METHODS

We used Grading of Recommendations Assessment, Development and Evaluation (GRADE) methodology [1] as a framework for the EULAR-ACR endorsed recommendation project.[Dejaco et al., ARD and A&R 2015 (in press)] The key questions for the systematic literature review (SLR) were framed in the PICO (=Population, Intervention, Comparator, Outcome) format (Table 1). For prognostic studies we considered the presence and absence of predictors as “interventions” and “controls”, respectively.

Relevant outcomes (deemed as being essential for clinical decisions) were categorized from a survey (as part of this project) among rheumatologists, general practitioners and patients. They are listed in Supplementary Box S1.

Supplementary Box S1. Outcome parameters used for the systematic literature review

- Disease remission
- Disease relapse
- Duration of glucocorticoid therapy
- Discontinuation of glucocorticoid therapy
- Development of giant cell arteritis
- Glucocorticoid side effects (diabetes mellitus/glucose intolerance, osteoporosis, cardiovascular disease, dyslipidemia, impaired wound healing, infections, osteonecrosis, myopathy, cataract, glaucoma, atherosclerosis, hypertension, peptic ulcer, weight gain, moon face, dyspnea, palpitations, fatigue, skin atrophy, bruising, mood disorders)
- Response to glucocorticoid therapy
- Cumulative glucocorticoid dose
- Acute phase reactants
- Patients assessment of global wellbeing
- Severity / duration of morning stiffness
- Lowest possible glucocorticoid dose (prednisone equivalent less than 5mg/day)
- Functional status (Health Assessment Questionnaire or other measures)
- Quality of life (Short Form-36, EQ5D etc.)
- Mortality
- Hospitalization (due to disease, its complications, co-morbidity and/or treatment related complications)
- Impact on patients' social environment
- Fatigue
- Imaging of shoulder/hip
- Healthcare resource use (health economics)
- Disease activity score

Literature search

A sensitive systematic literature search aimed at the retrieval of all articles on Polymyalgia rheumatica (PMR) was conducted between January 1970 and June 2013. An update of the search was performed in April 2014. The protocol of this project has been available since July 2013 on the homepage of the American College of Rheumatology (ACR) (www.rheumatology.org).

Role of authors in the systematic literature review

Three authors (CD, Rheumatologist, Graz, Austria; YS, Rheumatologist, Southend, UK; AH, clinical epidemiologist, London, UK) designed the key words, planned the search strategy, designed the data extraction sheet and conducted the search (counselled by PP, clinical epidemiologist, London, UK). CD and YS independently reviewed all articles identified by the literature search, performed data extraction and quality appraisal. Any disagreement was resolved by discussion. AH was consulted if CD and YS did not achieve a consensus (15.6% of articles) to make a final decision. Two other authors (SM, Rheumatologist, Leeds, UK and DC, Rheumatologist, Genova, Italy) helped with review and data extraction of non-English articles.

Search strategy including search terms used for literature search

Search strategies were designed using the following electronic databases: Ovid MEDLINE®, Embase, PubMed, CINAHL, Web of Science and the Cochrane Library. We used the thesauri for each database, text words, truncated text words and abbreviations as key words (see Supplementary Box S2 for key words used for Ovid

MEDLINE®, similar strategies were applied to other databases). “Giant cell arteritis” (and related terms) were not included as search terms due to the low expected yield and the possible large volume of non-relevant literature. The grey literature [reports by the Agency for Healthcare Research and Quality, conference abstracts from annual meetings of the ACR, EULAR and British Society of Rheumatology, as well as international PMR/Giant cell arteritis (GCA) and ANCA meetings) were reviewed, tracked to determine whether additional peer-reviewed articles not identified by the primary search were published. Trial registries, such as ClinicalTrials.gov, ISRCTN and EU Clinical Trials Register, were searched to identify ongoing and completed trials and we contacted sponsors and investigators to request any available results. Additional papers were retrieved searching the reference list of full papers and review articles and by contacting experts in the field.

Supplementary Box S2. Key words for Ovid MEDLINE®

1. Polymyalgia rheumatica #
2. Polymyalgia rheumatica $
3. polymyalgia rheumatica .mp
4. PMR NOT prenatal mortality rate $ .mp
5. PMR NOT premature mortality rate $ .mp
6. PMR NOT population mortality rate $ .mp
7. polymyalgi*
8. polymyalgia $ .mp
9. rheumatic polymyalgia $ .mp
10. polymyalgia arteritica $ .mp
11. forestier certonciny syndrome $ .mp
12. pseudopolyarthritis rhizomelica $ .mp
13. rheumatic myalgia $ .mp
14. rheumatism, inflammatory rhizomelic $ .mp
15. rhizomelic pseudopolyarthritis $ .mp

*truncation;# Mesh term; $ textword; mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword

Inclusion and exclusion of studies, data extraction

References and abstracts identified by the search were imported into bibliographic management software (Zotero Version 4.0.20, Fairfax, VA, USA) and duplicates were removed. Titles and abstracts were screened to remove editorials, commentaries and letters without patient data. The full text of each remaining article was then tested against the inclusion and exclusion criteria. The literature review team also made every effort to identify multiple publications from a single trial.

We excluded all articles that did not report original data, did not study patients with PMR, or that considered PMR and GCA patients as a single group. For PICOs on prognostic factors, we excluded all studies investigating factors that were not routinely available [e.g. cytokines other the interleukin (IL)-6, adhesion molecules ect) and/or trials with a follow-up of fewer than 6 months.

Study details (interventional studies: study design, setting, participating center(s), study duration, primary endpoint(s), criteria used to define PMR, number of patients included, proportion of patients randomized/receiving treatment, age and gender of participants, proportion (plus age and gender) of individuals completing the study, reasons for losses to follow-up, handling of missing data, intervention and control treatment including dose and administration details, flare/rescue medication; prognostic studies: Study design, setting, participating center(s), methods used to identify population, recruitment period, study duration, observational period, primary endpoint, criteria used to define PMR, number of patients included, proportion of patients in whom the prognostic factor was measured, age and gender of participants, proportion (plus age and gender) of individuals completing the study,

reasons for losses to follow-up, attempts to collect information on participants who dropped-out, differences between completers/non-completers, description of measurement of the prognostic factor), parameter required for quality assessment (including confounders detailed below) and results (related to the outcomes specified by the PICO questions) were extracted using a pre-specified data extraction sheet.

Quality assessment and generation of evidence tables

Data related to therapeutic interventions were quality-assessed using GRADE methodology. We used Review Manager (RevMan) and GRADE profiler (GRADEpro) to summarize the data on interventions and to produce the GRADE profile, respectively. The GRADE profile contains the number of studies for each intervention and outcome, the appraisal of the quality of evidence (QoE), the number of participants in intervention and control groups, the relative and absolute effect estimates and the overall quality rating (4 categories ranging from very low to high). Relative effect estimates include the relative risk (cumulative risk over an entire study using a defined endpoint) or the hazard ratio (corresponding to the instantaneous risk over the study period) for categorical outcomes. Absolute effect estimates include the number needed to treat/harm for categorical outcomes and the mean difference for continuous outcomes.

For the appraisal of studies on prognostic factors we used the Quality In Prognosis Studies (QUIPS) tool. Accordingly, we evaluated potential biases (3 levels: high, moderate, low) related to study participation, attrition, measurement of prognostic factors, measurement of/controlling for confounding variables, measurement of outcomes and statistical analysis. We additionally rated the possibility of confounding

by GCA and noted other reasons for bias, thus resulting in a total of 8 (instead of 6) categories (QUIPS+2). For the category on confounding variables the following parameters were addressed: age, gender, acute phase reactants [i.e. erythrocyte sedimentation rate (ESR) or C-reactive protein (CRP)], disease activity/severity (determined by pain, global disease activity, morning stiffness or similar), peripheral arthritis, symptom duration, comorbidities and constitutional symptoms.

Evidence tables on prognostic factors contain the study design, the quality appraisal, criteria used to define PMR, the definition and cut-offs for the prognostic factor, the number of patients, the proportion of patients with complete follow-up data, the outcomes addressed, the effect estimate (odds ratio or hazard ratio for categorical outcomes, mean difference for continuous outcomes) and the information whether the effect estimate was adjusted for confounders by statistical methods.

We attempted to perform meta-analyses (fixed effect methods) for interventional studies whenever possible to produce an overall effect estimate. Statistical heterogeneity was assessed with the chi-squared test (significance at $p<0.1$) and the I-squared inconsistency statistic (>50% indicating significant heterogeneity). For prognostic studies, meta-analysis was impossible because of the large heterogeneity between the studies (divergent study designs, different PMR case definitions, varying measurements and definitions of prognostic factors and outcomes as well as divergent quality). Therefore, results and quality appraisal are presented for each study and outcome separately.

REFERENCES

1 Guyatt G, Oxman AD, Akl EA, *et al.* GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 2011;**64**:383–94.