

Measurement Properties of the Polymyalgia Rheumatica Activity Score: A Systematic Literature Review

Thomas E. Bolhuis¹ , Lizanne E.A. Nizet², Claire Owen³ , Alfons A. den Broeder¹ ,
Cornelia H.M. van den Ende¹ , and Aatke van der Maas²

ABSTRACT. *Objective.* To perform a COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN)-based systematic literature review of measurement properties of the Polymyalgia Rheumatica Activity Score (PMR-AS).

Methods. PubMed, EMBASE, and CINAHL were broadly searched. English full-text articles, with (quantitative) data on ≥ 5 patients with PMR using the PMR-AS were selected. Seven hypotheses for construct validity and 3 for responsiveness, concerning associations with erythrocyte sedimentation rate, physical function, quality of life, clinical disease states, ultrasound, and treatment response, were formulated. We assessed the structural validity, internal consistency, reliability, and measurement error, or the hypotheses on construct validity or responsiveness of the PMR-AS based on COSMIN criteria.

Results. Out of the identified 26 articles that used the PMR-AS, we were able to use 12 articles. Structural validity, internal consistency, construct validity, and responsiveness were assessed in 1, 2, 8, and 3 articles, respectively. Insufficient evidence was found to confirm structural validity and internal consistency. No data were found on reliability or measurement error. Although 60% and 67% of hypotheses tested for construct validity and responsiveness, respectively, were confirmed, there was insufficient evidence to meet criteria for good measurement properties.

Conclusion. While there is some promising evidence for construct validity and responsiveness of the PMR-AS, it is lacking for other properties and, overall, falls short of criteria for good measurement properties. Therefore, further research is needed to assess its role in clinical research and care.

Key Indexing Terms: outcome assessment, outcomes, polymyalgia rheumatica, quality of life

Polymyalgia rheumatica (PMR) is an inflammatory rheumatic disease characterized by pain and stiffness of the hip and shoulder girdle.¹ Glucocorticoids (GCs), the mainstay of treatment, are gradually tapered to control disease activity, and at the same time try to minimize the risk of GC-related adverse events.² However, during GC tapering, disease relapses (or flares) occur in up to 55% of patients, typically necessitating therapy intensification.^{3,4}

Analogous to the Disease Activity Score in 28 joints in rheumatoid arthritis (RA), Leeb and Bird developed a disease activity score for PMR: the PMR Activity Score (PMR-AS).⁵ The PMR-AS is a composite score based on C-reactive protein (CRP), pain assessed on a visual analog scale (VASp), morning stiffness

(MS; in minutes), elevation of the upper limbs (EUL; active shoulder abduction ranging from 0 to 3), and physician disease activity assessed on a VAS (VASp). The PMR-AS is calculated as follows: $CRP \text{ (mg/dL)} + VASp \text{ (0-10)} + (MS \times 0.1) + EUL \text{ (0-3)} + VASp \text{ (0-10)}$. This outcome measure is the most frequently used instrument for disease activity in PMR studies, with up to 40% of publications selected by Duarte et al between 2007 and 2014 using the PMR-AS.⁶ The PMR-AS is being used in several ongoing trials. Moreover, a trial of baricitinib (ClinicalTrials.gov: NCT04027101) and abatacept (NCT03632187) are adjusting their trial treatment based on PMR-AS values. Therefore, considering this extended use, face validity may be inferred. However, so far, measurement properties of the PMR-AS have not been systematically scrutinized and summarized.⁶

Our objective was to systematically assess the measurement properties of the PMR-AS in available clinical studies in order to ultimately summarize reliability and validity for future clinical trials and clinical care.

METHODS

Guided by the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) taxonomy and methodology for systematic reviews, we chose to assess internal structure (by assessing model type, structural validity, and internal consistency), reliability (by assessing inter- and intrarater reliability and measurement error), and validity (by testing hypotheses for construct validity and responsiveness) of the PMR-AS

This is an investigator-initiated and -funded study.

¹T.E. Bolhuis, MD, A.A. den Broeder, PhD, C.H.M. van den Ende, MSc, PhD, Department of Rheumatology, Sint Maartenskliniek, Ubbergen, and Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, the Netherlands; ²L.E.A. Nizet, A. van der Maas, MD, PhD, Department of Rheumatology, Sint Maartenskliniek, Ubbergen, the Netherlands; ³C. Owen, MBBS, PhD, Department of Rheumatology, Austin Health, Melbourne, Australia.

The authors declare no conflicts of interest relevant to this article.

Address correspondence to Dr. T.E. Bolhuis, Sint Maartenskliniek, Department of Rheumatology, PO Box 9011, 6500GM, Nijmegen, the Netherlands. Email: t.bolhuis@maartenskliniek.nl.

Accepted for publication February 16, 2022.

in published studies with patients with PMR.⁷ The review protocol and an amendment (which excluded assessment of PMR-AS–based relapse criteria) were registered in PROSPERO (ID 187907). The Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines were used for reporting,⁸ as shown in Supplementary Material 1 (available from the authors upon request).

We first formulated hypotheses to test for construct validity and responsiveness, as shown in detail in Supplementary Material 2 (available from the authors upon request). To increase the relevance of these hypotheses, the team comprised rheumatologists with both PMR and research experience (CO, AAdB, AvdM) and researchers with experience in systematic reviews (AAdB, CHMvdE). To test construct validity, we hypothesized that increased PMR disease activity is generally accompanied by increased erythrocyte sedimentation rate (ESR), decreased physical function, and decreased quality of life (QOL), and therefore, the PMR-AS is (relatively) highly correlated with these. Further, we expected patients with an ultrasound (US) indicative of active PMR, or with clinical relapse or active disease, to have a significantly higher PMR-AS than those without. To test responsiveness, we hypothesized that PMR disease activity responds substantially (and quickly) to GC treatment, and an increase in disease activity over time may be accompanied by an increase in ESR or a decrease in physical function; therefore, change in PMR-AS is moderately correlated with change in ESR and change in physical function.

Search. PubMed/MEDLINE, EMBASE, and CINAHL were searched in May 2020 and again in May 2021, using a search strategy with variants of PMR (population) and PMR-AS (outcome) in title, abstract, or full text (Supplementary Material 3, available from the authors upon request). Further, investigator knowledge and the references of included articles were used to search for additional publications. Records were stored in a Mendeley database and duplicate records were removed.

Study selection. First, record titles and abstracts were screened by 2 independent investigators (TEB, LEAN) using Rayyan.⁹ Studies were eligible if they were in English, published in a peer-reviewed journal, had a controlled or noninterventional design, and included ≥ 5 patients diagnosed with PMR discernible from giant cell arteritis (GCA). English was chosen as the single language, as authors are not experienced enough in other languages and the PMR-AS is predominantly used in countries using English as lingua franca. PMR diagnosis could be based on either recognized classification criteria or clinical diagnosis.^{10,11,12,13,14} To increase sensitivity, articles included by ≥ 1 investigator, or with uncertainty regarding eligibility, were included for the next phase.

Second, full-text articles were assessed by 2 investigators (TEB, LEAN) and reasons for exclusion were hierarchically recorded as (1) no full text available, (2) not available in English, (3) not presenting original data (eg, reviews), (4) < 5 discernible PMR patients without GCA or not distinguishing between PMR and GCA, and (5) not using the PMR-AS as an instrument. Disagreement was resolved through consensus. The systematic reviews by Duarte et al and Dejaco et al were used as an additional check to confirm whether any relevant articles were missed.^{6,15} We did not apply snowballing, but we did contact authors regarding potential data on measurement error and reliability.

Third, the same 2 investigators assessed whether internal structure, reliability, or validity (by testing hypotheses) could be investigated with an article based on COSMIN criteria.^{7,16} Articles not providing information or data on these properties of interest were excluded. Additionally, multiple articles regarding the same study (data) were discussed, and articles not providing additional data on a measurement property were excluded.

Data extraction.

• **Study characteristics** — Standardized data extraction tables were established a priori through discussion and were based on COSMIN recommendations.⁷ Study characteristics included design, participants (number, inclusion and exclusion criteria), treatment, and follow-up duration. For studies with multiple study samples (eg, multiple cohorts for a validation study or different treatment groups in a trial) characteristics regarding each

subsample were extracted. Study characteristics were extracted by 2 investigators (TEB, LEAN) and disagreement was resolved through discussion and consensus.

• **Summary of measurement properties** — Methodological quality, measurement property results, and whether these properties met criteria for good measurement properties were described in COSMIN-based Summary of Measurement Properties (SOMP) tables. Methodological quality of each study was evaluated for the property it was used to assess specifically. For example, methodologic quality for a study used to assess construct validity was evaluated using different criteria than a study used to assess internal consistency. Methodologic assessment was based on the COSMIN risk of bias checklist, which uses a 4-point rating scale (ranging from inadequate to very good), with the lowest score corresponding to the overall quality regarding that property.⁷ Results of measurement properties were extracted (eg, correlations) and evaluated according to the COSMIN criteria for good measurement properties. Our hypotheses for construct validity and responsiveness are shown in more detail in Supplementary Material 2 (available from the authors upon request).⁷ The overall SOMP tables were evaluated using COSMIN recommendations for good measurement properties (eg, 75% of construct validity hypotheses should be met to accept construct validity).

Statement of ethics and consent. The paper has been seen and approved by all authors. They have given necessary attention to ensure the integrity of the work and are all meeting the ICMJE criteria for authorship.

RESULTS

Studies included. Our search resulted in 1052 records, with 853 remaining after removal of duplicates; there were 26 articles with a controlled or noninterventional design and using the PMR-AS as an outcome (Figure 1).^{5,17–41} Of note, an article by Björkman et al did not use the entire PMR-AS (VASph was left out) and was therefore excluded.⁴¹ Further, an article by Leeb et al on PMR response criteria was included, since data regarding this composite score could be used to assess structural validity and internal consistency of the PMR-AS.¹⁷ One article not included through the search strategy, as it did not mention the PMR-AS in the title, abstract, or full text, but mentioned “Polymyalgia Rheumatica Activity Score” instead, was added based on investigator knowledge.⁴²

From the 26 articles using the PMR-AS, we could use 12 articles to assess measurement properties of interest.^{5,17–22,24,25,29,35,42} Eight articles could not be used to assess measurement properties at all. Six articles did not provide additional data regarding measurement properties and were excluded. From the 2 articles by McCarthy and colleagues, using the same study population, the publication from 2014 was chosen because it also included additional results of patient-reported outcome measures.^{25,26} From the 2 articles by Kreiner and colleagues, both using the same study data, the publication from 2010 was chosen due to display of treatment group results.^{22,23} From 5 articles on the TENOR (Tolerance and Efficacy of tocilizumab in pOlymyalgia Rheumatica) study,^{27,29,33,38,39} we used the article by Devauchelle-Pensec²⁹ due to the longitudinal display of PMR-AS results.

Study characteristics. The 12 articles (with 12 study populations) used to assess measurement properties were mostly prospective (83%) and noninterventional (66%) in design (Table 1).^{5,17–22,24,25,29,35,42} Inclusion criteria varied from clinical (rheumatologist) diagnosis to formal classification criteria based on the

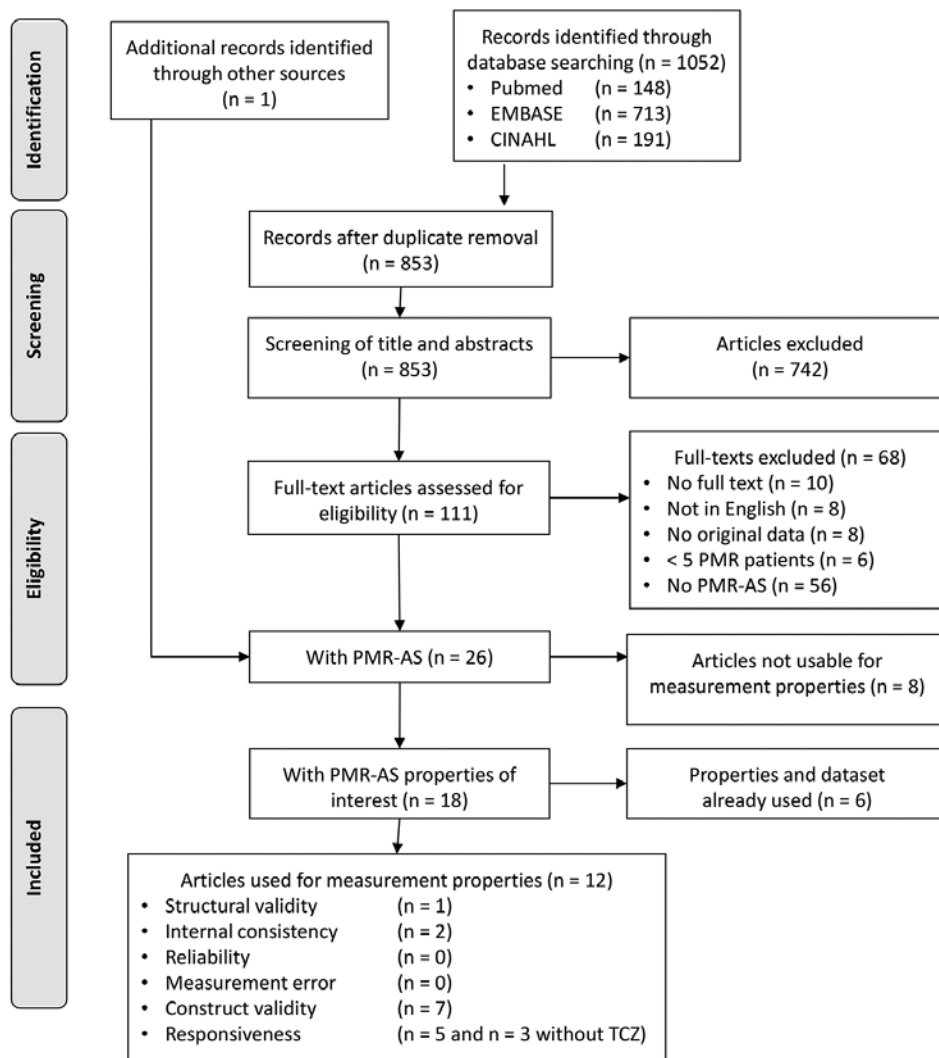


Figure 1. PRISMA 2009 adapted flow chart. PMR-AS: Polymyalgia Rheumatica Activity Score; PRISMA: Preferred Reporting Items for Systematic reviews and Meta-Analyses.

European Alliance of Associations for Rheumatology/American College of Rheumatology,¹⁴ Bird,¹⁰ Jones and Hazleman,¹¹ Chuang,¹² and Healey¹³ criteria. Exclusion criteria varied as well, from solely exclusion of GCA to exclusion of a range of disease with symptoms that may mimic PMR symptoms.

Summary of measurement properties.

- Structural validity and internal consistency — Although some evidence was found on structural validity and internal consistency, it was insufficient based on COSMIN criteria. We found no (theoretical) background information on internal structure of the PMR-AS and, therefore, we could not formally conclude whether the PMR-AS was developed as a formative or reflective instrument.^{43,44} Consequently, information on structural validity and internal consistency was assessed and summarized (Table 2).

One article by Leeb et al assessed structural validity and mentioned that the PMR-AS is a monodimensional instrument.¹⁸ Unfortunately, no information was provided on how dimensionality was assessed (eg, which kind of factor analysis

was used) and what the exact results were (eg, what the resulting comparative fit index was). Based on COSMIN criteria, we therefore considered methodology inadequate and exact (numerical) results unknown.^{7,43} Two articles, with a total of 3 study populations, assessed internal consistency, with standardized Cronbach α of 0.91, 0.88, and 0.81 (Table 2).^{5,18} However, considering that internal consistency requires consideration and acceptable unidimensionality of each separate subscale based on COSMIN criteria, methodological quality may be considered doubtful, and this property may not be accepted yet.

- Reliability and measurement error — No information was available on inter- and intrarater reliability and measurement error of the PMR-AS as a continuous measurement instrument for disease activity.

- Construct validity — Five out of 7 hypotheses concerning construct validity could be tested, from which 3 (60%) were confirmed with relatively limited sample sizes and adequate overall methodological quality (Table 3). These confirmed hypotheses included the association between the PMR-AS and

Table 1. Study characteristics.

Author	Year	Review Population	Design	Participants				
				Inclusion	Exclusion	n	Follow-up Duration	Treatment
Chino et al ³⁵	2019	1	OT	EULAR/ACR ¹⁴	GCA	13	104 w	TCZ + GC
Devauchelle-Pensec et al ²⁹	2016	2	OT	Chuang et al ¹²	GCA, uncontrolled dyslipidemia, or cardiovascular disease; chronic infection; hydroxyapatite crystal disease or chondrocalcinosis; severe OA of the hip or shoulder; symmetrical peripheral arthritis; active TD and drug-related myalgia	20	12 m	TCZ for 12 w followed by GC based on PMR-AS
McCarthy et al ²⁵	2014	3	PC	New; Jones and Hazleman ¹¹	RF or ACPA, CTD, systemic infection, abnormal CK or TSH, suspected malignancy	25	6 w	15 mg GC 6 w
				Asymptomatic; Jones and Hazleman ¹¹	RF or ACPA, CTD, systemic infection, abnormal CK or TSH, suspected malignancy	35	6 w	< 5 mg GC
Do-Nguyen et al ⁴²	2013	4	RC	RJ	GCA	135	12 m	Discretion ^a
Cleuziou et al ²⁴	2012	5	PC	RJ	Mimicking diseases per RJ	89	NS	GC
Kreiner and Galbo ²²	2010	6	RCT	Chuang et al ¹²	IST, GCA, systemic infections, hepatitis B/C, TB, uncontrolled DM or HT, severe HF, other inflammatory diseases, cancer in the past 5 years, NMD, TD, calcium disturbance	10 ^b	2 w	ETN
			RCT	Chuang et al ¹²	IST, GCA, systemic infections, hepatitis B/C, TB, uncontrolled DM or HT, severe HF, other inflammatory diseases, cancer in the past 5 years, NMD, TD, calcium disturbance	10 ^b	2 w	Placebo
Macchioni et al ²¹	2009	7	PC	RJ	GCA, prior GC treatment	57	NS	GC
Binard et al ²⁰	2008	5			c			
Catanoso et al ¹⁹	2007	8	OT	Healey ¹³ + > 12 m + > 7.5 mg GC	GCA, RA, uncontrolled DM or HT, systemic infection, malignancies, TB	6	36 w	ETN + GC
Leeb et al ¹⁸	2007	9	CS	Bird ¹⁰	GCA, RA	78	NS	GC ^d
		10	RC	Bird ¹⁰	GCA, RA	39	NS	GC
Leeb & Bird ⁵	2004	11	PC ^e	RJ	RA, shoulder capsulitis, cervical spondylosis, shoulder OA, polymyositis, thyrotoxicosis, myopathies, SLE, PAN, dermatomyositis, multiple myeloma, occult carcinoma, Parkinson disease	57	24 w	Discretion ^a
		12	PC ^f	RJ	RA, shoulder capsulitis, cervical spondylosis, shoulder OA, polymyositis, thyrotoxicosis, myopathies, SLE, PAN, dermatomyositis, multiple myeloma, occult carcinoma, Parkinson disease	24	33.6 w ^g	Discretion ^a
Leeb et al ¹⁷	2003	11			c			
		12			c			

^a Per discretion of investigator, rheumatologist, or treating physician. ^b Displayed are the number of PMR participants, without healthy controls. ^c Denotes if study uses the same study population as a previously mentioned review population. ^d Sixty-nine out of 78 patients were treated with GC. ^e European patient cohort. ^f Lower Austrian cohort. ^g Mean follow-up duration. ACPA: anticitrullinated protein antibodies; ACR: American College of Rheumatology; CK: creatine kinase; CS: cross-sectional; CTD: connective tissue disease; DM: diabetes mellitus; ETN: etanercept; EULAR: European Alliance of Associations for Rheumatology; GC: glucocorticoids; GCA: giant cell arteritis; HF: heart failure; HT: hypertension; IST: immunosuppressive therapy; m: months; NMD: neuromuscular disease; NS: not specified; OA: osteoarthritis; OT: open-label trial; PAN: polyarteritis nodosa; PC: prospective cohort; PMR-AS: Polymyalgia Rheumatica Activity Score; RA: rheumatoid arthritis; RC: retrospective cohort; RCT: randomized controlled trial; RF: rheumatoid factor; RJ: per rheumatologist judgment; SLE: systemic lupus erythematosus; TB: tuberculosis; TCZ: tocilizumab; TD: thyroid disease; TSH: thyroid-stimulating hormone; w: weeks.

Table 2. Summary of measurement properties: internal consistency and structural validity.

Author, Year	Structural Validity			Internal Consistency		
	n	Methodological Quality	Result	n	Methodological Quality	Rating
Leeb & Bird, 2004 ⁵				57 ^a	Doubtful	+ ^b
				24 ^c	Doubtful	+ ^d
Leeb et al, 2007 ¹⁸	39	Inadequate	Unknown ^e	39	Doubtful	+ ^f

Results (rating) are displayed as + and – and are either in line or not in line with criteria for good measurement properties, respectively. ^a Concerns the European patient cohort. ^b Cronbach α 0.91. ^c Concerns the Lower Austrian cohort. ^d Cronbach α 0.88. ^e Instead of exact numerical results (eg, comparative fit index) on structural validity, the following statement was given: “Factor analysis by principal component analysis demonstrated that the PMR-AS constitutes a monodimensional instrument.”¹⁸ ^f Cronbach α 0.81. n: no. of participants from study that were used to assess property; PMR-AS: Polymyalgia Rheumatica Activity Score.

physical function, QOL, and clinical disease state.^{5,18,20,22,24,25} The 2 hypotheses that were not confirmed explored the association between the PMR-AS and ESR, and US findings indicative of PMR.^{5,18,21} As a note, we used patient data provided by Catanoso et al to calculate correlations, although these were nonsignificant with a sample size of only 6.¹⁹ Thus, the construct validity could not be confirmed as the required 75% threshold for COSMIN construct validity was not met.⁷

• Responsiveness — All 3 hypotheses concerning responsiveness could be tested, and 2 (67%) were confirmed, although sample sizes were inadequate and methodological quality was doubtful at times (Table 4). The 2 hypotheses that were confirmed explored the change in PMR-AS following GC treatment (increase) and the association between change in PMR-AS and physical function.^{5,42} In addition, there were 2 open-label trials on tocilizumab,^{29,35} which could be used to explore overall

Table 3. Summary of measurement properties: construct validity.

Hypothesis	Author, Year	n	Methodological Quality	Rating ^a
ESR, $r_s \geq 0.60$	Leeb & Bird, 2004 ⁵	53	Adequate	– ^b
	Leeb et al, 2007 ¹⁸	78	Adequate	– ^c
	Catanoso et al, 2007 ^{19*}	6	Very good	– ^d
Physical function, $r_s \geq 0.60$	McCarthy et al, 2014 ²⁵	60	Very good	+ ^e
QOL, $r_s \geq 0.60$	Leeb & Bird, 2004 ⁵	53	Adequate	+ ^f
	Leeb et al, 2007 ¹⁸	78	Adequate	+ ^g
	McCarthy et al, 2014 ²⁵	60	Very good	+ ^h
Disease states ⁱ	Binard et al, 2008 ²⁰	89 ^j	Very good	+ ^k
	Kreiner and Galbo, 2010 ²²	40 ^l	Very good	+ ^m
	Cleuziou et al, 2012 ²⁴	89	Inadequate	+ ⁿ
US ⁱ	Macchioni et al, 2009 ²¹	44 ^o	Very good	– ^p

* Since analyses were not performed, but data were provided, a Spearman rank correlation was calculated by the research team (using Stata v13.1 for Windows, StataCorp). ^a Results are displayed as + and – and are either in line or not in line with criteria for good measurement properties, respectively. ^b Correlation between PMR-AS and ESR ($r_s = 0.48$; $P = 0.001$). ^c Correlation between PMR-AS and ESR ($r_s = 0.38$; $P = 0.01$). ^d Correlation between PMR-AS and ESR ($r_s = -0.23$, $P = 0.66$; and $r_s = 0.03$, $P = 0.96$ for 0 and 6 months, respectively). ^e Correlation between PMR-AS and HAQ ($r_s = 0.68$; $P < 0.001$). ^f Correlation between PMR-AS and VAS patient global ($r_s = 0.75$; $P = 0.0001$). ^g Correlation between PMR-AS and VAS patient global ($r_s = 0.76$; $P = 0.001$). ^h Correlation between PMR-AS and VAS QOL ($r_s = 0.80$; $P < 0.001$). ⁱ Patients with clinically defined active disease or inflammation suggestive of PMR should have a significantly higher PMR-AS than patients without. ^j Comparison is between 60 and 22 first patient visits with and without a flare, respectively, and 108 and 29 overall visits with and without a flare, respectively (with flare based on clinical judgment). ^k PMR-AS was significantly higher in a flare group when compared with a group without flares ($P = 0.0001$). ^l Comparison is between 20 patients with PMR and 20 HCs. ^m PMR-AS was significantly higher in patients with active PMR, than in HCs ($P < 0.0001$). ⁿ Correlation between PMR-AS and prednisone dosage change ($r_s = 0.58$; $P < 0.001$). ^o Comparison is between 26 patients with persistent inflammation and 18 patients with no evidence of inflammation on US. ^p No significant difference was found when comparing patients with vs without persistent inflammation on US. ESR: erythrocyte sedimentation rate; HAQ: Health Assessment Questionnaire; HC: healthy control; n: no. of participants per study per hypothesis; PMR: polymyalgia rheumatica; PMR-AS: PMR Activity Score; QOL: quality of life; US: ultrasound; VAS: visual analog scale.

Table 4. Summary of measurement properties: responsiveness.

Hypothesis	Author, Year	n	Methodological Quality	Rating ^a
GC response ^b	Leeb & Bird, 2004 ⁵	57 ^c	Doubtful	+ ^d
	Leeb & Bird, 2004 ⁵	24 ^e	Doubtful	+ ^f
	Do-Nguyen et al, 2013 ⁴²	135	Inadequate	? ^g
TCZ response ^h	Devauchelle-Pensec et al, 2016 ²⁹	20	Very good	+ ⁱ
	Chino et al, 2019 ³⁵	13	Very good	+ ^j
ESR, $r_s \geq 0.50$	Catanoso et al, 2007 ^{19,k}	6	Very good	- ^l
Physical function, $r_s \geq 0.50$	Catanoso et al, 2007 ^{19,k}	6	Very good	+ ^m

^a Results are displayed as + and – and are either in line or not in line with criteria for good measurement properties, respectively. ^b PMR-AS should decrease by at least 70% in 50% of patients after 4 weeks of either starting GC treatment or increasing GC dose by 5 mg. ^c Concerning the European patient cohort. ^d PMR-AS mean (SD) at baseline was 27.5 (12.5) and changed to 6.0 (4.7) at week 4 in the international cohort, corresponding to an average decrease of 78%. ^e Concerning the Lower Austrian cohort. ^f In the Lower Austrian cohort PMR-AS, mean (SD) at baseline was 29.1 (6.5) and changed to 9.0 (3.8) at week 4, corresponding to an average decrease of 69%. ^g The exact change, or percentage of change, in PMR-AS is not given. However, the article stated that 96% of patients showed a clinical response (and 67% a complete response) and that response was generally rapid (within days or weeks). ^h The response to TCZ monotherapy was not included as a hypothesis, and therefore was assessed separately from GC responsiveness. ⁱ PMR-AS median (IQR) decreased from 36.6 (30.4–43.8) at baseline to 11.0 (7.9–19.3) at week 4, corresponding to an average decrease of 70%. ^j The average PMR-AS decreased from 33.3 at baseline to 13.0 at week 4, corresponding to an average decrease of 61%. ^k Since analyses were not performed, but data were provided, a Spearman rank correlation coefficient was calculated by the research team (using Stata v13.1 for Windows; StataCorp). ^l Correlation between Δ PMR-AS and Δ ESR ($r_s = 0.14$, $P = 0.79$). ^m Correlation between Δ PMR-AS and Δ HAQ ($r_s = 0.83$; $P = 0.04$). ESR: erythrocyte sedimentation rate; GC: glucocorticoid; HAQ: Health Assessment Questionnaire; n: no. of participants from study that were used to assess property; PMR-AS: Polymyalgia Rheumatica Activity Score; TCZ: tocilizumab.

treatment responsiveness, and which further supported overall responsiveness of the PMR-AS to treatment. The hypothesis that was not confirmed explored the association with ESR, although assessment of this hypothesis (and the hypothesis on association with change in Health Assessment Questionnaire–Disability Index [HAQ-DI]) was possible by calculating correlations with the data of 6 patients provided by Catanoso et al.¹⁹ As an additional note, patients from the study by Do-Nguyen et al were retrospectively included (between 1999 and 2003) prior to publication of the PMR-AS.⁴² However, no information was given about this period and how potentially missing measurements were handled; therefore, methodological quality was rated inadequate. In summary, evidence to support responsiveness is limited, as it is based on a relatively small number of studies with small sample sizes, including some of doubtful methodological quality.

DISCUSSION

We performed a COSMIN-based systematic review of quantitative studies to assess and summarize measurement properties of the PMR-AS. Our limited restrictions and use of articles without measurement property assessment as a goal resulted in a range of designs and study populations that could be used to assess measurement properties. There is some promising evidence, although insufficient, for construct validity and responsiveness, but no data were found on measurement error and reliability, and further assessment of internal structure may be necessary.

Although it is not clear from background information which model is applicable, a formative multidimensional model may

be appropriate for the PMR-AS as an instrument for measuring disease activity.^{5,18} If disease activity is measured using multiple different constructs (eg, pain and inflammation are different and nonexchangeable constructs), then a formative multidimensional model is applicable.⁴³ A formative model may be supported further by the statement by Leeb et al, noting that acute-phase reactants and MS were items for disease activity independent of pain, and consequently do not seem to be exchangeable constructs.¹⁷ If this model is indeed applicable, then structural validity and internal consistency are not applicable. However, further assessment would be necessary to draw this conclusion.

There may be several reasons why some construct validity and responsiveness hypotheses were not met, whereas others were. Hypotheses regarding physical function, QOL, and clinical disease state were confirmed, but those regarding ESR and US were not, reflecting that items of the PMR-AS may gravitate more toward symptoms, functioning, and health perception, as opposed to biological and physiological variables when looking at the Wilson and Cleary model.⁴⁵ As a matter of fact, the association between PMR-AS and physical function, as well as between PMR-AS and QOL, seems even higher than for some disease activity measures for other rheumatic disorders.^{46,47} Concerning US specifically, the criteria of Macchioni et al for identifying persistent inflammation on US may not properly distinguish PMR disease activity states due to limited sensitivity and specificity.^{21,48} Indeed, previous studies show a lack of correlation between US and separate PMR-AS components, except perhaps MS,^{49,50} although comparison may be hindered due to use of different US procedures. Further assessment of relation with

other—particularly physiological—instruments combined with larger study sizes for some hypotheses (eg, physical function) may be needed to meet criteria for good measurement properties.⁷

Contrary to the promising validity, however, the current lack of information regarding the reliability and measurement error of the PMR-AS has some negative implications for current use. For example, it may be that a previously proposed relapse criterion (eg, an increase in PMR-AS of 4.2) falls outside the smallest detectable change; thus, a relapse may not be detected or may be incorrectly diagnosed.^{7,51} Further, information on reliability is needed among others to determine sample size for trials based on the PMR-AS. Some reliability studies have been performed regarding specific items of the PMR-AS, but not for all components and not for the PMR-AS as a whole.⁵² Further, Binard et al performed a reliability study of PMR-AS relapse criteria, but unfortunately did not assess the scale as a whole.⁵³

When generalizing these results on measurement properties of the PMR-AS to clinical practice, an important note should be made. In the development article by Leeb et al, it is noted that the original objective of the PMR response criteria (and presumably the PMR-AS thereafter), was to establish criteria that could be used in future clinical trials.¹⁷ Consequently, a range of conditions which may mimic PMR (eg, GCA, RA, osteoarthritis [OA] and local bursitis/tendinitis) were excluded in the developmental studies by Leeb et al.^{5,17} However, these comorbidities may be quite prevalent in the elderly PMR population, as shown by Do-Nguyen et al in a retrospective cohort of 137 patients with PMR, out of whom 45 had OA; these comorbidities may interfere with PMR disease activity assessment.⁴² Therefore, performance of the PMR-AS in either clinical practice or trials of a pragmatic nature may differ due to concomitant disorders influencing PMR-AS items.

Some strengths and limitations should be noted regarding this review process. Although a broad search strategy without design or property filter was used, 1 article—which used more general terms for disease activity as opposed to PMR-AS—was not found in the search.⁴² Further, not all data and studies that used the PMR-AS were usable to assess measurement properties. A main reason studies could not be used was that outcomes key to this review were not reported; for example, McCarthy et al measured change in HAQ and PMR-AS but reported no correlation between these.²⁵ Another reason studies could not be used was that some correlations that were reported were not anticipated when hypotheses were drafted; for example, no hypotheses were formulated regarding fluorodeoxyglucose positron emission tomography scans, since their exact role in PMR is still unclear.⁴⁸ However, a broad range of hypotheses were formulated and amended during initial extraction of study characteristics to optimize the number of hypotheses that could be tested. Finally, since the nature and cut-offs of hypotheses formulated by the research team are inherently subjective, they might be different, or less stringent, for another team. However, to increase relevance and feasibility of hypotheses, our research team comprised both rheumatologists with PMR research experience and researchers with experience in systematic literature reviews.

All in all, the PMR-AS shows promise as a measurement instrument for PMR disease activity, although evidence on certain measurement properties is still limited or absent. Further, new measurement property validation methodologies have been introduced since the development and initial validation of the PMR-AS. Therefore, stepwise reassessment of reliability, validity, and other properties not included in this review, such as interpretability, may be useful to assess the potential role of PMR-AS as an outcome for trials, as well as its applicability for clinical treat-to-target strategies. However, considering our findings, the PMR-AS seems the most appropriate measure for PMR disease activity in clinical trials as of yet.

REFERENCES

1. González-Gay MA, Matteson EL, Castañeda S. Polymyalgia rheumatica. *Lancet* 2017;390:1700-12.
2. DeJaco C, Singh YP, Perel P, et al; European League Against Rheumatism, American College of Rheumatology. 2015 recommendations for the management of polymyalgia rheumatica: a European League Against Rheumatism/American College of Rheumatology collaborative initiative. *Ann Rheum Dis* 2015;74:1799-807.
3. Smolen JS, Breedveld FC, Burmester GR, et al. Treating rheumatoid arthritis to target: 2014 update of the recommendations of an international task force. *Ann Rheum Dis* 2016;75:3-15.
4. Smolen JS, Schöls M, Braun J, et al. Treating axial spondyloarthritis and peripheral spondyloarthritis, especially psoriatic arthritis, to target: 2017 update of recommendations by an international task force. *Ann Rheum Dis* 2018;77:3-17.
5. Leeb BF, Bird HA. A disease activity score for polymyalgia rheumatica. *Ann Rheum Dis* 2004;63:1279-83.
6. Duarte C, De Oliveira Ferreira RJ, Mackie SL, Kirwan JR, Pereira de Silva JA; OMERACT Polymyalgia Rheumatica Special Interest Group. Outcome measures in polymyalgia rheumatica. A systematic review. *J Rheumatol* 2015;42:2503-11.
7. Mokkink LB, Prinsen CAC, Patrick DL, et al. COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs). [Internet. Accessed March 3, 2022.] Available from: https://www.cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018-1.pdf
8. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009;339:b2700.
9. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. *Syst Rev* 2016;5:210.
10. Bird HA, Esselinckx W, Dixon AS, Mowat AG, Wood PH. An evaluation of criteria for polymyalgia rheumatica. *Ann Rheum Dis* 1979;38:434-9.
11. Jones JG, Hazleman BL. Prognosis and management of polymyalgia rheumatica. *Ann Rheum Dis* 1981;40:1-5.
12. Chuang TY, Hunder GG, Ilstrup DM, Kurland LT. Polymyalgia rheumatica: a 10-year epidemiologic and clinical study. *Ann Intern Med* 1982;97:672-80.
13. Healey LA. Long-term follow-up of polymyalgia rheumatica: evidence for synovitis. *Semin Arthritis Rheum* 1984;13:322-8.
14. Dasgupta B, Cimmino MA, Maradit-Kremers H, et al. 2012 provisional classification criteria for polymyalgia rheumatica: a European League Against Rheumatism/American College of Rheumatology collaborative initiative. *Ann Rheum Dis* 2012;71:484-92.
15. DeJaco C, Duftner C, Cimmino MA, et al; International Work Group for PMR and GCA. Definition of remission and relapse in

- polymyalgia rheumatica: data from a literature search compared with a Delphi-based expert consensus. *Ann Rheum Dis* 2011;70:447-53.
16. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737-45.
 17. Leeb BF, Bird HA, Nesher G, et al. EULAR response criteria for polymyalgia rheumatica: results of an initiative of the European Collaborating Polymyalgia Rheumatica Group (subcommittee of ESCISIT). *Ann Rheum Dis* 2003;62:1189-94.
 18. Leeb BF, Rintelen B, Sautner J, Fassl C, Bird HA. The polymyalgia rheumatica activity score in daily use: proposal for a definition of remission. *Arthritis Rheum* 2007;57:810-5.
 19. Catanoso MG, Macchioni P, Boiardi L, Pipitone N, Salvarani C. Treatment of refractory polymyalgia rheumatica with etanercept: an open pilot study. *Arthritis Rheum* 2007;57:1514-9.
 20. Binard A, de Bandt M, Berthelot JM, Saraux A; Inflammatory Joint Disease Working Group of the French Society for Rheumatology. Performance of the polymyalgia rheumatica activity score for diagnosing disease flares. *Arthritis Rheum* 2008;59:263-9.
 21. Macchioni P, Catanoso MG, Pipitone N, Boiardi L, Salvarani C. Longitudinal examination with shoulder ultrasound of patients with polymyalgia rheumatica. *Rheumatology* 2009;48:1566-9.
 22. Kreiner F, Galbo H. Effect of etanercept in polymyalgia rheumatica: a randomized controlled trial. *Arthritis Res Ther* 2010;12:R176.
 23. Kreiner FF. Activity of the neuroendocrine axes in patients with polymyalgia rheumatica before and after TNF- α blocking etanercept treatment. *Arthritis Res Ther* 2012;14:R186.
 24. Cleuziou C, Binard A, De Bandt M, Berthelot JM, Saraux A. Contribution of the polymyalgia rheumatica activity score to glucocorticoid dosage adjustment in everyday practice. *J Rheumatol* 2012;39:310-3.
 25. McCarthy EM, MacMullan PA, Al-Mudhaffer S, et al. Plasma fibrinogen along with patient-reported outcome measures enhances management of polymyalgia rheumatica: a prospective study. *J Rheumatol* 2014;41:931-7.
 26. McCarthy EM, MacMullan PA, Al-Mudhaffer S, et al. Plasma fibrinogen is an accurate marker of disease activity in patients with polymyalgia rheumatica. *Rheumatology* 2013;52:465-71.
 27. Palard-Novello X, Querellou S, Gouillou M, et al. Value of (18) F-FDG PET/CT for therapeutic assessment of patients with polymyalgia rheumatica receiving tocilizumab as first-line treatment. *Eur J Nucl Med Mol Imaging* 2016;43:773-9.
 28. Mori S, Koga Y. Glucocorticoid-resistant polymyalgia rheumatica: pretreatment characteristics and tocilizumab therapy. *Clin Rheumatol* 2016;35:1367-75.
 29. Devauchelle-Pensec V, Berthelot JM, Cornec D, et al. Efficacy of first-line tocilizumab therapy in early polymyalgia rheumatica: a prospective longitudinal study. *Ann Rheum Dis* 2016;75:1506-10.
 30. Devauchelle-Pensec V, Saraux L, Berthelot JM, et al. Assessing polymyalgia rheumatica activity when C-reactive protein is unavailable or uninterpretable. *Rheumatology* 2018;57:666-70.
 31. Lally L, Forbess L, Hatzis C, Spiera R. Brief report: a prospective open-label phase IIa trial of tocilizumab in the treatment of polymyalgia rheumatica. *Arthritis Rheumatol* 2016;68:2550-4.
 32. Manzo C, Natale M. Relapse of polymyalgia rheumatica after a fall. *Reumatologia* 2017;55:251-5.
 33. Carvajal Alegria G, Devauchelle-Pensec V, Renaudineau Y, Saraux A, Pers JO, Cornec D. Correction of abnormal B-cell subset distribution by interleukin-6 receptor blockade in polymyalgia rheumatica. *Rheumatology* 2017;56:1401-6.
 34. Huwart A, Garrigues F, Jousse-Joulin S, et al. Ultrasonography and magnetic resonance imaging changes in patients with polymyalgia rheumatica treated by tocilizumab. *Arthritis Res Ther* 2018;20:11.
 35. Chino K, Kondo T, Sakai R, et al. Tocilizumab monotherapy for polymyalgia rheumatica: a prospective, single-center, open-label study. *Int J Rheum Dis* 2019;22:2151-7.
 36. Nakajima S, Chiba A, Makiyama A, et al. Association of mucosal-associated invariant T cells with different disease phases of polymyalgia rheumatica. *Rheumatology* 2020;59:2939-46.
 37. Owen CE, Poon AMT, Yang V, et al. Abnormalities at three musculoskeletal sites on whole-body positron emission tomography/computed tomography can diagnose polymyalgia rheumatica with high sensitivity and specificity. *Eur J Nucl Med Mol Imaging* 2020;47:2461-8.
 38. Carvajal Alegria G, Cornec DYK, Renaudineau Y, Saraux A, Devauchelle-Pensec V. Inflammatory markers are quickly improved by tocilizumab in early polymyalgia rheumatica and might predict early response to interleukin-6 blockade. *Rheumatol Ther* 2021;8:751-60.
 39. Carvajal Alegria G, Garrigues F, Bettacchioli E, et al. Tocilizumab controls bone turnover in early polymyalgia rheumatica. *Joint Bone Spine* 2021;88:105117.
 40. Owen CE, McMaster C, Liew DFL, Leung JL, Scott AM, Buchanan RRC. Neutrophil to lymphocyte ratio predicts glucocorticoid resistance in polymyalgia rheumatica. *Int J Rheum Dis* 2021;24:56-62.
 41. Björkman MP, Tilvis RS. Muscle functions in polymyalgia rheumatica and giant-cell arteritis. *Healthy Aging Clin Care Elder* 2010;2:1-8.
 42. Do-Nguyen D, Inderjeeth CA, Edelman J, Cheah P. Retrospective analysis of the clinical course of patients treated for polymyalgia. *Open Access Rheumatol* 2013;5:33-41.
 43. Fleuren BPI, van Amelsvoort L, Zijlstra FRH, de Grip A, Kant I. Handling the reflective-formative measurement conundrum: a practical illustration based on sustainable employability. *J Clin Epidemiol* 2018;103:71-81.
 44. Streiner DL. Being inconsistent about consistency: when coefficient alpha does and doesn't matter. *J Pers Assess* 2003;80:217-22.
 45. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *JAMA* 1995;273:59-65.
 46. Boyd TA, Bonner A, Thorne C, et al. The relationship between function and disease activity as measured by the HAQ and DAS28 varies over time and by rheumatoid factor status in early inflammatory arthritis (EIA). Results from the CATCH Cohort. *Open Rheumatol J* 2013;7:58-63.
 47. Castrejón I, Silva-Fernández L, Bombardier C, Carmona L. Clinical composite measures of disease activity for diagnosis and followup of undifferentiated peripheral inflammatory arthritis: a systematic review. *J Rheumatol Suppl* 2011;87:48-53.
 48. Mackie SL, Koduri G, Hill CL, et al. Accuracy of musculoskeletal imaging for the diagnosis of polymyalgia rheumatica: systematic review. *RMD Open* 2015;1:e000100.
 49. Jiménez-Palop M, Naredo E, Humbrado L, et al. Ultrasonographic monitoring of response to therapy in polymyalgia rheumatica. *Ann Rheum Dis* 2010;69:879-82.
 50. Miceli MC, Zoli A, Peluso G, Bosello S, Gremese E, Ferraccioli G. Baseline shoulder ultrasonography is not a predictive marker of response to glucocorticoids in patients with polymyalgia rheumatica: a 12-month followup study. *J Rheumatol* 2017;44:241-7.
 51. Binard A, De Bandt M, Berthelot JM, Saraux A; Inflammatory Joint Disease Working Group CRI of the French Society for Rheumatology. Usefulness of the disease activity scores for polymyalgia rheumatica for predicting glucocorticoid dose changes: a study of 243 scenarios. *Arthritis Rheum* 2007;57:481-6.
 52. Twohig H, Owen C, Muller S, et al. Outcomes Measured in polymyalgia rheumatica and measurement properties of instruments considered for the OMERACT core outcome set: a systematic review. *J Rheumatol* 2021;48:883-93.
 53. Binard A, Lefebvre B, De Bandt M, Berthelot JM, Saraux A; Club "Rhumatismes et Inflammation". Validity of the polymyalgia rheumatica activity score in primary care practice. *Ann Rheum Dis* 2009;68:541-5.