?
There is also no consensus on the role of imaging for the evaluation of suspected PMR. Magnetic resonance imaging (MRI) in PMR patients with active disease showed that bilateral subacromial and subdeltoid bursitis was present in 100% of PMR patients. Gleno-humeral synovitis and long head biceps tenosynovitis were detected in 77% and 54% of the cases. Ultrasound evaluations revealed evidence of subacromial/subdeltoid bursitis in almost all patients with PMR and in only 22% of controls. The lesions were bilateral in 96% of the PMR patients and in only 4% of controls. A recent systematic review of ultrasound in GCA clearly finds it a useful investigation.

Needless to say, there is an urgent need to develop consensus criteria for PMR. The one advantage of the investigators of this proposal is the body of experience with criteria development in other rheumatic diseases. International, gold standard guidelines for identifying a core set of outcomes to be measured to define improvement in rheumatic diseases, OMERACT, have now been established in adult RA, juvenile arthritis, ankylosing spondylitis and osteoarthritis and osteoporosis. The

Table 1. Remission criteria applied in the literature

<table>
<thead>
<tr>
<th>History criteria</th>
<th>Clinical criteria</th>
<th>Laboratory criteria</th>
<th>Therapy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Martinez-Taboada</td>
<td>Absence of clinical symptoms</td>
<td>Normal laboratory values</td>
<td></td>
</tr>
<tr>
<td>Meliconi</td>
<td>Absence of clinical symptoms</td>
<td>&lt; 40</td>
<td>&lt; 0.5</td>
</tr>
<tr>
<td>Cantini</td>
<td>Absence of clinical symptoms</td>
<td>&lt; 40</td>
<td>&lt; 0.5</td>
</tr>
<tr>
<td>Dasgupta</td>
<td>Reduction of Pain (VAS ≥ 50%) Morning stiffness &lt;30 min</td>
<td>&lt; 20</td>
<td>Hb&gt;12 g/dl</td>
</tr>
<tr>
<td>Feinberg</td>
<td>Absence of pain</td>
<td>Normal values</td>
<td></td>
</tr>
<tr>
<td>Ferraccioli</td>
<td>Absence of myalgia</td>
<td>Normal values</td>
<td>Normal values</td>
</tr>
<tr>
<td>Mertens</td>
<td>Absence of clinical symptoms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Krogsgaard</td>
<td>Absence of muscular pain morning stiffness</td>
<td></td>
<td>Normal values</td>
</tr>
<tr>
<td>Wagner</td>
<td>morning stiffness for just some minutes</td>
<td>Absence of clinical symptoms</td>
<td>&lt; 20</td>
</tr>
<tr>
<td>Delecoeuillerie</td>
<td>Absence of clinical symptoms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ayoub</td>
<td>Absence of clinical symptoms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behn</td>
<td>Absence of clinical symptoms</td>
<td>&lt; 30</td>
<td></td>
</tr>
<tr>
<td>Chuang</td>
<td>Absence of clinical symptoms</td>
<td>Normal laboratory values</td>
<td>Discontinuation of steroids or NSAID other than Aspirin: If treated with Aspirin: dose less than 1.8g/ day</td>
</tr>
</tbody>
</table>

ESR, erythrocyte sedimentation rate; CRP, C-reactive protein
OMERACT guidelines were developed through a series of evidence-based, consensus conferences, the aim of which was to identify the most scientifically rigorous measures to evaluate clinical and patient-based outcomes in clinical trials. There is a clear need for similar guidelines to define a core set of outcomes for use in future research on PMR. The first step in this process is to obtain validated classification criteria, rigorously measured outcome data (clinical and patient-based), a standard validated definition of disease activity and remission and improvement in a large cohort of patients in a prospective study of outcomes in PMR. The goal of this proposal is that each outcome measure in PMR meets the standards of the "OMERACT Filter" of truth (face, content, construct, and criterion validity), discrimination (reliability and sensitivity to change), and feasibility. We will adopt a data-driven process to achieve this goal.

B.4 Significance
This project is significant in a number of aspects. First, this will be the first comprehensive assessment of outcome measures in patients with PMR. The results emanating from this study will provide reliable and quantifiable criteria for early distinction of PMR from other similar conditions, initiation of timely therapeutic interventions and consideration of clinically relevant outcomes that matter to the patients and also define the course of the disease. Second, it brings together the expertise and commitment of an international group of investigators. Third, validated and standardized classification and outcome criteria will open the way to well-designed clinical studies and new therapeutic targets with the aim of improving patient care in PMR. Fourth, the proposed biospecimen storage will facilitate identification of therapeutic targets. Lastly, the benefits are not just limited to the clinical research setting. These standardized criteria will be used to monitor health status and response to treatment in the clinical setting, case management and adjudication in litigation environments, for benchmarking activities, systematic reviews, cost-effectiveness analyses, clinical decision making and estimation of the burden of PMR to patients and the society.

C PRELIMINARY STUDIES
This project will specifically build on a strong infrastructure of international network of rheumatologists with interest in PMR and who are committed to development of standardized PMR outcomes criteria. Specifically, the PMR work group was convened in 2005 and already demonstrated their commitment to this effort and efficiency in working together with the series of activities over the last year. These are briefly summarized below and more extensively in a manuscript.

Literature review: An extensive literature review was performed by the PMR work group in collaboration with the British Society for Rheumatology. Briefly, this literature review indicated that there was a dearth of clinical studies of PMR conducted within the primary care and rheumatology setting and very few randomized controlled treatment trials. Methodological problems encountered in the published literature included: small sample sizes that were not powerful enough to measure differences in clinically important outcomes with high precision, diversity of measurement instruments.
that were mostly unvalidated, problems with internal validity, absence of strict randomization and blind assessment of observer-rated outcomes, inadequate information on content and quality of interventions and protocol adherence, lack of patient data prior to steroid interventions and use of a variety of diagnostic entry criteria with blurred distinction between isolated PMR and GCA with polymyalgia. There were only a few observational studies of long-term outcomes.

**A Delphi survey of imaging including ultrasound in PMR** was conducted in 2005 and indicated that physicians did not believe that currently, there was a role for routine use of MRI or FDGPET scans in the investigation of PMR. However musculoskeletal ultrasound had promise due to widespread availability (65% of experts considered musculoskeletal ultrasound to be routinely available), feasibility and good research evidence, and the utility of ultrasound in PMR diagnosis. The PMR Consortium imaging work group met in the Netherlands in June 2006 to standardize the examination of shoulders and hips by ultrasound for the purposes of the current study.

**Development of consensus-based initial candidate criteria by the PMR Consortium** The PMR work group first convened in 2005 and discussed several issues on the nature of the diagnostic and evaluation process in PMR. The results of the work group’s deliberations were presented to the attendees of the 3rd International Conference on GCA and PMR held at St. Johns College, Cambridge and feedback was received. At the end of the conference, the work group members agreed that the diagnosis of PMR was a stepped process using core and supportive inclusion/exclusion criteria and a standardized initial steroid response. Over 50 items were rated. Of these, 43 receiving at least 50% agreement, including 10 core items that were agreed upon by expert consensus. These 43 items were evaluated in a subsequent extended Delphi survey.

**Testing for wider acceptability of consensus based initial candidate criteria in an extended Delphi survey**. A wider Delphi survey was used to determine if there is widespread acceptance of the views of the PMR work group and to receive input from other rheumatologists, primary care physicians and non-rheumatologists. Responses were received from 111 practicing rheumatologists from a mix of academic and private practice settings (49 US, 62 from 15 countries in Northern/Western Europe and Canada) and 53 generalists/internists (29 US, 24 UK). Over 70% of Delphi survey respondents agreed with the experts on 7 of 10 core candidate criteria. These were age >50 years, duration >2 weeks, bilateral shoulder and/or pelvic girdle aching, morning stiffness duration of >45 minutes, erythrocyte sedimentation rate (ESR), C-reactive protein (CRP) and rapid steroid response (> 75% global response within one week to prednisolone/prednisone 15-20 mg). Little agreement was observed on the remaining 3 core items: abrupt onset in <1 week (25%), systemic signs/symptoms (38%) and neck aching (35%). Among physical signs, >70% of respondents agreed with the experts on the importance of assessing shoulder (84%) and/or hip (76%) pain on motion, but agreement was low for peripheral signs like carpal tunnel, tenosynovitis and peripheral arthritis. Seventy five % or more of both groups agreed that a diagnosis of rheumatoid arthritis, lupus, vasculitis, inflammatory myopathy, septic arthritis, active neoplasia, active thyroid disease or drug-related myalgia would exclude the PMR diagnosis in a classification criteria study of the polymyalgic syndrome. Therefore, this Delphi survey clearly demonstrated that there is agreement between experts and practicing rheumatologists and generalists/internists on the importance of the majority of potential PMR criteria examined.

In conclusion, these preliminary studies demonstrate the previous research experience, competence and commitment of the research team who have already laid the framework for the proposed research to develop outcome measures in PMR. A comprehensive and sensible list of potential criteria had been developed through an extensive literature review, consensus meetings and 3 separate Delphi surveys. The selected criteria to be evaluated as part of this
Proposal are the most relevant, crucial clinical characteristics of PMR and likely to facilitate classification and comprehensive assessment of patients. Therefore, the initial face and content validity of the selected list of outcome criteria have already been evaluated according to the OMERACT filter and more recent recommendations by the Classification and Response Criteria Subcommittee of the ACR Quality Measures Committee. The next steps are to establish the reproducibility, redundancy and validity of these criteria for defining disease activity, remission and response in a prospective observational study. Still, we do recognize that criteria developed as part of this effort will be considered preliminary until tested in the context of a clinical trial.

D RESEARCH DESIGN AND METHODS

The core component of this project is a prospective observational study. General research methods including descriptions of the study design, recruitment process and study measurements are outlined in section D.1. Measurements and our rationale for inclusion of each of the measurements are described in section D.2. The statistical methods and power considerations are then described, according to each specific aim, in sections D.3 - D.5.

D.1 Overview of study design

The specific aims of this proposal will be fulfilled using a prospective observational study. Potential PMR cases (total 120 subjects) will be identified among consecutive patients referred to rheumatology clinics with suspected polymyalgia, i.e., new-onset proximal pain and morning stiffness and followed up over a 6-month period (Figure 1). Patients will be recruited prior to initiation of steroid treatment (i.e. patients should not have been treated with steroids within the last 12 weeks prior to study entry). Steroid treatment will be initiated according to a pre-defined protocol. Various clinical, laboratory and patient-based outcome measurements will be collected using standard questionnaires translated to different languages. Evaluations will be performed at baseline, 1 week, 4 weeks, 12 weeks, 26 weeks (6 months). Control patients will be evaluated at baseline and at 26 weeks. A main purpose of these follow-up evaluations until 26 weeks is to determine whether the initial PMR diagnosis is sustained.

Every effort will be made to choose PMR cases across the whole spectrum of PMR disease severity. Polymyalgia with clinical suspicion of GCA will be studied as part of the inflammatory control group since these patients require different steroid doses.

At each follow-up visit, clinical evaluation will also include response
to steroid therapy (see also measurements section) and opinion on emergence of alternative
diagnoses. Patients not considered as having PMR at any time during the study would be evaluated
and treated according to accepted clinical practice and they will be excluded from the pool of PMR
cases and will continue to be evaluated as non-PMR controls. It is expected that approximately 15% of
potential PMR subjects (and potentially up to 17%) will be reclassified as controls, and
approximately 10% of controls are expected to qualify as PMR cases by the end of the 26 weeks of
follow-up.

**Selection of controls** (total 240 subjects) is important for the development of classification criteria
because it influences both the sensitivity and specificity. In this study, we propose to include control
patients from a variety of “suspected” conditions, in order to evaluate the discriminating validity of the
PMR criteria in different patient populations. The control patients will have conditions likely to be
representative the types of patients that need to be distinguished from PMR patients (Figure 2). This is
essential in order to produce criteria that are useful, both for clinical studies and in clinical
practice. We plan to include consecutive elderly (>50 years) patients with new-onset proximal
musculoskeletal pain and with a suspected diagnosis of either inflammatory (120 control subjects)
or non-inflammatory (120 control subjects) conditions. The controls will be included in the study on the
basis of clinician diagnosis. Formal criteria for control conditions will not be developed as part of this study.

---

**Inclusion-exclusion criteria.** All patients who present with new onset (≤ 3 months) proximal bilateral
shoulder/thigh pain will be eligible for inclusion into the study. Patients with active cancer or those
diagnosed with selected conditions prior to the enrollment date will be excluded (see Table 2). A
diagnosis other than PMR should be suspected in all controls. No guidelines will be provided for
treatment of conditions in the control group.

**Inflammatory conditions (should be previously undiagnosed or condition present for ≤3 months)**

- New onset rheumatoid arthritis (60 subjects, 3 per center)
- New onset other sero-negative arthritides including spondyloarthritis, psoriatic arthritis etc. (30
  subjects, 1-2 per center)
- New onset connective tissue diseases & vasculitides, e.g. ANCA positive vasculitis, GCA with
  polymyalgia, SLE, inflammatory myopathies (30 subjects, 1-2 per center)

**Non-inflammatory conditions (should be previously undiagnosed or condition present for ≤3 months)**

---

* Others include endocrinopathies (thyroid disorders) and neurological disorder (parkinsons disease), malignancies
  (myeloma, solid cancers esp. metastatic disease presenting as musculoskeletal symptoms, infections
  (ideally systemic) e.g. chest infections, UTIs, possibly bacterial endocarditis, septic arthritis
PMR Classification Study Protocol

- Shoulder conditions: bilateral rotator cuff syndrome and/or adhesive capsulitis, rotator cuff tear, OA (60 subjects, 3 per center)
- Chronic pain e.g. fibromyalgia, back and hip pain (30 subjects, 1-2 per center)
- Either one of the below conditions (30 subjects, 1-2 per center)
  - Endocrinopathies, neurological or drug-related disorders, e.g. Thyroid disorders, Parkinson’s disease, statin-associated myopathy
  - Malignancies e.g. myeloma, solid cancers esp. metastatic disease presenting as musculoskeletal ‘red flags’
  - Infections (ideally systemic) e.g. chest infections, urinary tract infections, possibly bacterial endocarditis, septic arthritis

Recruitment: Patients will be recruited initially from 20 rheumatology referral clinics in eight countries: USA, UK, Italy, Spain, Sweden, France, Austria, Germany; up to 22 centers in 10 countries may ultimately be used (including Hungary, Israel). Each center will be recruiting 6 PMR cases and 12 non-PMR controls. Recruitment is expected to last 12 months with an anticipated recruitment rate of two patients per center per month. Full written consent will be obtained from all patients. Participating general practices which have agreed to refer the suspected PMR cases and non-PMR control subjects will send letters of support agreeing to refer untreated patients to the study centers. Recruitment will be monitored closely in the first few months to determine whether additional practices need to be identified in order to maintain the target recruitment level.

Steroid treatment: All eligible PMR patients in the study will receive a standard course of steroid therapy with oral prednisolone taken according to the schedule shown in Table 3. Tapering should be performed according this schedule. All patients will receive an initial dose of 15 mg of prednisone equivalent per day. Subsequent treatment will be individualized on the basis of treatment response, comorbidities, and other relevant clinical considerations. In accordance with recognized treatment methods, deviations from the treatment schedule can be made at the discretion of the clinician to allow steroid doses to be tailored to individual patient requirements. All steroid treatment will be recorded.

Re-evaluation of PMR diagnosis and response: The development of classification and outcomes 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 the diagnosis and outcome assessment of PMR, we will perform a diagnostic re-evaluation at the end of the study. Each investigator will be randomly assigned to a set of data from PMR cases and non-PMR controls enrolled in the study by different investigators. Each reviewer will be asked to re-evaluate the diagnosis on a 5-point scale. Note that only patients who were originally diagnosed and

<table>
<thead>
<tr>
<th>Table 3. Steroid schedule</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time</strong></td>
</tr>
<tr>
<td>Week 1-2</td>
</tr>
<tr>
<td>Week 3-5</td>
</tr>
<tr>
<td>Week 6-11</td>
</tr>
<tr>
<td>Week 12-15</td>
</tr>
<tr>
<td>Week 16-25</td>
</tr>
<tr>
<td>Week 26 and thereafter</td>
</tr>
</tbody>
</table>

---

* Known indicates condition present for >3 months*
PMR will have complete information. Furthermore, blinding is not feasible because steroid treated patients are the ones who originally received PMR diagnosis. Nonetheless, during the re-evaluation, physicians will have information on the course of the disease for these patients, not originally available at the time of initial diagnosis. Therefore, the gold standard for the diagnosis of PMR will be the “clinical diagnosis” on this 5-point scale based upon the patients’ entire course by at least 2 participating physicians with an expertise in PMR. This evaluation will ensure the homogeneity of decision to diagnose PMR and also constitute a reproducibility assessment of the newly developed criteria (i.e. to estimate the reliability of classification based on only the first visit).

D.2 Study Measurements

Several measurements, corresponding to 9 domains of disease activity (see table 4), required to accomplish the 3 specific aims, are described below. Our justification for inclusion of each of these measurements is described in this section. Briefly, each of these domains are deemed important based on either (a) inclusion in existing criteria (pain, stiffness, acute phase reactants, steroid response), (b) expert consensus and Delphi survey (clinical examination and ultrasound imaging), (c) evidence from the literature (functional status, QOL). Values and change in values of each of these measurements over time will be assessed as disease activity and remission criteria (Aim 1), response criteria (Aim 2) and classification criteria (aim 3). All measurements, except ultrasound, will be collected at each visit. Ultrasound evaluations will be performed at baseline and at 26 weeks.

Pain domain: Pain is the dominant feature of PMR and a mandatory criteria in previously proposed EULAR PMR response criteria\textsuperscript{67}. Almost all patients report bilateral shoulder or pelvic girdle pain at initial presentation and pain resolves quickly (50% of patients within weeks) with initiation of steroid therapy\textsuperscript{28}. In our study, we will assess pain using a horizontal 100 mm visual analogue scale in 4 separate locations (shoulder, pelvic, neck and overall, 0=no pain, 100=worst pain).

Morning Stiffness: This is another prominent domain in PMR and had been included in previous diagnostic and response criteria. We will assess the duration of morning stiffness (during the past 24 hours) at each study visit. Stiffness will be assessed by direct questioning of the patient, e.g. “How long did you feel stiff this morning?” with 0 indicating no stiffness.

Functional status & quality of life (QOL): QOL is increasingly recognized as a major contributor of the burden of chronic diseases, especially rheumatic diseases. QOL takes account of the effect of impairments and disability on the patient in addition to other influences including personality, social and physical environment, economic resources and culture. Several generic and disease-specific measures of QOL and functional disability have been validated and are being used in rheumatic diseases\textsuperscript{68-71}. However, evidence is scanty in PMR, as described below. In this study, we propose to use 2 QOL instruments, as described below.

SF-36 (Medical Outcome Study Short Form 36 item health status survey questionnaire) is the most widely used generic health status measure to assess impairment and disability\textsuperscript{72}. The SF-36 is a 36-item questionnaire which measures physical functioning, social functioning, role limitation due to physical health problems, role limitation due to emotional problems, mental health, vitality/energy, bodily pain and general health perceptions. It is the current gold standard measure of QOL and has been shown to be a reliable, valid and responsive measure of outcome in a number of diseases\textsuperscript{73-76} including rheumatoid arthritis\textsuperscript{69, 77}. The majority of studies that utilized SF-36 demonstrate that patients with rheumatic diseases have significantly lower scores on all SF-36 dimensions than those without, especially for physical functioning and pain. Furthermore, SF-36 scores significantly correlate with disease activity and severity measures\textsuperscript{69}. Another important advantage of SF-36 in PMR is that it can provide additional information about comorbid conditions or treatment side effects that cannot be easily
captured by other measures such as pain, stiffness or clinical examination. This is particularly important in the elderly with several comorbid conditions. Thus, this generic instrument will be complementary to other disease specific measures in our study. We will evaluate SF-36 according to established scoring systems\textsuperscript{72, 78, 79}. The SF-36 is scored to produce eight dimension scores and two physical and mental component summary scores.

Table 4. Domains and measurements

<table>
<thead>
<tr>
<th>Domains</th>
<th>Assessor</th>
<th>Scale/measures</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pain</strong></td>
<td>Self-reported</td>
<td>(VAS 100 mm) Pelvic girdle pain Neck pain Overall pain</td>
</tr>
<tr>
<td><strong>Morning Stiffness</strong></td>
<td>Self-reported</td>
<td>Duration of morning stiffness (minutes)</td>
</tr>
<tr>
<td><strong>Functional status/disability &amp; QOL</strong></td>
<td>Self-reported</td>
<td>MHAQ SF-36</td>
</tr>
<tr>
<td><strong>Global well-being measures</strong></td>
<td>Self-reported</td>
<td>Patient global assessment (VAS 100 mm) Physician global assessment (VAS 100 mm)</td>
</tr>
<tr>
<td><strong>Fatigue</strong></td>
<td>Self-reported</td>
<td>VAS-Fatigue 100 mm</td>
</tr>
<tr>
<td><strong>Clinical examination</strong></td>
<td>Physician</td>
<td>Shoulder tenderness, pain on motion &amp; limitation (dichotomous) Hip tenderness, pain on motion &amp; limitation (dichotomous)</td>
</tr>
<tr>
<td><strong>Steroid therapy</strong></td>
<td>Physician</td>
<td>Steroid response Steroid dose Steroid dose change Steroid discontinuation (dichotomous)</td>
</tr>
<tr>
<td><strong>Acute phase reactants</strong></td>
<td>Laboratory</td>
<td>ESR CRP</td>
</tr>
<tr>
<td><strong>Imaging</strong></td>
<td>Physician</td>
<td>Shoulder ultrasound (dichotomous) Pelvic ultrasound (dichotomous)</td>
</tr>
</tbody>
</table>

VAS, visual analog scale; QOL, quality of life; mHAQ, modified Health Assessment Questionnaire; SF-36, Medical Outcomes Study Short Form; ESR, erythrocyte sedimentation rate; CRP, C-reactive protein

We are only aware of \textbf{2 unpublished studies that have used SF-36 in PMR patients}. In an unpublished study by one of the co-investigators\textsuperscript{28}, the QOL of PMR patients at presentation, as measured by mean SF-36 physical and mental component summary scores (PCS=31.5, MCS=38.9), was substantially lower than general population norms (PCS=44.7, 53.2 MCS) for 65-74 year olds. By the end of 12 months of follow-up, physical and mental scores had, on average, improved by 12.6 and 11.2 points, respectively. Higher levels of disease activity in all markers were associated with lower physical and mental scores. SF-36 scores correlated significantly with pain, stiffness and inflammatory acute phase reactants. In another unpublished survey conducted by the PI of this application (Maradit Kremers, unpublished data), 105 PMR patients filled the SF-36 questionnaire several years after the resolution of their PMR. Findings indicated that the PMR subjects had worse physical functioning but better mental functioning compared with their peers of the same age and sex. The age- and sex-adjusted physical composite score was 48.45 (SD: 10.69) and mental composite score was 52.85 (SD: 7.27). There was no correlation between the two scores. These findings suggest that physical functioning declines with advancing age but mental health remains stable. Age, sex, ESR at baseline, marital status and time since diagnosis (time from incidence date to survey date) were not associated with physical and mental scores. The lack of any trend with time since diagnosis points to the
conclusion that PMR, although devastating during the course of the disease, has no long-term impact on physical and mental functioning of patients shortly after full remission of symptoms.

**MHAQ** (Stanford Health Assessment Questionnaire (HAQ)) is the most common and widely used disease-specific instrument in rheumatology. HAQ focuses on two dimensions of health: physical disability and pain. In this study, we propose to use the modified HAQ (MHAQ) questionnaire which is a shortened version of the HAQ, with reduction of the 20 items (8 categories) in the original questionnaire to 8 items. The MHAQ is an 8-item, disease-specific questionnaire which measures dressing/grooming, rising, eating, walking, hygiene, reach, grip and activities on 2- and 3-pt scales to yield an overall score. The MHAQ score is calculated as the mean of the scores for each activity. MHAQ was used in 2 separate PMR studies by one of the co-investigators. HAQ was responsive to change and correlated well with conventional indices of disease activity in PMR. However, fixed disabilities like osteoarthritis, shoulder capsulitis and systemic diseases affected its interpretation. The sections of the HAQ measuring disability related to inflammatory stiffness/proximal involvement showed greater responsiveness to change than other sections, and hence may have a greater role in evaluating disease activity in PMR. In another larger study, by the end of 12 months of follow-up of PMR patients, the mean change in HAQ scores in PMR patients was -0.83 (-0.95 to -0.72) and the percentage of patients with a score of less than 1 increased from 30% at baseline to 88% at month 12. Higher levels of ESR, CRP and morning stiffness were all independently associated with lower HAQ scores. Therefore, these 2 studies justify the feasibility of using MHAQ in PMR patients and demonstrate the correlation of MHAQ with disease activity measures and sensitivity to change during the active period of the disease.

**Global well-being measures:** Global measures are designed to measure QOL in a comprehensive manner, usually in a single question. Indeed, physician and patient assessment of global status are included in the ACR core dataset for rheumatoid arthritis. Both the patients and the physicians assessment of global status and overall disease activity will be evaluated on a 100 mm VAS scale. Although the patients' perception of their health status is poorly assessed by their physicians, we believe it is important to collect both the patient and physician assessments in this validation study, in order to better capture the clinical practice and treatment decisions of the physicians with respect to their assessment of disease activity at each visit.

**Fatigue:** In addition to disability, physical parameters such as fatigue are also measurable features of rheumatic diseases and increasingly recognized as a significant contributor to decreased QOL. Fatigue was also demonstrated to improve with treatment indicating its sensitivity to change over time. We chose to assess fatigue as a potential criterion because pain, the most prominent symptom in PMR, has the strongest association with fatigue. Disease-related fatigue can be evaluated by using a variety of instruments. We will use a single item 100 mm VAS scale since it performs as well as or better than longer scales with respect to sensitivity to change, and is at least as well correlated with clinical variables as longer scales.

**Clinical examination:** The only clinical examination measures to be assessed in our study (by physician) are tenderness, pain on motion and limitation (dichotomous) in shoulders and hip. According to expert consensus, there was >80% agreement on importance of these criteria but Delphi respondents rating was only 50%.

**Steroid therapy:** The mainstay of therapy in PMR is oral steroids. Prednisone, 15 to 25 mg, usually suppresses inflammation dramatically. However, up to 60% of patients experience disease relapse during steroid tapering, and several studies indicate that steroid treatment can only rarely be discontinued before 2 years. Most clinicians consider a rapid resolution of symptoms with steroid therapy a diagnostic hallmark of PMR, though some studies report slower response rates in many patients. Lack of response may also indicate an incorrect diagnosis. Therefore, we will assess the
validity of steroid therapy characteristics as potential criteria for classification, disease activity and response.

In our Delphi survey, 99% of respondents agreed that steroid response is an important criterion in PMR. There was variation in respondents’ opinion of what the initial dose should be, what constitutes a rapid response and time to response. In our study, steroid therapy will be examined in a number of ways as potential response and remission criteria. We will specifically consider important components of steroid therapy: **rapid, complete and sustained response to low dose steroid therapy** (i.e. 15 mg daily). Evaluation at 1 week will define whether the clinical response is rapid and complete. Evaluation at 4 weeks will define whether the laboratory response is complete. Even if there is clinical response at 1 week the laboratory parameters may take another 2-3 weeks to normalize. The global clinical response along with normalization of laboratory abnormalities of acute phase response will distinguish the non-specific steroid response from the ‘PMR’ response. Evaluation at 26 weeks will define whether the response is sustained. We will also examine daily dose (continuous), dose changes (increase, decrease, unchanged) and steroid discontinuation (e.g. time to discontinuation) as potential disease activity and response criteria (e.g. remission while on or off steroid therapy). Steroid therapy will be assessed at each visit and tapering details between visits will be collected through record review and questionnaires.

**Acute phase reactants:** Erythrocyte sedimentation rate (ESR) by the Westergren method and C-reactive protein (CRP) are part of current practice to evaluate diagnosis and response in PMR. Although they are non-specific, they are the most widely used laboratory tests in the absence of more specific biomarkers. These tests can be performed at all participating centers. The test results will be standardized based on the normal values provided by the centers.

**Imaging-ultrasound:** Ultrasound has been used for the investigation and management of patients with musculoskeletal diseases for at least 30 years. Technology improvements in the last decade greatly improved and more recently, its use has been directed towards the assessment of patients with inflammatory arthritis. Ultrasound has recently been used in PMR patients as well. It depicts characteristic pathologic findings of shoulders and hips that can aid in distinguishing PMR from other diseases that may mimic PMR. Typical findings on ultrasound include subdeltoid bursitis and biceps tendon tenosynovitis at the shoulders, and less frequently synovitis of the glenohumoral joint. In the hips, ultrasound often reveals synovitis and trochanteric bursitis. Inflammatory shoulder lesions were observed even in patients with normal ESR values and it has been suggested that ultrasound may facilitate the proper diagnosis in patients with the typical proximal symptoms of PMR who also have normal ESR.

In our study, we propose to employ bilateral ultrasound examinations of hip and shoulders at baseline visit and at 26 weeks of follow-up. Evaluations will be made according to EULAR guidelines. In particular, we will assess inflammatory changes, such as subdeltoid bursitis, tenosynovitis of the biceps tendons, hip joint synovitis and trochanteric bursitis.

A rheumatologist or a radiologist who is experienced with musculoskeletal ultrasound of shoulders and hips will perform the ultrasound examination. For the shoulders linear probes are necessary that provide a frequency in the range between 6 and 10 MHz. For the hips linear or curved array probes are necessary that provide a frequency in the range between 5 and 8 MHz. Table 5 describes the core criteria to be evaluated. Table 6 outlines the scans that are to be performed.
Table 5. Core ultrasound criteria to be evaluated

<table>
<thead>
<tr>
<th>Joint</th>
<th>Pathology</th>
<th>Definition*</th>
<th>Parameters to be evaluated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shoulder</td>
<td>Subdeltoid bursitis</td>
<td>Anechoic or hypoechoic thickened tissue between rotator cuff and deltoid muscle that may be compressible and displaceable and that may exhibit color signals. Minimum sagittal diameter at any region in any position from internal to external rotation: &gt; 2.0 mm</td>
<td>Present: yes / no Maximum sagittal diameter</td>
</tr>
<tr>
<td>Shoulder</td>
<td>Tenosynovitis at long head of biceps tendon</td>
<td>Anechoic or hypoechoic thickened tissue within tendon sheath that may exhibit color signals.</td>
<td>Present: yes / no Maximum diameter</td>
</tr>
<tr>
<td>Shoulder</td>
<td>Effusion / synovitis</td>
<td>Anechoic or hypoechoic thickened tissue that may be compressible or displaceable and that may exhibit color signals. Axillary region: Sagittal diameter: &gt; 3.9 mm and / or joint capsule not parallel to the bone Posterior region: Sagittal diameter: &gt; 2.1 mm in any position from internal to external rotation. Synovitis is present if detected in axillary and / or posterior region</td>
<td>Present: yes / no Maximum sagittal diameter at axillary recess and at posterior region</td>
</tr>
<tr>
<td>Hip</td>
<td>Trochanteric bursitis</td>
<td>Anechoic or hypoechoic thickened tissue lateral to the greater trochanter</td>
<td>Present: yes / no Volume (longitudinal, transverse, and sagittal diameters)</td>
</tr>
<tr>
<td>Hip</td>
<td>Synovitis / effusion</td>
<td>Anechoic or hypoechoic thickened tissue that may be compressible or displaceable and that may exhibit color signals.</td>
<td>Present: yes / no Sagittal diameter between bone and joint capsule at the middle of the concavity of femoral head / neck</td>
</tr>
</tbody>
</table>

* The text refers to the definitions of the OMERACT special interest group on musculoskeletal ultrasonography and to standard reference values for musculoskeletal US.

Table 6. Scans to be performed (according to EULAR guidelines)

<table>
<thead>
<tr>
<th>Joint</th>
<th>Scan</th>
<th>Pathologies to evaluate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shoulder</td>
<td>Biceps tendon transverse</td>
<td>Biceps tendon tenosynovitis</td>
</tr>
<tr>
<td></td>
<td>Biceps tendon longitudinal</td>
<td>Biceps tendon tenosynovitis</td>
</tr>
<tr>
<td></td>
<td>Shoulder anterior transverse</td>
<td>Rotator cuff, bursitis, impingement, erosion</td>
</tr>
<tr>
<td></td>
<td>Shoulder anterior longitudinal</td>
<td>Rotator cuff, bursitis, impingement, erosion</td>
</tr>
<tr>
<td></td>
<td>Shoulder lateral transverse</td>
<td>Rotator cuff, bursitis, impingement, erosion</td>
</tr>
<tr>
<td></td>
<td>Shoulder lateral longitudinal</td>
<td>Rotator cuff, bursitis, impingement, erosion</td>
</tr>
<tr>
<td></td>
<td>Shoulder posterior transverse</td>
<td>Rotator cuff, bursitis, impingement, erosion, synovitis</td>
</tr>
<tr>
<td></td>
<td>Shoulder posterior longitudinal</td>
<td>Rotator cuff, bursitis, impingement, erosion, synovitis</td>
</tr>
<tr>
<td>Acromioclavicular joint,</td>
<td>Longitudinal to joint</td>
<td>Osteoarthritis, synovitis</td>
</tr>
<tr>
<td>Axillary</td>
<td></td>
<td>Synovitis, erosion, osteoarthritis, axillary arteritis</td>
</tr>
<tr>
<td>Hip</td>
<td>Anterior longitudinal</td>
<td>Synovitis, iliopsoas bursitis, osteoarthritis</td>
</tr>
<tr>
<td></td>
<td>Anterior transverse</td>
<td>Synovitis, iliopsoas bursitis, osteoarthritis</td>
</tr>
<tr>
<td></td>
<td>Greater trochanter longitudinal</td>
<td>Trochanteric bursitis</td>
</tr>
<tr>
<td></td>
<td>Greater trochanter transverse</td>
<td>Trochanteric bursitis</td>
</tr>
</tbody>
</table>
**Additional measurements** include demographics (age, sex, race, ethnicity), vital signs and symptoms, weight, height, family history of rheumatic/autoimmune diseases, comorbidities, comedications and several laboratory measurements (hemoglobin, complete blood count, lipids, serum protein electrophoresis, rheumatoid factor (RF) and anti-cyclic citrullinated peptide (anti-CCP) antibodies (for differential diagnosis).

**D.3 Methods for Specific Aim 1. To develop a reliable and valid composite disease activity score and remission criteria in PMR**

**Overview:** Under aim 1(a), we will evaluate the validity, reliability and reproducibility of each of the core disease activity domains (Table 4) based on patient-reported (pain, stiffness, function, global assessment, fatigue), physician-reported (global assessment, tenderness, pain on motion and limitation of hip and shoulders, steroid therapy), laboratory (ESR, CRP) and ultrasound-based measures. We will consider all of these domains and measures for potential inclusion in the definition of disease activity. We will then derive (aim 1(b)) a composite disease activity score and remission criteria incorporating these core domains. These analyses will be restricted to the estimated 126 subjects who are confirmed to have PMR (see figure 1 in section D.1.). Aim 1(c) will involve assessment of discriminant validity of the core domains and the composite disease activity score in distinguishing PMR patients from other similar conditions mimicking PMR (see figure in section D.1.).

**D.3.A Methods for Specific Aim 1(a)**

This aim involves evaluation of the validity, redundancy and reproducibility of each of the domains and measures included in Table 4. **Validity** is the extent to which an instrument specifically measures the phenomenon of interest. The initial face and content validity of our proposed set of domains have already been tested through consensus meetings and Delphi surveys (see Preliminary Studies section). Furthermore, most of the domains are part of previously proposed diagnostic and response criteria for PMR (see Background section). Therefore, aim 1(a) will mainly focus on establishing criterion and construct validity for these domains and associated measures, consistent with several guidelines including the OMERACT filter \(^{55, 56, 101}\).

Criterion validity of each of the domains in our study is the correlation of that domain or a measure of the domain with another domain or measure of PMR. Concurrent criterion validity is the correlation of the measures assessed at the same point in time. Concurrent criterion validity is essentially a correlational measure of convergent construct validity. In our study, these correlation assessments will be performed for all measures performed during a study visit. We will use standard techniques applied in rheumatology and psychometric criteria development \(^{102, 103}\). Convergent construct validity will be assessed by the Spearman’s rank correlations. Values greater than 0.7 will be considered high correlation, 0.4-0.7 moderate and less than 0.4 low correlation. Construct validity of ultrasound measures will be similarly assessed by examining Spearman correlations with patient’s and physician’s global assessment of disease activity as the surrogate measures. **Redundancy**, or colinearity of measures will be investigated by the Pearson’s correlation coefficient values, where a coefficient >0.7 will be considered to represent evidence of colinearity or redundancy. Steroid treatment characteristics are an essential component of construct validity, as any definition of disease activity is related to the treatment decision-making. Under aim 1(a), we will examine the correlation of steroid dose and dose changes (increased, decreased, unchanged, discontinued) with other disease activity domains.

It is quite likely that the form of relationship or correlation between domains (measures) is different at different levels. For example, pain scores and ESR may be perfectly correlated at high levels but the correlation and the functional form of the relationship may change at low pain levels (i.e. patient has
little pain but the ESR values remain high). We will use scatter plots to investigate the functional form of relationships at different time points and identify change points, if there are any. Furthermore, subgroup analyses will be performed based on baseline patient characteristics, i.e. patients with low ESR, as the correlations between measures may differ in these patients.

**Internal consistency** is a measure of the degree to which various measures hold together, or the degree to which they measure the same underlying construct. This will be determined by Cronbach’s alpha with cut-off values of >0.9 as indication of perfect consistency and <0.6 as poor consistency. Feasibility and acceptability of measures will be determined through an examination of ease of scoring, simplicity and item non-response rate and endorsement frequencies across response categories.

**Reproducibility** (reliability, consistency) is the extent to which a measurement yields the same result on repeated measurements by the same observer (intra-observer) or by different observers (inter-observer), as compared to the measurement differences between patients. Since we anticipate that the measurements over time are rarely exactly the same in PMR (especially early in the disease course), our analyses will focus on intra- and inter-observer reproducibility of physician-based measures (global assessment, clinical examinations and ultrasound). In a small number of stable patients, it may also be feasible to assess the test-retest reproducibility of patient-based measures by administering questionnaires with a 2-5 days interval. When developing a composite disease activity criteria (as in aim 1b), each measure in the criteria must demonstrate reliability, since an unreliable measure only adds noise to the scale. Therefore, the reliability of each of the individual measures needs to be assessed prior to development of a composite score.

Intra- and inter-observer reproducibility of clinical examinations will be assessed at 2-3 participating sites and will involve at least 2 rheumatologists and 2 generalists at each site. A total of 20 patients with PMR will be evaluated independently by the 4 physicians twice and 1-2 days apart. A similar method will be used to assess the reproducibility of ultrasound assessments. A total of 20 patients with PMR will be evaluated independently by 4 observers at 2 time points (baseline and at 6 months). Ultrasound reproducibility assessment will be performed at 2-3 participating centers. Effort will be made to ensure that the observers have different levels of experience. In both reproducibility assessments, observers will meet ahead and review the study protocol and less experienced observers will receive basic training. The observers will also document the time required for each assessment (feasibility). Statistical analysis of intra-observer and inter-observer reproducibilities will be evaluated using a two-way mixed effect model, leading to an overall components of variance decomposition (between patients, within a patient at different times, between observers, between items, etc). Individual aspects of the decomposition can also be summarized using intra-class correlation coefficients (ICC). Cronbach’s alpha is related to the variance of the per-subject latent trait versus the total variance in this model.

**D.3.B Methods for Specific Aim 1(b). Derivation of a composite disease activity score**

A composite disease activity score will be derived using data from Aim 1(a). A composite disease activity score is more useful than an individual measure of disease activity because it can incorporate several aspects of disease manifestation into a single descriptive indicator. This composite score can provide a uniform outcome measure of disease activity with greater face validity, can avoid multiplicity and can enhance precision.

Since this is a longitudinal study with repeated measures, it is feasible to develop a continuous disease activity score. This continuous score can then be dissected to an ordered categorical score, if deemed necessary, depending on prediction of future outcomes, such as the prognostic value of disease activity scores to predict physician’s decision to initiate or change steroid dose or, likelihood of
reversal of disease activity back to a higher level. Undoubtedly, PMR subjects typically experience periods of inactive disease followed by disease flares. Thus, any PMR disease activity score must allow patients to shift from one disease activity state to a higher disease activity state. Given our extensive analyses under Aim 1(a), disease activity scores will likely perform well in distinguishing patients with high, medium, low and no disease activity states. However, misclassification bias is always possible, especially for those patients on the boundary.

The ultimate test of validity is the ability of an instrument or score to predict phenomena of clinical interest\textsuperscript{105, 106}. The distinction between true PMR cases and the other diseases chosen for study controls is an endpoint of critical importance. We will create candidate models for the composite score using multiple regression, with the 0-5 disease rating scores as the dependent variable. Variables will be grouped into 3 broad classes: imaging measurements, other information known at the initial visit, and follow-up information such as early response. We will assess how important each item is within each group, in terms of both statistical significance and decrease in overall classification error, as well as the overall utility that both imaging and later evaluations add to the initial core measurements.

Note, the objective is not to simply derive a model based upon stepwise regression techniques but to consider various models based upon these different predictor sets and to evaluate these models in terms of their association with both physician and patient assessments, their robustness to changes in a single measure, content validity scores of the individual core measures as found in aim 1a and the content validity scores of the various composite measures as obtained from consensus (i.e. experts will be asked to rank the easiness and credibility of the various candidate composite disease activity scores). For a multiple linear regression model which already includes 5 covariates (e.g. one for each domain) with a squared multiple correlation R-square of 0.60, a sample size of 120 will have 90% power to detect at $\alpha = 0.050$ an increase in R-square of 0.033 due to including 1 additional covariate.

An alternative approach, more in keeping with psychometric tradition, is to first normalize each variable to the same standard deviation, and then form a simple sum of the items. For questionnaire data whose responses are all concordant, e.g. a 5 point Likert, the scaling is unnecessary, but it is mandatory with the different item ranges found in this study. If all the items used have been shown to be reliable and valid, this will decrease the variance of the score, at the possible cost of some precision that could be gained by using "optimal" weights. We will also explore a hierarchical random effects model on the scaled data, where the individual coefficients $\beta_i$ are assumed to come from a common $N(\mu, \sigma)$ distribution. This strikes a balance between unrestricted regression ($\mu = \infty$) and the simple summation ($\sigma = 0$) model; the optimal value for the coefficient shrinkage will be based on Akaike's Information Criteria (AIC).

In additional analyses, we will assess the construct validity of the composite disease activity scores by examining the correlation between the composite measure and its components. The primary summary will be a guided factor analysis of the individual items, where the first factor is constrained to be the composite score. The spearman correlation matrix will provide augmental data.

One important issue in derivation of the composite disease activity score in PMR is the \textit{weighting} of individual domains or measures, because the validity of each domain in defining disease activity in PMR is likely different depending on the time since initial diagnosis. For example, early in the disease course, patient-derived measures such as pain and QOL may be important and the main drivers of the disease activity. In contrast, later in the disease course, therapeutic changes typically target high acute phase reactants and therefore, physician-derived measures may be more important. Therefore, the relative weight of each of the measures in the composite disease activity score will differ depending on time since diagnosis. Technically these differences may be thought of as time by modifier interactions. The dependency of repeated measures form the multiple time points for individual patients will be accounted for in the analyses by either longitudinal or resampling methods.
We anticipate that there will be some discordance between patient’s and physician’s assessments of disease activity. Patients and physicians typically think differently about how to define disease activity. In judging the level of disease activity, patients may rate their complaints higher than, for example, abnormal laboratory results, whereas physicians will tend to give weight to the latter observations, irrespective of patient complaints. Therefore, models examining these outcomes will have different results, depending on the extent of discordance. Examination of the weights will identify in a sense the characteristics yielding the discordance, but because of colinearities these will be difficult to interpret. We will thus also consider regression techniques to identify the most significant predictors for any discordances.

The ideal validation of the composite score is, of course, data from a new study. We plan to assess the reproducibility of the composite disease activity score in 2 separate longitudinal studies. One of these is a prospective observational study by one of the co-investigators and includes all (except fatigue) of the measures required for validation. The other dataset is retrospective observational data from the Mayo Clinic PMR cohort. Although this retrospective dataset does not include all measures, patients included in this cohort had been followed up for a period of up to 5 years following diagnosis and available data include visit level details for all patients, especially steroid therapy details abstracted at every visit during the 5 years of follow-up. In the current grant, we will get first estimates of predictive validity in two ways, a statistical approach using the main endpoint of aim 1b, and investigation of other related endpoints that were not directly used in the creation of the regression model. For the first, we will use bootstrap calibration. This repeatedly sets aside a portion of the sample, rebuilds the model of aim 1b on the subset (with the same y and x), and assesses it on the left out portion. The resultant estimates can be used to shrink or recalibrate the coefficients of the regression model so as to provide an unbiased estimate of the prediction error. The second, more direct approach is to consider prediction of other outcome variants, e.g. how does the composite score, developed using only PMR yes/no as the target, work for predicting patient assessment of disease or physician assessment of severity.

**D.3.C Methods for Specific Aim 1(c). Development of PMR classification criteria**

Under this aim, we will develop classification criteria for PMR. It will involve assessment of discriminant validity of each of the domains and the composite score (developed in aim 1(b)) in distinguishing PMR patients from other similar conditions mimicking PMR. The study population for this aim is both the PMR cases and controls (see figure in section D.1.)

In statistical analyses, sets of classification criteria will be developed. One set will best distinguish PMR cases from the pooled set of controls (i.e. all controls combined) and other sets will best distinguish PMR cases from each control group separately. This analytic approach will provide insight as to what measures are important in a situation of general diagnosis for PMR as well as what measures are important for distinguishing between the PMR diagnosis versus another disease with similar presentation. This is especially relevant in this study since the diagnosis of PMR in different clinical settings (e.g. primary care v rheumatology clinic v other non-rheumatology clinics) may involve different sets of differential diagnoses. Three specific tools will be considered to develop the classification criteria: random forests, logistic regression and a sum of the individual scores. Random forests employ recursive partitioning to identify which measures are important. A random forests approach is an improvement over rpart alone in that it produces classifiers with higher accuracy. It also provides a measure of the importance of each measure in terms of being able to distinguish between cases and controls.

Our rationale for using random forests as the primary tool for developing the classification criteria is as follows: The proposed random forest classification technique is based upon rpart methodology. rpart is
a classification procedure that can handle both continuous and categorical variables, simultaneously. It also objectively determines cut-off values for continuous variables, which are needed when performing classification using continuous variables. \textit{rpart} also performs well when there is missing data because it has a “built-in” method for imputing the missing values. Finally, it also automatically allows for synergistic effects among the variables. All these characteristics make it an extremely useful and flexible tool for developing classification criteria.

Ideally, in developing classification criteria, one would have a training data set, upon which to develop the criteria, and an independent test data set from which to estimate the performance of the classification criteria (i.e. estimate sensitivity and specificity). Currently such an independent data set is not available (i.e. PMR cases and non-PMR controls). Naively using the data that was used to develop a classification criteria to also estimate the sensitivity and specificity of the performance of the classification criteria results in biased estimates of accuracy. Specifically, it results in estimates of sensitivity and specificity that are too high. The random forest technique allows us to obtain estimates of sensitivity and specificity that are unbiased or closer to the true sensitivity and specificity without having an independent test data set. Furthermore, random forests will also indicate the relative importance of each variable with respect to its contribution to the classification criteria. In other words, it provides a measure of the usefulness of the variable in terms of distinguishing PMR cases from subjects with other conditions.

The measures identified by random forests will be used in a logistic regression model. Logistic regression will provide another measure of the importance of each measure in the presence of the other measures in the model. It will also help refine the classifiers in that there may be measures that can be omitted without substantially reducing the classifier’s accuracy. Finally, the logistic regression model will produce a score for each subject based upon the measures in the model (note that the measures will not have equal weight). A subject’s logistic regression score can be used to classify her/him as a PMR case or not. A receiver operator characteristic (ROC) curve will be generated for the scores to determine a cut-off that provides satisfactory sensitivity and specificity.

Since our aim is to develop classification criteria for PMR rather than to test a specific hypothesis, a \textbf{sample size} calculation based on power is not appropriate. On the other hand, the metrics by which classification criteria are judged typically are sensitivity and specificity. Sample size determines how precise we are able to estimate these quantities. Hence, the sample size justification arises from the level of desired precision for estimating sensitivity and specificity. Another factor that influences the level of precision for our estimates of sensitivity and specificity are the true values of sensitivity and specificity for our classification criteria. Table 8 indicates the level of precision for our estimates of sensitivity and specificity for different sample sizes.

We anticipate that the classification criteria will achieve a minimum sensitivity and specificity of 0.70 (70%). If this cannot be achieved, the criteria will not be very useful. The sample size for the sensitivity analysis will be the number of cases, which could range from \(N = 100\) to 120. For a classification criterion that yielded a true sensitivity of 0.70, we would be able to estimate it to within \(\pm 0.04\) (confidence interval size = 0.08). For larger levels of sensitivity, the precision of the sensitivity estimate increases (i.e. the size of the confidence interval decreases). Similarly, the worst-case scenario for specificity would be a sample size of 240 and specificity of 0.70. Our estimate of specificity would be within \(\pm 0.03\) of the true specificity. As the true specificity and sample size

<table>
<thead>
<tr>
<th>Table 8. Size of confidence interval* for sensitivity and specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
</tr>
<tr>
<td>-------------</td>
</tr>
<tr>
<td>120</td>
</tr>
<tr>
<td>240</td>
</tr>
</tbody>
</table>

* two-sided 95% confidence intervals assumed
increases, the precision of our estimate will improve (i.e. the size of the confidence interval for the estimate will decrease). The proposed sample sizes yield levels of precision such that our estimates will be within at least ±0.04 of the true sensitivity and ±0.03 of the true specificity values.

**D.4 Specific Aim 2. To develop a preliminary definition of improvement for the evaluation of response to therapy in PMR**

**Overview:** Under aim 2a, we will evaluate within individual sensitivity to change (responsiveness) and discriminant validity of core disease activity domains (as measured by patient-based, physician-based, laboratory and ultrasound-based measures). We will then derive a preliminary composite definition of improvement based on disease activity (Aim 1b) and change in core disease activity domains. We will use a combination of statistical criteria and consensus-formation techniques to derive a validated definition of improvement. These analyses will be restricted to the estimated 126 subjects who are confirmed of having PMR (see figure 1 in section D.1.).


**Sensitivity to change** (responsiveness) refers to the ability of each measure to detect clinically important changes between baseline and at different follow-up time points. Clinically important change is a change that either the patients or the treating physicians think is discernible and important. These evaluations will be performed with individual measures as well as the disease activity score derived in aim 1(b). Anchor-based methods will be used to examine sensitivity to change or responsiveness. Anchor-based methods compare changes in measure values on an anchor, which we will define as a patients’ global rating of change. A patient will be considered to have improved (or deteriorated) only when he/she meets the anchor-based criteria and distribution-based criteria for establishing clinically meaningful change.

**Discriminant validity** of each measure will be assessed by evaluating the capacity of the measure to distinguish patients who experienced improvement from those who did not. At each follow-up visit, patients will be asked to judge whether the disease had improved, remained stable or had worsened as compared to the previous visit. Treating physician’s decision to change steroid therapy (increase dose, keep same dose, reduce dose, discontinue) can also be considered as a surrogate measure of physician assessment of improvement. Patients who were judged as improving will be compared with those who were judged as not improving (i.e., disease remained stable or worsened) by t-tests or the Mann-Whitney U test, as appropriate. Moreover, the level of agreement between physicians and patients in the evaluation of response to therapy will be assessed with the kappa statistic.

Convergent construct validity (of the changes over time) will be examined by assessing whether improvement (which is the construct in question) is related to other measures (change in pain scores, ESR) in a manner consistent with prior expectations. Pearson's correlation coefficients will be calculated for the absolute change in measure values over time (e.g. baseline to 6 months). Values >0.7 will be considered as evidence of colinearity or redundancy whereas, values in the mid range 0.4-0.7 will be considered as good convergent construct validity. The internal consistency of the various measures will be determined by Cronbach's alpha.

Based on these analyses, the various domains of response to therapy will be finalized. Once this is complete, we will use a combination of consensus methods and statistical criteria to define improvement, as outlined in Aim 2(b).
D.4.B Methods for Specific Aim 2(b). Derivation of a composite improvement criteria

This aim involves a combination of consensus-driven and statistical methods to derive the composite definition of improvement in PMR.

First, we will aim to derive consensus-based gold standard definitions of improvement using expert ratings of patient profiles. Each patient’s profile over time will be presented to the experts and they will be instructed to rate each profile as clinically improved or not improved. In order to achieve a gold-standard consensus definition, we will require 75%-80% consensus among the experts. Since the physicians assessments are frequently in disagreement, we can also adopt probability of agreement approach, using performance characteristic curves based on computer simulations (as previously applied in SLE)\textsuperscript{112}.

Second, we will derive various combinations of improvement criteria (based on consensus-derived definitions) and evaluate their ability to classify patients as improved or not improved. For each consensus-derived improvement criteria evaluated, we will calculate the chi-square test, the sensitivity, specificity and false-positive/negative rates against the gold standard of expert consensus. In these analyses, sensitivity will demonstrate the ability of an improvement definition to identify a patient as improved who had been classified as improved by the gold standard (i.e. consensus definition). Specificity refers to the ability of the definition to identify a patient as not improved who had been classified as not improved by the gold standard. We will also calculate kappa statistics to measure the strength of agreement between the definitions and consensus of the physicians. Only definitions with substantial agreement (e.g. kappa statistics >0.7) and high sensitivity and specificity values and low false-positive and false-negative rates will be evaluated further.

Next we will consider multiple composite models for improvement using multivariable logistic regression. The approach will be similar to that of Aim 1b deriving candidate weights for the core measures and then evaluating the models in terms of their association with improvement, content validity, easiness and credibility. Further, along with the results of Aim 1(a) we will attempt to identify a model with good performance both as a disease activity score and sensitive to improvement.

D.5 Specific Aim 3. To create an infrastructure for storage of biospecimens in order to facilitate future research on novel disease-specific biomarkers and their therapeutic potential

All PMR subjects will be asked to give 10 cc of whole blood for future studies at each of the 5 visits, and controls will be asked to give 10 cc of whole blood for these studies at baseline only. The samples will be stored and processed at the Mayo Clinic (Rochester, MN). Each center will collect one EDTA tube (whole blood for DNA extraction) and two “dry” tubes for serum. The EDTA tube will be frozen on site. The dry tubes will be centrifuged, and serum will be aliquoted on site. Aliquotes will be labeled and frozen at -80°C. Specimens will be transported to Mayo Clinic twice a year from each center. DNA extraction and storage of DNA aliquots and the serum aliquots (without thawing), will be managed at the Mayo Clinic.

D.6 Approach to missing data

We expect that the number of missing values will be small in our study, due to short duration of follow-up. Nevertheless, for each analysis, we will first examine the distribution of missing values, and will consider multiple imputation and propensity score methods to deal with those. Multiple imputation corrects the variance deflation aspect of simple imputation by using a distribution of fill-in values\textsuperscript{113}. The propensity score methods for the Cox model were developed by Pugh et al., and further developed by Robins\textsuperscript{114-116}. If the probability of missingness differs between subgroups, but is random within a group (as might occur if PMR subjects on steroids were more likely to receive a certain test
than those who were not on steroids, but the probability of missingness was random within each group of subjects) then this is a very effective analytic strategy. We will compare the results from both the complete case analyses and the multiple imputation analyses and if a difference exists, we will report results from the multiple imputation analyses.

D.7 Strengths and Limitations

The strengths of our proposed study are several-fold. First, the acceptability of the study findings should be high owing to the multinational collaboration. Second, the infrastructure already exists as the participating centers are willing and able to recruit patients within a relatively short frame. Third, the study builds upon strong validation methods developed in other rheumatic diseases. Fourth, biospecimen storage will promote future studies of the proteomics and genomics of PMR in a well-characterized patient cohort.

This proposal does, however, have potential limitations. First, our ability to perform external validation of the newly developed criteria is limited as there are not many clinical trials performed in PMR patients. Second, given the uncertainties associated with the etiology of PMR, these criteria represent clinical remission and not biological remission. Third, the study subjects in this study will be recruited from rheumatology referral centers, rather than a community based patient population. Although this may raise concerns regarding representativeness, we believe this is unlikely because GPs within the coverage area of the participating rheumatology centers will be asked to refer their suspected patients during the duration of the study. Furthermore, there are wide practice variations in individual countries and therefore, it is challenging to ensure compliance with the study protocol in a GP based recruitment plan. Finally, the duration of follow-up (6 months) may be too short to identify sufficient number of patients in complete remission, especially while off steroid therapy. However, future clinical trials in PMR are unlikely to be longer than 6 months and therefore, the criteria developed should be satisfactory.

In conclusion, this will be the first comprehensive effort to develop outcome definitions for PMR. The study infrastructure and methodology are robust and achievable. The proposed approach will take advantage of a unique set of circumstances and expertise. Findings emanating from this study will lay the foundation for international guidelines for evaluating outcomes in PMR.

D.8 Timelines

The 2-year study plan includes recruitment, follow-up and closeout. Project milestones are as follows:

<table>
<thead>
<tr>
<th>Time (months)</th>
<th>Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2</td>
<td>Staff training, IRB, preparation of study materials;</td>
</tr>
<tr>
<td>2-14</td>
<td>Patient recruitment/enrolment;</td>
</tr>
<tr>
<td>8-20</td>
<td>Patient follow-up, data entry;</td>
</tr>
<tr>
<td>21-22</td>
<td>Initial data analysis;</td>
</tr>
<tr>
<td>23-24</td>
<td>Data analysis and preparation of reports.</td>
</tr>
</tbody>
</table>
E HUMAN SUBJECTS RESEARCH

E.1 Protection of Human Subjects

Human subjects involvement and characteristics: This study involves prospective recruitment of 120 subjects with polymyalgia symptoms and 240 control subjects, who are ≥ 50 years and are not pregnant or lactating. Based on our preliminary data, we anticipate that the majority of subjects will be white (>90%) with a mean age of 70 years. We will include in our study every person at the selected recruitment sites who qualifies for inclusion during the duration of the study regardless of race, ethnicity, or socioeconomic status (age ≥ 50 years is an inclusion criteria).

Source of research materials: The research material will involve a physical examination, a venous blood sample, an ultrasound examination and questionnaire data collected at each study visit.

Potential risks: This project does not involve any experimentation upon human subjects. The potential risks of this study are minor and unlikely to occur. Subjects will undergo a venipuncture for blood collection and an ultrasound examination. These procedures will expose subjects to minimal physical risks only. The potential physical risks associated with venipuncture include pain, bruising, hematoma formation, superficial phlebitis, and extremely rare instances of infection or cellulitis. There is no known risk of musculoskeletal ultrasound. There is a potential for incidental ultrasound findings which will require management discussion with the patients.

Recruitment and informed consent: Potential PMR subjects and controls will be identified among consecutive patients referred to secondary care rheumatology clinics with suspected polymyalgia at 20 rheumatology referral centers in eight countries: USA, UK, Italy, Spain, Sweden, France, Austria, Germany. Each center will be recruiting 6 PMR cases and 12 non-PMR controls. Recruitment is expected to last 12 months with an anticipated recruitment rate of two patients per center per month. Written informed consent will be obtained by the study coordinator or by one of the investigators and this will be documented as part of the study documentation. Use of stored blood samples for future analyses will be explained to subjects while informed consent is obtained.

Procedures to minimize risk: The Institutional Review Board of each participating center (some of which are foreign centers in Europe) will review all aspects of this study and the contact procedures will be approved through institutional channels prior to use. The confidentiality will be maintained by using a study patient number for each participant. Only the center staff will know the identity of the patient. Questionnaires will be sent to the data management center (Mayo Clinic) using an identification number. At the data management center, all information on study subjects will be stored in locked file cabinets. Electronic information will be identified only through the non-identifiable study number. Only persons with appropriate authorization will have access to computer files. All members of the staff at the data coordinating center (Mayo Clinic) are trained about the importance of confidentiality and procedures by which this can be maintained. All data will be managed by study number and analyzed anonymously. All reports will be of a summary nature and no individual will be identified.

Potential benefits of the proposed research to the subjects and others: If abnormalities are noted on the ultrasound, it will be recommended that further evaluation be undertaken by physician. No other significant direct benefits to the subjects involved are anticipated, but neither is there any appreciable health risk other than those outlined above. Benefits to future patients with PMR could be substantial through early detection and treatment of this condition.

Importance of the knowledge to be gained: This application will be the first large-scale international study that examines a comprehensive list of outcomes in PMR patients. The knowledge to be gained
will clearly offset the minimal risks to subjects. The importance of the knowledge to be gained is extensively detailed throughout the application.

**E.2 Inclusion of Women and Minorities – Inclusion of Children**

In our study, we will include every person greater than 50 years old and who fulfills the inclusion criteria, regardless of race, ethnicity, or socioeconomic status. Based on the epidemiology of PMR, we expect to recruit a slightly higher number of women (60-70%). Since one of the inclusion criteria of the study mandates >50 years of age, there are no major concerns regarding women of childbearing potential, or pregnant or lactating women. Similarly, children are not included.

**E.3 Data and safety monitoring plan**

Not applicable (observational study)

**F LITERATURE CITED**


59. Dasgupta B, Hutchings A, Matson E. Polymyalgia rheumatica: the mess we are now in and what we need to do about it. *Arthritis & Rheumatism.* 2006;in press.


75. Garratt AM, Ruta DA, Abdalla MI, Buckingham JK, Russell IT. The SF36 health survey questionnaire: an outcome measure suitable for routine use within the NHS?[see comment]. BMJ. 1993;306(6890):1440-1444.


