

EXTENDED REPORT

A comparison of the sensitivity of diagnostic criteria for polymyalgia rheumatica

H A Bird, B F Leeb, C M Montecucco, N Misiuniene, G Nesher, S Pai, C Pease, J Rovensky, B Rozman

Ann Rheum Dis 2005;64:626–629. doi: 10.1136/ard.2004.025296

See end of article for authors' affiliations

Correspondence to: Professor Howard Bird, Clinical Pharmacology Unit, Chapel Allerton Hospital, Chapeltown Road, Leeds LS7 4SA, United Kingdom; Howard.Bird@leedsth.nhs.uk

Accepted 22 August 2004

Objective: To compare the performance of the several different diagnostic criteria sets currently in use for polymyalgia rheumatica (PMR).**Methods:** 213 patients attending eight rheumatological centres in eight different European countries were studied. All had recently been referred and were considered by the senior investigator at each centre, selected because of their experience in treatment of PMR, to have this condition. By use of a standard international proforma, the requisite diagnostic points in each criteria set were sought. Sensitivity for each criterion from each set was then calculated, as well as the sensitivity of each criteria set as a whole.**Results:** Of four criteria sets compared, the Bird (1979) criteria performed best with a sensitivity of 99.5%, and the Hunder (1982) criteria second best, with sensitivity of 93.3%. These both performed significantly better than the two other criteria sets, though each of these was admittedly developed for rather specialised reasons.**Conclusions:** Although this study compares homogeneity, we suggest the Bird 1979 or Hunder 1982 criteria should be used whenever possible. Studies that have used alternative criteria may have less sensitivity in diagnosis.

Although the disease first described by Bruce in 1888¹ was probably polymyalgia rheumatica (PMR), this term was not used formally until first suggested by Barber in 1957.² Working from a spa centre in the north of England, Barber drew on his clinical experience with large numbers of patients to define a condition for which no diagnostic test existed. Unless PMR is associated with giant cell arteritis and its typical histological features, this situation still holds true. As a result, a variety of clinical diagnostic criteria sets have been suggested over the last 25 years to aid future research. To date, there has been no formal comparison of these diagnostic criteria sets in a clinical setting.

The first criteria set to be formally proposed was a multi-collaborative one from 11 United Kingdom rheumatology units in 1979, which led to the Bird/Wood criteria.³ This was soon followed by two further criteria sets (from Jones and Hazleman in 1981⁴ and from Hunder and colleagues in 1982⁵), both of which were based more on clinical expertise than on epidemiological analysis. Two further criteria sets were developed for specific purposes. That from Wilke (1985)⁶ was specifically designed for use in giant cell arteritis so has not been included in this study, and the Nobunaga criteria of 1989⁷ were designed specifically for a Japanese population and their requirements.

On an initiative of the European League Against Rheumatism's Standing Committee on Clinical Trials Including Therapeutic Trials (ESCISIT), a European collaborating PMR group has been established. As a result, an opportunity arose to compare the several different diagnostic criteria sets. This study was undertaken from eight different rheumatology centres in eight countries across Europe, which felt able to participate in the protocol that had been circulated to representatives on the committee as well as to all European centres that had published extensively on PMR in the previous decade.

This study formed part of a larger study from the European collaborating PMR group, using mainly the same group of

patients, which has led to the definition of formal response criteria for the condition.⁸

METHODS

It was a prerequisite for participating centres that each should have a senior physician with substantial previous experience in the diagnosis and management of PMR, preferably with previous publications in the field, and that adequate facilities should exist for follow up and reasonable investigations to exclude conditions that might mimic PMR, though the extent of such investigation was left to the discretion of the contributor. However, where there was clear diagnostic ambiguity, these investigations were mandatory. This method of selection of centres was judged preferable to attempting a true pan-European distribution. Centres were only included for analysis if they contributed at least 10 patients over a four year period starting in 1998. On this basis, data from several other centres that could only produce fewer numbers were not used.

At first visit a full history was taken from each patient with suspected PMR. Blood was taken for measurement of erythrocyte sedimentation rate (ESR) in the clinic and, where possible, for C reactive protein (providing this assay was available in the local laboratory).

A proforma was completed, allowing an assessment of the extent to which symptoms, signs, and initial results accorded with the different criteria sets. If there was diagnostic confusion, investigations were done to exclude the presence of alternative diseases, as stipulated in table 1. Where diagnosis was in doubt patients returned after up to one week for these results to be scrutinised.

Where the only diagnosis, on the basis of these tests, was still felt to be PMR in the opinion of the senior recruiting physician, the patient was recruited to the study and treatment was started immediately. On ethical grounds this

Abbreviations: PMR, polymyalgia rheumatica

Table 1 Conditions that might mimic polymyalgia rheumatica and investigations stipulated to exclude them

Mimicking condition	Investigations to exclude
Rheumatoid arthritis	Rheumatoid factor, x ray of hands, possibly feet, and any other affected joint
Capsulitis of shoulder	Full examination of shoulder with clinical history
Cervical spondylosis	Clinical history, x ray of cervical spine
Osteoarthritis of shoulder	Shoulder x ray
Polymyositis	CPK, electromyography where clinically indicated
Thyrotoxicosis	Thyroid function tests
Myopathies	Clinical examination, compatible electromyography
Systemic lupus erythematosus	ANF, DNA binding
Polyarteritis nodosa	Clinical picture
Dermatomyositis	Clinical picture, CPK
Multiple myeloma	Serum and urine electrophoresis
Neoplasm	Careful clinical history, chest x ray, occult blood
Parkinson's disease	Clinical history and clinical signs
Subacute bacterial endocarditis	Heart murmurs, blood cultures

ANF, antinuclear factor; CPK, creatine phosphokinase.

was left to the discretion of the physician but was normally prednisolone (non-enteric coated) 20 mg/day until the symptoms responded, when the dose was reduced in steps according to the judgement of the clinician.

It was mandatory for patients to return within two weeks of starting treatment for the physician to determine whether a significant response to the steroid had occurred. Further follow up was encouraged (though it was not mandatory) at eight and 16 weeks for adjustment of the steroid dose as felt necessary by the rheumatologist. At each subsequent visit the diagnosis of PMR was further reviewed. At six months, if there was any doubt about the diagnosis or if any other condition, such as malignancy, had developed, the patient was excluded from analysis.

Those patients from this group for whom longer (up to two years) and more frequent follow up could be provided were also included (at the discretion of the investigator) in the parallel running "response criteria" study.⁸

Data recorded on the proforma comprised the following: initials, date of birth and sex, shoulder stiffness, neck involvement, shoulder involvement, upper arm involvement, hip involvement, thigh involvement, and whether these were unilateral or bilateral, early morning stiffness, rapid onset of symptoms, associated depression and/or weight loss, visual disturbance, headache, jaw ache, and neurological features. Early morning stiffness was estimated in minutes and tenderness of the shoulders or upper arm and swelling of the shoulder joint both recorded. ESR was measured (in one centre, Leeds, viscosity was substituted) and C reactive protein was also estimated where practical.

Comparison of the data was then made with the four selected diagnostic criteria sets. Sensitivity (the proportion of patients with the disease who are positive for the feature when related to all individuals with the disease) for each

attribute and for each criteria set was calculated. Comparison of the separate sets was made by this method.

Local ethical approval was obtained at each participating centre.

RESULTS

The number of patients participating in the study from each centre is shown in table 2.

The performance of each of the criteria sets and their component parts is shown in tables 3 to 6. The first column lists the criterion, the second the number of patients possessing this criterion set against the number of patients in the study for whom this criterion was evaluable on the data provided, and the third lists this figure as a percentage (the sensitivity of the criterion). The sensitivity of the performance of the criteria set as a whole is also given at the bottom of each table.

Some of the criteria sets studied used the absence of a particular disease as one of the criteria required. Because exclusion of such conditions was inherent to the study, we have assumed that this criterion was present in all patients

Table 2 Participating centres and number of patients entered

Leeds, United Kingdom	76
Piestany, Slovak Republic	31
Jerusalem, Israel	31
Ljubljana, Slovenia	19
Stockerau, Austria	18
Pavia, Italy	17
Kaunas, Lithuania	11
Tartu, Estonia	10
Total	213

Table 3 Bird/Wood criteria (1979)

Bilateral shoulder pain/stiffness	178/196	90.8%
Duration onset <2/52	148/196	75.5%
Initial ESR >40 mm/h	158/165	95.7%
Stiffness >1 h	147/173	84.9%
Age >65 years	171/213	80.2%
Depression and/or weight loss	85/213	40.0%
Bilateral upper arm tenderness	147/195	75.5%
Probable: 3 or more	212/213	99.5%

ESR, erythrocyte sedimentation rate.

Table 4 Jones/Hazleman criteria (1981)

Shoulder or pelvic girdle pain	178/196	90.8%
Morning stiffness >1 h	147/173	84.9%
Disease duration >2/12	Not known	
ESR >30 mm/h	162/165	98.1%
or CRP >6 µg/ml	Not known	
Absence of RA	Presumed 100.0%	
Absence of muscle disease	Presumed 100.0%	

Overall performance of criteria (since presence of all is required) no greater than 84.9%.
CRP, C reactive protein; ESR, erythrocyte sedimentation rate; RA, rheumatoid arthritis.

included, though this was not always documented in respect of each individual patient. We have indicated where we have made this assumption.

Unfortunately it proved particularly hard to evaluate the Jones/Hazleman criteria, first because our proforma did not distinguish disease duration longer than two weeks, making a precise estimate of their two month cut off point impossible, and second because it uses C reactive protein. Even when C reactive protein determinations could be obtained, the varying normal range of the many different assays used throughout Europe prevented us from checking this criterion accurately.

DISCUSSION

The proposal of diagnostic criteria remains a problem, particularly for diseases where at present no clear diagnostic test exists. Even though more information may be available by way of laboratory and radiological investigations of diagnostic validity, ultimately diagnosis is still based on the consensus of a group of experienced clinicians. The ARA diagnostic criteria for rheumatoid arthritis⁹ and for systemic lupus erythematosus¹⁰ provide similar examples. In the absence of a pathological gold standard (except for cranial arteritis) we are unable to propose any alternative method of formulating criteria, and our methodology for the comparison of existing criteria sets was based upon this premise. As important as diagnostic criteria are response criteria. The American College of Rheumatology has moved in this direction more recently for more accurate assessment of response in clinical trials.¹¹ Our collaborating PMR group recognises this, and the assessment of diagnostic criteria formed only one part of a larger study in which response criteria were evaluated (for the first time in the literature),⁸ largely from patients who participated in this and other studies but were available for longer term follow up.

A full assessment of the value of diagnostic criteria needs consideration of both the sensitivity and the specificity of each criterion. Each criterion is then selected on the basis of a combination of sensitivity and specificity. To behave well a criterion should have high values for each. Sometimes sensitivity and specificity values are added together to produce a "relative value", which many find easier to understand than the Youden index.³ A fundamental weakness of this study remains that the data available did not allow the calculating of specificity for each attribute as it was not possible to enrol a control group of patients with conditions that mimic PMR for identical study in each of the many participating centres, largely on the grounds of expense. Therefore, our results might be better described as a "test of homogeneity" and it is possible our results are biased towards the study of "true" PMR rather than a clinical syndrome that might appear indistinguishable initially. In addition, patients with giant cell arteritis were not specifically sought and, by implication, excluded, though this reflects conventional practice throughout Europe whereby the

Table 5 Hunder criteria (1982)

Patient's age >50 years	213/213	100.0%
Bilateral aching/tenderness for one month or more of:		
neck or torso		
shoulders or upper arms	183/196	93.3%
hips or thighs		
ESR >40 mm/h	158/165	95.7%
Exclusion of other diagnoses	Presumed 100.0%	
Definite: all of above	183/196	93.3%

ESR, erythrocyte sedimentation rate.

Table 6 Nobunaga criteria (1989)

Bilateral muscle pain for two weeks in		
>2 of:		
neck	181/196	92.3%
shoulders	93/213	43.7%
upper arm	178/196	90.8%
hips/pelvis	170/213	50.2%
thighs	82/213	38.5%
thighs	138/213	64.8%
Normal myogenic enzymes	Estimate 100.0%	
ESR >40 mm/h	158/165	95.7%
No swelling hand joints	166/195	85.1%
The presence of all four criteria is required for diagnosis	101/149	67.8%

ESR, erythrocyte sedimentation rate.

majority of such patients are initially referred to ophthalmologists rather than to rheumatologists.

It is also accepted that our selection of centres (and therefore the geographical distribution of patients) is a little arbitrary, based as it was on the countries and centres available for collaboration at the initiation of the study. For diagnostic criteria, we felt it important to insist on the diagnostic opinion of a senior physician experienced in this disease. However, the inclusion of patients from eight different centres ensured that no single opinion predominated. We also excluded centres that were unable to contribute 10 or more patients, mainly for logistical reasons and partly on the grounds of expense.

With recruiting centres selected primarily for these reasons, we were then dependent upon the local customs at each centre. Some European countries have greater access to longer term follow up than others, and the mechanics of the study deterred participation of centres in several other European countries that were unable to provide long term follow up or which had experienced increasing difficulty in recruitment, for PMR is a disease now often treated by the primary care physician and not necessarily referred to hospital. The facilities for diagnosis also varied between centres, restricting the collection of investigations on patients with diseases that mimic PMR, and preventing an analysis by specificity as well as by sensitivity. A particular difficulty was the lack of C reactive protein at many of the participating centres, which relied upon ESR alone. This made comparison of the Jones/Hazleman 1981 criteria particularly difficult, a point that had not been appreciated when the study was designed and the centres recruited.

Although pan-European collaboration is to be encouraged, we also have slight anxiety that all participants adequately understood the English required or, whenever translation by the clinician was required, individuals were adequately able to separate "stiffness" from "pain", which is a feature of some criteria sets.

One potential source of error, requiring clarification, is that the senior physician in each of the recruiting centres might have had preconceptions about the various criteria sets used, inevitably defining PMR by comparison (conscious or subconscious) with an existing criteria set. A specific concern is the large number of patients recruited from a single centre (Leeds), the address of the principal author of this paper and the first author on one of the papers defining a criteria set studied. In order to avoid any bias, for the purposes of this study only patients treated by a consultant colleague (CP) were used in this analysis. The first author (HB) participated in a coordinating role only. Although this still may not completely exclude such bias, we think it unlikely that this has occurred. Moreover, the weight of the many other participating centres included in the study would also reduce it.

Those criticisms apart, to our knowledge this represents the first systematic comparison of the several different criteria sets currently in use. It is of interest that criteria sets based on the option of including some of a group of criteria perform better than those for which the presence of all of a set of criteria are required. The optimum study, which would have allowed study of both sensitivity and specificity (the best criteria performing well on both of these attributes), is perhaps better suited to a more detailed study of larger numbers of patients in just one or two centres in the most "Westernised" countries, providing patients are seen in the hospital rather than in primary care. Therefore, a "best buy" criteria set for clinical trials may have eluded us but at least our analysis, based only on sensitivity, ensures that with the use of the sets that perform best, patients with the condition will not be missed in clinical practice.

The study has also provided a salutary insight into pan-European collaboration, a laudable aim that is one of the founding principles behind EULAR. This study, which had the support of ESCISIT, was one of the early endeavours initiated soon after this committee's reconstitution in the mid-1990s. A firm policy decision was taken to encourage participation on the basis of enthusiasm and an interest in the disease rather than on the basis of scientific credibility alone, which might have restricted the study to western Europe. As a result, several countries in what was then "eastern" Europe participated even before the enlargement of the European Union. The study therefore provides an interesting exercise in the advantages and disadvantages of this particular method of selecting participating centres. Inevitably, with the meagre resources available at certain centres, the benefits of multinational collaboration have here taken precedence over the benefits of scientific homogeneity, which would have been more restrictive.

ACKNOWLEDGEMENTS

The following workers in the following centres provided additional clinical support: Roberto Caporali (Pavia), Dusan Logar (Ljubljana), Michael Mates and Moshe Sonnenblick (Jerusalem), Alena Tuchynova (Piestany), and Raili Vaas (Tartu).

Authors' affiliations

H A Bird, University of Leeds, Leeds, UK

B F Leeb, Lower Austrian Centre for Rheumatology, Stockerau Hospital, Stockerau, Austria

C M Montecucco, Rheumatology Unit, IRCCS, S Matteo University Hospital, Pavia, Italy

N Misiuniene, Rheumatology Clinic, University of Medicine, Kaunas, Lithuania

G Nesher, Shaare-Zedek Medical Centre, Jerusalem, Israel

S Pai, Department of Internal Medicine, University of Tartu, Tartu, Estonia

C Pease, Department of Rheumatology, General Infirmary at Leeds, Leeds, UK

J Rovensky, Research Institute of Rheumatic Diseases, Piestany, Slovak Republic

B Rozman, Department of Rheumatology, University Medical Centre, Ljubljana, Slovenia

REFERENCES

- 1 **Bruce W**. Senile rheumatic gout. *BMJ* 1888;**2**:811–13.
- 2 **Barber HS**. Myalgic syndrome with constitutional effects: polymyalgia rheumatica. *Ann Rheum Dis* 1957;**16**:230–7.
- 3 **Bird HA**, Esselinckx W, Dixon ASTJ, Mowat AG, Wood PHN. An evaluation of criteria for polymyalgia rheumatica. *Ann Rheum Dis* 1979;**38**:434–9.
- 4 **Jones JG**, Hazleman BL. Prognosis and management of polymyalgia rheumatica. *Ann Rheum Dis* 1981;**40**:1–5.
- 5 **Chuang T-Y**, Hunder GG, Ilstrup DM, Kurland LT. Polymyalgia rheumatica: a 10-year epidemiologic and clinical study. *Ann Intern Med* 1982;**97**:672–80.
- 6 **Wilke WS**, Wysenbeck AJ, Krall PL, Segal AM. Masked presentation of giant cell arteritis. *Cleve Clin Q* 1985;**52**:155–9.
- 7 **Nobunaga M**, Yoshioka K, Yasuda M, Shingu. Clinical studies of polymyalgia rheumatica: a proposal of diagnostic criteria. *Jpn J Med* 1989;**28**:452–6.
- 8 **Leeb BF**, Bird HA, Andel I, Hueber W, Logar D, Montecucco CM, *et al*. EULAR response criteria for polymyalgia rheumatica: results of an initiative of the European Collaborating Polymyalgia Rheumatica Group (sub-committee of ESCISIT). *Ann Rheum Dis* 2003;**62**:1189–94.
- 9 **Ropes NW**, Bennett GA, Cobb S, Jacox R, Jassar RA. Proposed diagnostic criteria for rheumatoid arthritis. *Ann Rheum Dis* 1957;**16**:118–25.
- 10 **Cohen AL**, Reynolds WE, Franklin EC, Kulka JP, Ropes NW, Shulman LE, *et al*. Preliminary criteria for the classification of systemic lupus erythematosus. *Bull Rheum Dis* 1971;**21**:643–8.
- 11 **Felson DT**, Anderson JJ, Boers M, Bombardier C, Chernoff M, Fried B, *et al*. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. *Arthritis Rheum* 1993;**36**:729–40.