RMD Open
Rheumatic & Musculoskeletal Diseases

ORIGINAL RESEARCH

Instruments for measuring fatigue in people with rheumatic and musculoskeletal diseases: a systematic review of measurement properties

Eduardo José Ferreira Santos ,[1,2,3] Bayram Farisogullari ,[3,4] Katie Fishpool ,[5] George Young,[5] Coziana Ciurtin ,[6] Fiona Cramp,[5] Emmanuel Oghenetejiri Erhieyovwe ,[7] Gary J Macfarlane ,[8] Jen Pearson,[5] Emma Dures ,[5,9] Pedro M Machado [3,10,11]

To cite: Santos EJF, Farisogullari B, Fishpool K, *et al*. Instruments for measuring fatigue in people with rheumatic and musculoskeletal diseases: a systematic review of measurement properties. *RMD Open* 2025;**11**:e006079. doi:10.1136/rmdopen-2025-006079

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/rmdopen-2025-006079).

EJFS and BF contributed equally.

Received 1 July 2025
Accepted 23 September 2025

Check for updates

For numbered affiliations see end of article.

Correspondence to
Dr Pedro M Machado; p.machado@ucl.ac.uk

ABSTRACT

Objective To summarise the measurement properties of instruments used to assess fatigue in people with rheumatic and musculoskeletal diseases (RMDs).

Methods A systematic review (SR) of measurement properties was conducted in children, adolescents/young adults and adults with RMDs, following Joanna Briggs Institute and COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) guidelines. Searches were performed in Medline, Embase, CINAHL and Cochrane Library. Risk of bias assessment, data extraction and synthesis were conducted independently by two reviewers. Instruments were assessed according to Outcome Measures in Rheumatology (OMERACT) criteria.

Results Out of 16 657 records, 109 articles underwent full-text review, and 60 met inclusion criteria. These studies evaluated the psychometric properties of 27 instruments. Most studies focused on construct validity (54/60, 90%) and intermethod reliability (45/60, 75%), with an overall low risk of bias. In contrast, test–retest reliability (13/60, 21.7%) and responsiveness (14/60, 23.3%) were less frequently assessed, but also with an overall low risk of bias. Evidence regarding clinical trial discrimination and thresholds of meaningful change was limited or absent, indicating the need for further research in these domains. Only five instruments—the Functional Assessment of Chronic Illness Therapy (FACIT) Fatigue, the 36-Item Short Form Survey Instrument (SF-36) Vitality, the Bristol Rheumatoid Arthritis Fatigue Multi-Dimensional Questionnaire (BRAF-MDQ), the BRAF Numerical Rating Scales (BRAF-NRS) and the Fatigue Severity Scale (FSS)—were rated as valid, reliable and low risk of bias, fulfilling OMERACT endorsement criteria.

Conclusions This SR comprehensively supports the use of several well-validated instruments to assess fatigue, particularly FACIT-Fatigue, SF-36 Vitality, BRAF-MDQ, BRAF-NRS and FSS, in both clinical and research settings.

PROSPERO registration number CRD42024507112.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Fatigue is a recognised priority and unmet need among people with rheumatic and musculoskeletal diseases (RMDs). Recent evidence and European Alliance of Associations for Rheumatology recommendations highlight the urgent need for guidance on appropriate measurement instruments for fatigue.

WHAT THIS STUDY ADDS

⇒ This review summarises the current evidence on the measurement properties of instruments used to assess fatigue in people with RMDs.
⇒ It confirms the importance of fatigue assessment and identifies 27 instruments, of which only five were fully endorsed based on their psychometric robustness.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Clinicians and researchers are advised to use one of fully endorsed instruments: Functional Assessment of Chronic Illness Therapy-Fatigue, 36-Item Short Form Survey Vitality, Bristol Rheumatoid Arthritis Fatigue (BRAF) Multi-Dimensional Questionnaire, BRAF-Numerical Rating Scales or Fatigue Severity Scale. These instruments demonstrate strong validity and reliability.
⇒ The review also highlights the need for further research into provisionally endorsed tools, particularly to address gaps in less-studied psychometric properties.

INTRODUCTION

Rheumatic and musculoskeletal diseases (RMDs) are commonly characterised by acute or chronic pain, significant disability and a substantial impact on quality of life, often limiting individuals' ability to function in society.[1] An estimated 1.71 billion people are affected by these conditions globally,[2] making them the leading cause of chronic pain and disability worldwide. They include a range of long-term inflammatory and non-inflammatory disorders, with low back pain, neck pain, osteoarthritis (OA) and rheumatoid arthritis (RA) among the most disabling.[3]

Fatigue is a frequent and often one of the most debilitating symptoms experienced by individuals with RMDs across the life course. It is pervasive, invisible, unpredictable[4 5] and affects approximately 40-80% of people with RMDs,[6 7] with about half reporting severe fatigue.[6 8–10] Fatigue is described as being overwhelming, distressing, extreme, not restored by sleep, multidimensional and part of a vicious circle where fatigue feeds and fosters itself.[11–13] Despite its burden, no universally accepted definition exists.[14] Fatigue also significantly affects children and adolescents with RMDs, with up to 75% experiencing it. Of those, around 25% experience disabling fatigue with major consequences for their performance at school, social life, sports and hobbies.[15 16]

Over the past two decades, fatigue has gained increasing recognition in rheumatology research.[17] In 2007, an international consensus recommended its inclusion in all RA clinical trials,[18] and delegates at the Outcome Measures in Rheumatology (OMERACT) eighth meeting endorsed fatigue as an addition to the 'core set', and therefore a core outcome for future RA studies.[18] However, its assessment remains challenging due to its subjective nature and close ties to emotional and cognitive domains.[19] Furthermore, condition-specific characteristics may impose the need for tailored assessment tools. Generic instruments, for instance, may conflate fatigue with limitations caused by inflammation or physical disability, particularly in inflammatory arthritis, potentially yielding unreliable or misleading results.[19]

A preliminary search of PROSPERO, Medline (PubMed), CINAHL (EBSCO) and Joanna Briggs Institute (JBI) Evidence Synthesis on 22 October 2023 identified several relevant reviews.[4 20–24] While these highlight the importance and diversity of fatigue measurement instruments, they have notable limitations: most focus on a single condition, rely on general reviews rather than systematic evaluations and lack robust assessment of psychometric properties. Consequently, they offer limited guidance for selecting appropriate tools. The recently published European Alliance of Associations for Rheumatology recommendations for management of fatigue in people with inflammatory RMDs also underscored the urgent need to identify suitable instruments, highlighting this gap in their research agenda and calling for standardisation of fatigue measurement in both research and clinical practice.[17]

In response, we conducted a systematic review (SR) of measurement properties to identify and appraise instruments for assessing fatigue in people with RMDs, with the aim of recommending the most appropriate tools for use in both research and clinical settings.

METHODS

This SR was conducted according to the JBI[25] and the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN)[26] approach, and reported following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines.[27] The SR protocol was established and registered in PROSPERO (CRD42024507112).

The research question was 'What are the measurement properties of instruments for evaluating fatigue in people with RMDs across the lifespan?'.

This question was formulated and structured using the 'Population, Instruments and Construct, Outcome, Type of study (PICOT)' format, as follows:

Population

A study was eligible for inclusion if the included participants were children, adolescents/young adults or adults with a rheumatic or musculoskeletal (MSK) disease, specifically, RA, axial spondyloarthritis (axSpA), peripheral spondyloarthritis, psoriatic arthritis (PsA), gout, systemic lupus erythematosus (SLE), systemic sclerosis (SSc), Sjögren's syndrome (recently renamed Sjögren's disease), idiopathic inflammatory myopathy (IIM, namely dermatomyositis, polymyositis, immune-mediated necrotising myopathy, antisynthetase syndrome, inclusion body myositis) and primary systemic vasculitis (large-vessel vasculitis: giant cell arteritis (and the related condition polymyalgia rheumatica), Takayasu's arteritis; medium-vessel vasculitis: polyarteritis nodosa; small-vessel vasculitis limited to the Anti-Neutrophil Cytoplasmic Antibody (ANCA)-associated vasculitis: granulomatosis with polyangiitis (GPA, previously named Wegener's granulomatosis), microscopic polyangiitis and eosinophilic GPA (previously named Churg-Strauss); and variable-vessel vasculitis: Behçet syndrome, also known as Behçet disease); OA, fibromyalgia (FM), chronic non-specific low back pain, hypermobility syndromes including Ehlers-Danlos syndrome and juvenile idiopathic arthritis (JIA). Only studies in which individuals were formally diagnosed with the condition or who satisfied internationally accepted disease classification criteria were included to maximise accuracy.[28–31] Whenever possible, information related to the participants' condition was stratified by age.

Instruments and construct

Regarding the construct and the instruments to measure it, the core concept was fatigue, which is a complex and multifaceted phenomenon. Fatigue associated with chronic diseases, unlike everyday fatigue, has no obvious 'explanation', does not improve with rest, varies in severity, is unpredictable and integrates the experience of profound or overwhelming physical and/or mental exhaustion.[32] In that sense, we took a broad approach and investigated any (semi-) quantitative self-reported fatigue score used to evaluate MSK fatigue. In adult populations, this could include instruments such as the Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F),[33] Rheumatoid Arthritis Impact of Disease-Fatigue (RAID-F),[34 35] Fatigue-Visual Analogue Scale (VAS),[36] 36-Item Short Form Survey (SF-36) vitality scale,[37] the Multidimensional Assessment of Fatigue (MAF),[38] Profile of Mood States-subscale fatigue,[39] Checklist Individual Strength,[40] Bristol Rheumatoid Arthritis Fatigue Multi-Dimensional Questionnaire (BRAF MDQ),[41 42] BRAF Numerical Rating Scales for severity, effect and coping (BRAF NRS),[41 42] among others. In paediatric populations, this could include instruments such as the Paediatric Quality of Life Multidimensional Fatigue Scale (PEDsQL-MFS)[43] and the Kids Fatigue Severity Scale (K-FSS), among others.

Outcome (measurement properties)

All the psychometric properties of the measurement instruments evaluated in the original articles were included in the SR.

Reliability (ie, the degree to which the instrument is free from measurement error) can include the assessment of internal consistency, intraobserver and interobserver agreement, and measurement error. Reliability can be reported as either Cronbach's alpha and McDonald's omega coefficients, intraclass correlation coefficients (ICC), weighted or unweighted Kappa statistics, SE of measurement (SEM), limits of agreement, smallest detectable change or concordance correlation coefficients.[25 26]

Validity (ie, the degree to which the instrument measures the construct it purports to measure) can include the assessment of content validity, face validity, structural validity, hypothesis testing, cross-cultural validity and criterion-related validity. Content validity can be evaluated by relevant items for the construct (eg, Content Validity Index), purpose, target population, the comprehensiveness of the instrument and floor or ceiling effects.[25 26] Structural validity can be evaluated by factor analysis and measures are the comparative fit index, Tucker-Lewis index, root mean square error of approximation and standardised root mean residuals, among others goodness-of-fit statistics.[25 26] The differential item functioning could be reported for cross-cultural validity. Criterion and construct validity measures are typically reported as either correlations, area under receiver operating curves (ROCs) or as sensitivity and specificity.

Responsiveness (ie, the ability of the instrument to detect change over time in the construct to be measured)[25] can be reported as either absolute or relative correlations or differences of the change scores, area under the ROC, sensitivity and specificity or standardised response mean (SRM).[25 26]

Type of study

Any quantitative (including experimental, quasi-experimental and cross-sectional) studies were included. Only studies focusing on the development and/or validation of measurement instruments were selected. Protocols, existing systematic and scoping reviews, meta-analyses, commentaries and editorials, other than psychometric data, were excluded. Additionally, studies that only used the measurement instrument as an outcome measure were also excluded. Studies duplicating the validation data of an instrument in a previous study were also excluded. For studies reporting the same analyses in a larger population than a previously reported population (partial duplication of data), only the most recent report was considered. Regarding the context, there were no constraints.

Search strategy and study selection

A search strategy was run by one of the authors (EJFS) in Medline through PubMed, Embase; CINAHL Complete and Cochrane Library (through EBSCO). The start date was the date of inception of the database, and the end date was 3 November 2023. Studies published in English, French, Portuguese, Spanish and Turkish language, with no restriction on the publication date, were considered for inclusion. Details on complete search strategies are provided in online supplemental material S1 and included COSMIN filters.[44]

All identified citations were uploaded into an EndNote X9 (Clarivate Analytics, Pennsylvania, USA) library and the duplicates removed. Titles and abstracts were screened by two independent reviewers (BF and EJFS) to assess eligibility criteria. The full articles were retrieved for all studies that met or had insufficient information to assess the inclusion criteria, and two reviewers (BF and EJFS) independently examined them in detail. Any disagreements between the reviewers were resolved through discussion or adjudication by a third reviewer (PMM). The study selection was performed using Rayyan.

Risk of bias (quality) assessment

Two reviewers (BF and EJFS) assessed the risk of bias in each included study using the COSMIN-OMERACT Good Metrics Checklist.[45] Quality assessment was rated using a colour code: 'Green' if good methods were used, 'Amber' if there were some methodological concerns but the data were acceptable for inclusion and 'Red' if there was a high risk of bias, as indicated by OMERACT. Any disagreements between the two reviewers (BF and EJFS) were resolved through discussion or adjudication by a third reviewer (PMM).

To evaluate the robustness of measurement properties, each study was assessed using the OMERACT provisional standards for adequate performance and assigned ratings of + (if positive support for the measurement property), ± (if ambivalent support, inconclusive) or − (if the instrument did not reach performance standards for that measurement property).[45]

The assessment of bias risk and the robustness of measurement properties were subsequently integrated into a summary ratings of individual measurement properties per domain (Summary of Measurement Properties—SOMP table). All studies rated amber or green were synthesised to generate an overall rating for the individual measurement properties for each domain based on the OMERACT Handbook—'Criteria for Final Rating'.

This summarises the quality and quantity of studies, the consistency of the results and the performance of individual instruments. A colour grading was used as recommended: GREEN indicates 'Good to go' (good methods were used in the studies, there are at least two studies, there are consistent findings and adequate performance in the property), RED indicates 'Stop, do not continue' (applicable when at least one of the previous criteria is not met), WHITE indicates 'No evidence' (absent) and AMBER indicates 'There is a concern, or caution, or weakness, but it is good enough to go forward, perhaps with a research agenda to move it to GREEN or RED' (this applies to all other situations where the final rating is not GREEN, RED or WHITE).[45]

Finally, to determine the level of endorsement, the OMERACT algorithm proposes that the instrument be: ENDORSED, when all SOMP columns have a synthesis rating of GREEN; PROVISIONALLY ENDORSED, when there is a mixture of GREEN and AMBER synthesis ratings across the measurement properties; and NOT ENDORSED, when any of the columns have either RED or WHITE synthesis ratings.[45]

Data extraction and synthesis

Data were extracted from the selected reports by the same two independent reviewers (BF and EJFS), and disagreements were discussed until consensus was achieved, or with adjudication by the third reviewer (PMM), whenever necessary. The reviewers had the option to contact the authors of papers to request missing or additional data, where required.

Since statistical pooling was not possible, the findings were presented in narrative form, including tables and figures, where appropriate.

RESULTS

Out of a total of 16 657 records (14 425 non-duplicate and 2232 duplicate records), 109 were selected for full-text review, and 60 studies fulfilled the inclusion criteria and were included in this SR. There was no need to contact the authors of the papers to request additional information. The results of the searches are shown in a flow diagram (figure 1).

Assessment of risk of bias

The critical appraisal of results (assessment of risk of bias) for each study is summarised in table 1. There was agreement among the reviewers to include all the studies that were appraised. In general, the eligible publications were predominantly focused on the study of construct validity (54 out of 60, 90%) and intermethod reliability (45 out of 60, 75%), with an overall low risk of bias (GREEN). To a lesser extent, they focused on test-retest reliability (13 out of 60, 21.7%) and responsiveness (14 out of 60, 23.3%), with an overall low risk of bias (GREEN). As far as clinical trial discrimination and thresholds of meaning were concerned, there were few studies addressing these aspects (6 out of 60, 10%; and 3 out of 60, 5%, respectively), so the results should be interpreted with caution, even though these studies mostly have a low bias (GREEN).

Characteristics of included studies and narrative synthesis

Study characteristics and the individual results of the psychometric and study-specific properties are available and detailed in online supplemental material S2.

The 60 studies included made it possible to assess the psychometric properties of 27 instruments: Patient-Reported Outcomes Measurement Information System (PROMIS)-Fatigue,[46–56] Pittsburgh Fatigability Scale (PFS),[57] FACIT-F,[58–70] SF-36 Vitality,[62 63 69 71–73] Lupus Quality of Life (LupusQoL) Fatigue,[62] VAS-Fatigue,[71 73–75] Fibromyalgia Impact Questionnaire (FIQ) Fatigue,[71] BRAF-MDQ,[42 72 76–80] BRAF-NRS,[42 72 76–78 81] Multidimensional Fatigue Inventory (MFI),[63 82 83] FSS,[63 84–90] Fatigue-NRS,[91] MAF,[63 73] Chalder Fatigue Scale (CFS),[63] Sjögren's Syndrome Symptom Diary (SSSD),[64] RAID-F,[77 80] Fatigue Assessment Scale (FAS),[92 93] Fatigue Impact Scale (FIS),[94] Multidimensional Daily Diary of Fatigue (MDF)-Fibro-17,[95 96] Worst Fatigue (WF)-NRS,[97] BRAF-VAS,[42] computerised adaptive test (CAT),[98] PedsQL-Fatigue,[43 99 100] Warwick Axial Spondyloarthritis faTigue and Energy questionnaire (WASTEd),[101] Profile of Fatigue (ProF),[102] Fatigue-Patient-Reported Outcome (PRO)[103] and Brief Fatigue Inventory (BFI).[73]

These instruments were applied to 13 different RMDs. The most commonly studied RMDs were RA (16 studies), SLE (15 studies) and FM (nine studies). Other diseases represented included axSpA and PsA (four studies each), OA, IIMs and Sjögren's disease (two studies each) and single studies for chronic back pain, SSc, sarcoidosis, JIA, familial Mediterranean fever and mixed rheumatological conditions with childhood-onset.

Reliability

48 studies evaluated reliability.[42 43 48 49 51 52 54–62 66 68–71 73–94 96 98–100 102 103] The different instruments used showed moderate to excellent indices in the different psychometric domains of reliability. More specifically, the studies showed good to excellent internal consistency (Cronbach's α=0.60–0.99; McDonald's omega >0.93); moderate to strong significant correlations between instruments and/or between the specific domains of the different instruments (r=0.40–0.94); good SEM per cent (<10–30%); acceptable to excellent ICC for instruments' domains (ICC=0.66–0.99); Kappa coefficients ranging from moderate to substantial agreement (Kappa=0.51–0.80); all corrected item-to-total correlations were ≥0.40 for all domains of the studies; none of the measures showed problems with floor effects or ceiling effects; among other measures. Overall, internal consistency was assessed in 38 studies, ICC in 28 studies and measurement error (Smallest Detectable Difference, SDD) in 7 studies.

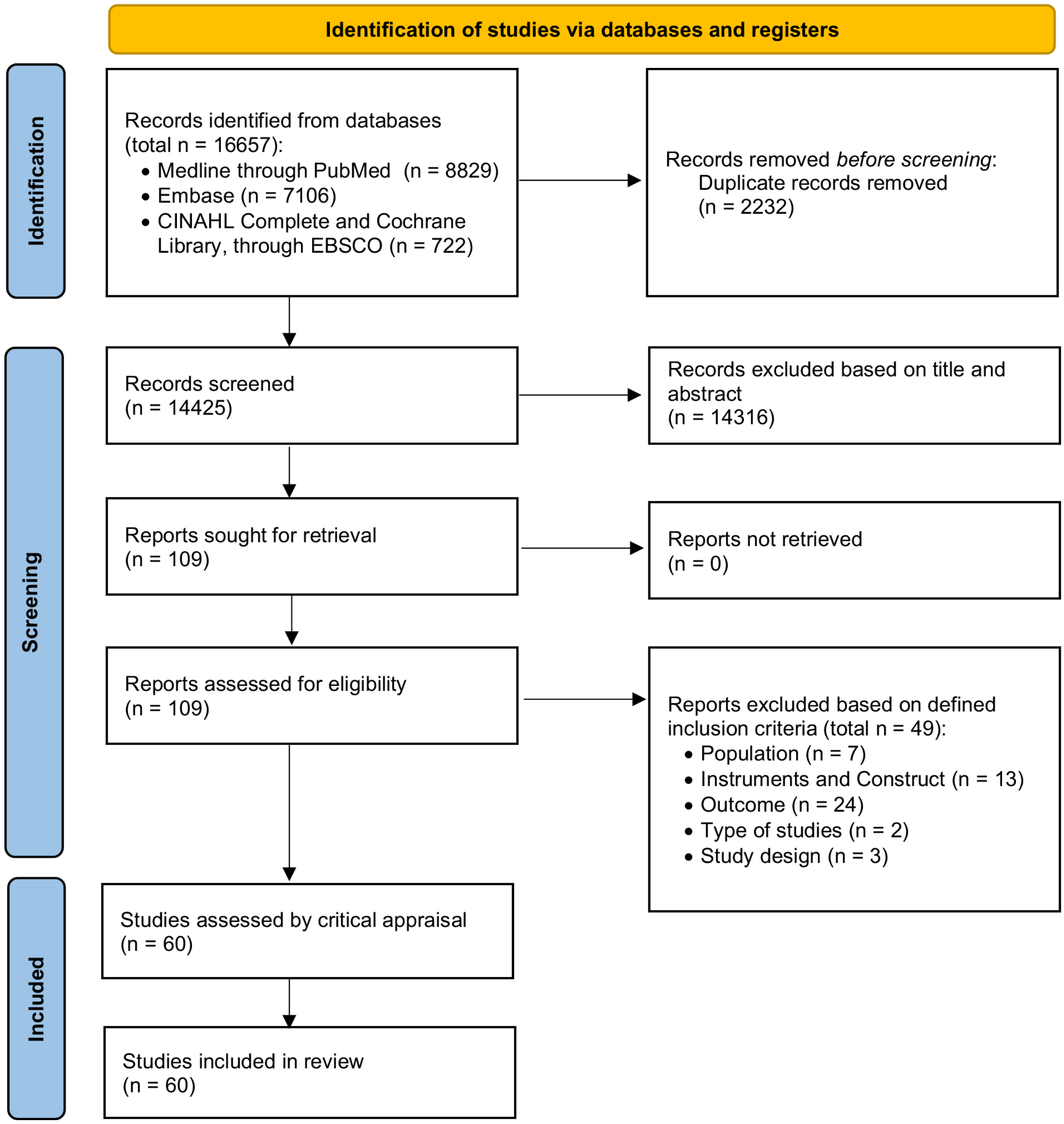

Figure 1 Flowchart of the study selection and inclusion process.

Validity

50 studies assessed validity.[42 43 46–48 51–55 57 59–61 64–70 72–74 77–99 101–103] The different instruments demonstrated moderate to strong significant correlations between instruments and/or between the specific domains of the different instruments (r=0.52–0.95). Construct validity ranged from weak to strong, with both negative and positive correlations reported. They also showed significant change scores and used exploratory and confirmatory factor analysis. Content validity was consistently reported as good and satisfactory across studies and was established by qualitative interviews, Delphi processes and other methods. Known-groups validity was available in four studies. Factor analysis (exploratory and/or confirmatory) was reported in five studies, providing evidence for structural validity of the instruments. More specifically, content validity was assessed in 13 studies, construct validity in 43 studies, criterion validity in 4 studies and external validity in 2 studies, while hypothesis testing was not reported.

Responsiveness

15 studies reported assessments of responsiveness.[47 50 55 58 62–66 68 69 74 76 91 96] Overall, they reported statistically significant mean changes, measures of central tendency, SRM and effect sizes, among other measures.

Instrument-level reporting of measurement properties

Across the included studies, each instrument was assessed separately for measurement properties. PROMIS-Fatigue demonstrated internal reliability in three studies, ICC in three studies and measurement error in three studies; validity assessments included content validity in two studies, construct validity in six studies and external validity in two studies, with responsiveness reported in two studies. FACIT-F showed internal reliability in eight studies and ICC in five studies, while validity included

Table 1 Assessment of risk of bias

Author	Instrument(s)	Construct validity (hypothesis testing)	Intermethod reliability	Test–retest reliability	Responsiveness	Clinical trial discrimination	Thresholds of meaning
Bartlett *et al*[46]	PROMIS Fatigue SFs 7a and 8a						
Beaumont *et al*[47]	PROMIS Fatigue 7a-item SF						
Bingham III *et al*[48]	PROMIS-Fatigue SF—4, 7 and 8-item						
Broderick *et al*[49]	PROMIS-Fatigue						
Carlozzi *et al*[57]	Pittsburgh Fatigability Scale						
Cella *et al*[50]	PROMIS-Fatigue						
Cella *et al*[58]	FACIT-Fatigue						
Cella *et al*[59]	FACIT-Fatigue						
Cella *et al*[60]	FACIT-Fatigue						
Chandran *et al*[61]	FACIT-Fatigue						
Christodoulou *et al*[51]	PROMIS Fatigue						
Cleanthous *et al*[62]	FACIT-F, SF-36 Vitality and LupusQoL Fatigue						
Crawford *et al*[71]	VAS-Fatigue, FIQ Fatigue and Tiredness and the SF-36 Vitality						
DiRenzo *et al*[52]	PROMIS-Fatigue 7a V.1.0						
Dures *et al*[76]	BRAF-MDQ and BRAF-NRS						
Ericsson and Mannerkorpi[82]	MFI-20						
Esfandiary *et al*[53]	PROMIS Fatigue Short Form 7a, Form 13a						
Feng *et al*[84]	FSS						
Gazzali *et al*[85]	FSS						
Gencay-Can and Can[86]	FSS						
Gladman *et al*[91]	Fatigue NRS						
Goligher *et al*[63]	FSS, VT, SF-36, MAF, FACIT-F, CFS, MFI						
Griffiths *et al*[64]	Sjögren's Syndrome Symptom Diary and FACIT-F						
Haroon *et al*[81]	BRAF-NRS						
Hewlett *et al*[77]	BRAF- MDQ revised (V2) BRAF-NRS and RAID						
Horisberger *et al*[92]	FAS						
Jones *et al*[54]	PROMIS-Fatigue						

Continued

Table 1 Continued

Author	Instrument(s)	Construct validity (hypothesis testing)	Intermethod reliability	Test–retest reliability	Responsiveness	Clinical trial discrimination	Thresholds of meaning
Junker *et al*[94]	Modified FIS						
Khanna *et al*[74]	Fatigue VAS						
Kiss *et al*[65]	FACIT-Fatigue						
Kratz *et al*[55]	PROMIS Fatigue-FM Profile						
Krupp *et al*[87]	FSS						
Lai *et al*[66]	FACIT-Fatigue						
Li *et al*[96]	MDF-Fibro-17						
Lorentzen *et al*[88]	FSS						
Magro *et al*[89]	FSS						
Mattsson *et al*[90]	FSS						
Michielsen *et al*[93]	FAS						
Morel *et al*[103]	Fatigue-PRO						
Morris *et al*[95]	MDF-Fibro-17						
Munguía-Izquierdo *et al*[83]	MFI						
Naegeli *et al*[97]	Worst Fatigue—NRS						
Nicklin *et al*[41]	BRAF-MDQ, BRAF-VAS and BRAF-NRS						
Nikolaus *et al*[98]	Multidimensional CAT						
Paulo *et al*[99]	PedsQL-Fatigue						
Pearson *et al*[101]	WASTEd						
Primdahl *et al*[78]	BRAF-MDQ and BRAF NRS version 2 (NRSv2)						
Raymond *et al*[67]	FACIT-Fatigue						
Rendas-Baum *et al*[68]	FACIT-Fatigue						
Revicki *et al*[69]	SF-36 and FACIT-Fatigue						
Rodríguez-Muguruza *et al*[70]	FACIT-Fatigue						
Saraç *et al*[100]	PedsQL-MFS						
Sari *et al*[79]	BRAF-MDQ						
Strömbeck *et al*[102]	ProF						
Studenic *et al*[75]	VAS-Fatigue						
Varni *et al*[43]	PedsQL-MFS						
Voshaar *et al*[72]	BRAF-MDQ, SF-36, BRAF-NRS						

Continued

Table 1 Continued

Author	Instrument(s)	Construct validity (hypothesis testing)	Intermethod reliability	Test–retest reliability	Responsiveness	Clinical trial discrimination	Thresholds of meaning
Voshaar *et al*[80]	BRAF-MDQ, RAID	Green	Green				
Wolfe[73]	VAS-Fatigue, MAF, SF-36, BFI	Green	Green				
Yost *et al*[56]	PROMIS-fatigue item bank (FIB)	Amber					

Green indicates good methods (ie, good avoidance of risk of bias), amber a non-critical limitation (ie, some concern of risk of bias) and red unsuitable evidence (ie, a warning for risk of bias).
BFI, Brief Fatigue Inventory; BRAF, Bristol Rheumatoid Arthritis Fatigue; CAT, computerised adaptive test; CFS, Chalder Fatigue Scale; FACIT, Functional Assessment of Chronic Illness Therapy; FAS, Fatigue Assessment Scale; FIQ, Fibromyalgia Impact Questionnaire; FIS, Fatigue Impact Scale; FM, fibromyalgia; FSS, Fatigue Severity Scale; LupusQoL, Lupus Quality of Life; MAF, Multidimensional Assessment of Fatigue; MDF, Multidimensional Daily Diary of Fatigue; MDQ, Multi-Dimensional Questionnaire; MFI, Multidimensional Fatigue Inventory; MFS, Multidimensional Fatigue Scale; NRS, Numeric Rating Scale; PedsQl, Paediatric Quality of Life Inventory; PedsQL, Paediatric Quality Of Life; PRO, patient-reported outcome; ProF, Profile of Fatigue; PROMIS, Patient-Reported Outcomes Measurement Information System; RAID, Rheumatoid Arthritis Impact of Disease; SF-36, 36-Item Short Form Survey Instrument; SF, Short Form; VAS, Visual Analogue Scale; VT, Vitality scale; WASTEd, Warwick Axial Spondyloarthritis faTigue and Energy questionnaire.

content validity in two studies, construct validity in nine studies and criterion validity in two studies; responsiveness was assessed in seven studies. SF-36 Vitality reported internal reliability in two studies and ICC in two studies, with construct validity in three studies and responsiveness in three studies. FSS was evaluated for internal reliability in seven studies and ICC in six studies; validity testing included content validity in three studies, construct validity in seven studies and criterion validity in one study, with responsiveness assessed in one study. BRAF-MDQ showed internal reliability in five studies, ICC in three studies and measurement error in one study; validity included content validity in one study, construct validity in seven studies and criterion validity in one study, with responsiveness reported in one study. BRAF-NRS reported internal reliability in three studies and ICC in two studies, construct validity in four studies and responsiveness in one study. VAS-Fatigue demonstrated internal reliability in two studies and ICC in two studies, with construct validity in two studies and responsiveness in one study. Other instruments, including the PFS, Fatigue-NRS, MAF, CFS, SSSD, RAID, MDF-Fibro-17, WF-NRS, BRAF-VAS, M-CAT, MFI, Modified Fatigue Impact Scale (MFIS), Lupus-QoL, FIQ, Peds-QoL, ProF and BFI, were each assessed in two or fewer studies for reliability or validity.

Summary ratings

According to the evaluation of the measurement properties of all the instruments included, we found that 5 out of 27 (18,5%) were rated GREEN and can therefore be 'endorsed', based on the use of OMERACT criteria: FACIT-F,[58–70] SF-36 Vitality,[62 63 69 71–73] BRAF-MDQ,[42 72 76–80] BRAF-NRS[42 72 76–78 81] and FSS.[63 84–90]

The remaining instruments, 22 out of 27 (81.5%) have been rated as AMBER and can be consequently 'provisionally endorsed', based on the use of OMERACT criteria: PROMIS-Fatigue,[46–56] PFS,[57] LupusQoL Fatigue,[62] VAS-Fatigue,[71 73–75] FIQ Fatigue,[71] MFI,[63 82 83] Fatigue-NRS,[91] MAF,[63 73] CFS,[63] SSSD,[64] RAID-F,[77 80] FAS,[92 93] FIS,[94] MDF-Fibro-17,[95 96] WF-NRS,[97] BRAF-VAS,[42] CAT,[98] PedsQL-Fatigue,[43 99 100] WASTEd,[101] ProF,[102] Fatigue-PRO[103] and BFI.[73]

The SOMP (summary of findings) is detailed in online supplemental material S3 and is shown in table 2.

DISCUSSION

This SR evaluated the psychometric properties of 27 instruments used to measure fatigue in RMDs. Of these, five instruments—FACIT-F, SF-36 Vitality, BRAF-MDQ, BRAF-NRS and FSS—were found to be valid, reliable and of low risk of bias, and can therefore be 'fully endorsed', based on OMERACT criteria. The remaining instruments received 'provisional endorsement' due to the limited number of dedicated studies and/or incomplete evaluation of key psychometric properties, restricting their confident use. This underscores the clear need for robust

Table 2 Summary of measurement properties (summary of findings)

Instruments studied	Number of studies	Proposed level of endorsement	References
PROMIS-Fatigue	11	Provisionally endorsed	46–56
PFS	1	Provisionally endorsed	57
FACIT-Fatigue	13	Endorsed	58–70
SF-36 Vitality	6	Endorsed	62 63 69 71–73
LupusQoL Fatigue	1	Provisionally endorsed	62
VAS-Fatigue	4	Provisionally endorsed	71 73–75
FIQ Fatigue	1	Provisionally endorsed	71
BRAF-MDQ	7	Endorsed	42 72 76–80
BRAF-NRS	6	Endorsed	42 72 76–78 81
MFI	3	Provisionally endorsed	63 82 83
FSS	8	Endorsed	63 84–90
Fatigue-NRS	1	Provisionally endorsed	91
MAF	2	Provisionally endorsed	63 73
CFS	1	Provisionally endorsed	63
SSSD	1	Provisionally endorsed	64
RAID-Fatigue	2	Provisionally endorsed	77 80
FAS	2	Provisionally endorsed	92 93
FIS	1	Provisionally endorsed	94
MDF-Fibro-17	2	Provisionally endorsed	95 96
WF-NRS	1	Provisionally endorsed	97
BRAF-VAS	1	Provisionally endorsed	42
CAT	1	Provisionally endorsed	98
PedsQL-Fatigue	3	Provisionally endorsed	43 99 100
WASTEd	1	Provisionally endorsed	101
ProF	1	Provisionally endorsed	102
Fatigue-PRO	1	Provisionally endorsed	103
BFI	1	Provisionally endorsed	73

OMERACT Endorsement criteria are based on the synthesis of the findings on all measurement properties. Profile solid green=full endorsement, good to go; profile is mix of amber and green=amber level, provisional endorsement, go ahead and use but also advance where more work needs to be done to get to full green; profile includes at least one red or white=do not use, unable to endorse with current body of evidence.

BFI, Brief Fatigue Inventory; BRAF, Bristol Rheumatoid Arthritis Fatigue; CAT, computerised adaptive test; CFS, Chalder Fatigue Scale; FACIT, Functional Assessment of Chronic Illness Therapy; FAS, Fatigue Assessment Scale; FIQ, Fibromyalgia Impact Questionnaire; FIS, Fatigue Impact Scale; FSS, Fatigue Severity Scale; LupusQoL, Lupus Quality of Life; MAF, Multidimensional Assessment of Fatigue; MDF, Multidimensional Daily Diary of Fatigue; MDQ, Multi-Dimensional Questionnaire; MFI, Multidimensional Fatigue Inventory; NRS, Numeric Rating Scale; OMERACT, Outcome Measures in Rheumatology; PedsQL, Paediatric Quality of Life; PFS, Pittsburgh Fatigability Scale; PRO, Patient-Reported Outcomes; ProF, Profile of Fatigue; PROMIS, Patient-Reported Outcomes Measurement Information System; RAID, Rheumatoid Arthritis Impact of Disease; SF, Short Forms; SF-36, 36-Item Short Form Survey Instrument; SSSD, Sjögren's Syndrome Symptom Diary; VAS, Visual Analogue Scale; WASTEd, Warwick Axial Spondyloarthritis faTigue and Energy; WF, Worst Fatigue.

research, particularly in children, adolescents and young adults, where data remain very scarce.

Despite the absence of a gold standard for fatigue measurement,[17] several instruments were able to discriminate between baseline and post-treatment states, as well as between treatment and placebo/usual care groups. The most frequently assessed instruments were FACIT-F (13 studies), PROMIS-Fatigue (11 studies), FSS (8 studies), BRAF-MDQ (7 studies), SF-36 Vitality (6 studies) and BRAF-NRS (6 studies). Many other instruments were assessed in four or fewer studies, limiting the strength of the evidence base for these tools, namely VAS-Fatigue (four studies), PedsQL-Fatigue (three studies), MFI (three studies), RAID-F (two studies), FAS (two studies), MAF (two studies), MDF-Fibro-17 (two studies), PFS (one study), LupusQoL Fatigue (one), FIQ-Fatigue (one), Fatigue-NRS (one study), CFS (one study), FIS (one study), WF-NRS (one study), BRAF-VAS (one study), CAT (one study), WASTEd (one study), ProF (one study), Fatigue-PRO (one study), SSSD (one study) and BFI (one study).

The heterogeneity in diseases and measurement instruments poses a significant barrier to establishing a universal standard for fatigue assessment. While strong supporting evidence exists for some tools in specific diseases—such as FACIT-F and PROMIS-Fatigue in RA and SLE—data remain limited for others, especially in conditions like vasculitis, SSc, connective tissue diseases and non-inflammatory MSK conditions, or paediatric rheumatic conditions. These gaps reinforce the need for further research before fatigue measurement can be standardised across the spectrum of RMDs. Despite variability in methodological robustness and differences in the number of studies conducted across various RMDs, it remains possible to confidently recommend certain instruments that are 'fully endorsed' according to OMERACT criteria for specific conditions. These include: FACIT-F for RA, SLE, Sjögren's disease, PsA and axSpA; SF-36 Vitality for RA, SLE and axSpA; BRAF-MDQ and BRAF-NRS for RA; and finally, FSS for SLE.

Although this review thoroughly examined aspects such as reliability and construct validity, important gaps persist. In particular, there is limited evidence on responsiveness, structural validity and measurement error for many instruments. These gaps may contribute to the proliferation of fatigue measures in clinical practice and the continued absence of a widely accepted, standardised instrument for use in both clinical and research settings.

Content validity is considered the most critical measurement property when selecting a PRO measure. This is best assessed through head-to-head studies where individuals of all ages affected by fatigue evaluate the relevance, comprehensiveness and clarity of multiple instruments.[104 105] In this review, construct validity was the most frequently assessed property, reported in 43 of the 60 included studies (72%). Instruments that showed sufficient evidence for construct validity included FACIT-F, BRAF-MDQ, BRAF-NRS, FSS, SF-36 Vitality, PedsQL-Fatigue, MDF-Fibro-17 and RAID-F.

In contrast, instruments such as PROMIS-Fatigue, BFI, Fatigue-PRO, ProF, WASTEd, CAT, BRAF-VAS, WF-NRS, FIS, FAS, SSSD, MAF, Fatigue-NRS, MFI and the PFS raised concerns regarding construct validity. Furthermore, for CFS, FIQ-Fatigue and LupusQoL-Fatigue, construct validity had not been evaluated at all. This points to the need for a careful reappraisal of these tools before their continued use is recommended.

Reliability was assessed in 48 studies (80%), with findings ranging from moderate to excellent, lending support to the consistency of many fatigue instruments. Responsiveness was assessed in only 15 studies (25%) but showed encouraging results. Nonetheless, further research is needed to better evaluate responsiveness, feasibility, measurement error and other less-studied psychometric properties.

In terms of feasibility in routine clinical practice, it is worth noting that all instruments included in this review are patient self-reports, designed to be easy to use and typically take between 1 and 8 min to complete. Most are also freely available. As this study did not aim to evaluate the clinical usability of the instruments, aspects such as the availability of translations, completion time, operationalisation, scoring procedures and licensing status, among other aspects, were not systematically analysed. For more information on these topics, we recommend consulting publications that specifically address the clinical application of individual instruments.[4 23]

This review has several notable strengths. It applied the COSMIN methodology, which adds transparency and rigour to the evaluation of psychometric properties and strengthens the confidence in its recommendations. Another strength is the inclusion of a diverse and representative sample of individuals with fatigue across a wide range of RMDs, including children, young people and adults, which enhances the generalisability of the findings. However, the review also has limitations. Heterogeneity in study design, patient populations, definitions of fatigue and instruments limited comparability across studies. Some degree of selection bias may also be present in the included studies, given that we only selected publications in English. Additionally, a meta-analysis could not be performed due to insufficient data for pooling, especially for condition-specific or lesser-used instruments.

In conclusion, this SR comprehensively assessed the psychometric properties of instruments used to measure fatigue in both children and adults with RMDs. It supports the use of several well-validated tools, particularly FACIT-F, SF-36 Vitality, BRAF-MDQ, BRAF-NRS and FSS, in both clinical and research settings. However, further high-quality studies are needed, particularly in children, adolescents, young adults and under-researched conditions such as vasculitis, SSc and non-inflammatory MSK disorders, to establish a standardised approach to fatigue measurement. These findings have important implications for advancing both the assessment and management of fatigue in rheumatology.

Author affiliations
[1]School of Health, Polytechnic Institute of Viseu, Viseu, Portugal
[2]Nursing School of Coimbra (ESEnfC), Health Sciences Research Unit: Nursing (UICiSA:E), Coimbra, Portugal
[3]Department of Neuromuscular Diseases, University College London, London, UK
[4]Department of Internal Medicine, Hacettepe University, Ankara, Turkey
[5]College of Health, Science and Society, University of the West of England, Bristol, UK
[6]Centre for Adolescent Rheumatology, Division of Medicine, University College London, London, UK
[7]General Internal Medicine, Withybush General Hospital, Haverfordwest, UK
[8]Aberdeen Centre for Arthritis and Musculoskeletal Health (Epidemiology Group), University of Aberdeen, Aberdeen, UK
[9]Academic Rheumatology, Bristol Royal Infirmary, Bristol, UK
[10]NIHR University College London Hospitals Biomedical Research Centre, University College London Hospitals NHS Foundation Trust, London, UK
[11]Department of Rheumatology, Northwick Park Hospital, London North West University Healthcare NHS Trust, London, UK

Social media Eduardo José Ferreira Santos, X @EduardoJFSantos

Acknowledgements PMM and CC are supported by the National Institute for Health Research (NIHR), University College London Hospitals (UCLH), Biomedical Research Centre (BRC).

Contributors EJFS and BF were the reviewers. PMM was the methodologist and principal/senior researcher. All authors contributed to data interpretation, reviewed and critically revised the manuscript and approved the final version. PMM is responsible for the overall content as the guarantor.

Funding This study was funded by Versus Arthritis and The Kennedy Trust for Rheumatology Research (project reference 23140).

Competing interests PMM has received consulting/speaker's fees from AbbVie, BMS, Celgene, Eli Lilly, Galapagos, Greywolf, Janssen, MSD, Novartis, Orphazyme, Pfizer, Roche and UCB, all unrelated to this manuscript. The other authors have no disclosures to report.

Patient and public involvement statement The research proposal included a patient co-applicant with lived experience of RMD-related fatigue. The patient co-applicant discussed the proposal, including the systematic review design and protocol, with the rest of the research team. The patient co-applicant has collaborated on several fatigue projects because they believe the symptom is a priority and that further research and changes to clinical practice are needed. One of the projects the patient co-applicant worked on was the EULAR recommendations for fatigue management, alongside patient partners from Germany and Italy. It was during this project that issues around how to measure fatigue and the need to understand the theory underpinning existing interventions emerged. Those discussions shaped the design of this research.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information. Not applicable.

ORCID iDs
Eduardo José Ferreira Santos https://orcid.org/0000-0003-0557-2377
Bayram Farisogullari https://orcid.org/0000-0002-9394-1103
Katie Fishpool https://orcid.org/0009-0000-7260-9577
Coziana Ciurtin https://orcid.org/0000-0002-8911-4113
Emmanuel Oghenetejiri Erhieyovwe https://orcid.org/0009-0003-4511-0369
Gary J Macfarlane https://orcid.org/0000-0003-2322-3314
Emma Dures https://orcid.org/0000-0002-6674-8607
Pedro M Machado https://orcid.org/0000-0002-8411-7972

REFERENCES

1 Murillo C, Vo T-T, Vansteelandt S, *et al*. How do psychologically based interventions for chronic musculoskeletal pain work? A systematic review and meta-analysis of specific moderators and mediators of treatment. *Clin Psychol Rev* 2022;94:102160.
2 Cieza A, Causey K, Kamenov K, *et al*. Global estimates of the need for rehabilitation based on the Global Burden of Disease study 2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* 2020;396:2006–17.
3 Vos T, Flaxman AD, Naghavi M, *et al*. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012;380:2163–96.
4 Santos EJF, Duarte C, da Silva JAP, *et al*. The impact of fatigue in rheumatoid arthritis and the challenges of its assessment. *Rheumatology (Oxford)* 2019;58:v3–9.
5 Dures E, Cramp F, Hackett K, *et al*. Fatigue in inflammatory arthritis. *Best Pract Res Clin Rheumatol* 2020;34:101526.
6 Pollard LC, Choy EH, Gonzalez J, *et al*. Fatigue in rheumatoid arthritis reflects pain, not disease activity. *Rheumatology (Oxford)* 2006;45:885–9.
7 Hewlett S, Chalder T, Choy E, *et al*. Fatigue in rheumatoid arthritis: time for a conceptual model. *Rheumatology (Oxford)* 2011;50:1004–6.
8 Overman CL, Kool MB, Da Silva JAP, *et al*. The prevalence of severe fatigue in rheumatic diseases: an international study. *Clin Rheumatol* 2016;35:409–15.
9 Repping-Wuts H, Fransen J, van Achterberg T, *et al*. Persistent severe fatigue in patients with rheumatoid arthritis. *J Clin Nurs* 2007;16:377–83.
10 Druce KL, Bhattacharya Y, Jones GT, *et al*. Most patients who reach disease remission following anti-TNF therapy continue to report fatigue: results from the British Society for Rheumatology Biologics Register for Rheumatoid Arthritis. *Rheumatology (Oxford)* 2016;55:1786–90.
11 Feldthusen C, Björk M, Forsblad-d'Elia H, *et al*. Perception, consequences, communication, and strategies for handling fatigue in persons with rheumatoid arthritis of working age--a focus group study. *Clin Rheumatol* 2013;32:557–66.
12 Sharpe M. Cognitive behavior therapy for chronic fatigue syndrome: efficacy and implications. *Am J Med* 1998;105:104S–109S.
13 Primdahl J, Hegelund A, Lorenzen AG, *et al*. The Experience of people with rheumatoid arthritis living with fatigue: a qualitative metasynthesis. *BMJ Open* 2019;9:e024338.
14 Seifert O, Baerwald C. Impact of fatigue on rheumatic diseases. *Best Pract Res Clin Rheumatol* 2019;33:101435.
15 Armbrust W, Siers NE, Lelieveld OTHM, *et al*. Fatigue in patients with juvenile idiopathic arthritis: A systematic review of the literature. *Semin Arthritis Rheum* 2016;45:587–95.
16 Nijhof LN, van de Putte EM, Wulffraat NM, *et al*. Prevalence of Severe Fatigue Among Adolescents With Pediatric Rheumatic Diseases. *Arthritis Care Res (Hoboken)* 2016;68:108–14.
17 Dures E, Farisoğulları B, Santos EJF, *et al*. 2023 EULAR recommendations for the management of fatigue in people with inflammatory rheumatic and musculoskeletal diseases. *Ann Rheum Dis* 2024;83:1260–7.
18 Kirwan JR, Minnock P, Adebajo A, *et al*. Patient perspective: fatigue as a recommended patient centered outcome measure in rheumatoid arthritis. *J Rheumatol* 2007;34:1174–7.
19 Hewlett S, Cockshott Z, Byron M, *et al*. Patients' perceptions of fatigue in rheumatoid arthritis: overwhelming, uncontrollable, ignored. *Arthritis Rheum* 2005;53:697–702.
20 Beckers E, Hermans K, Van Tubergen A, *et al*. Fatigue in patients with rheumatic and musculoskeletal diseases: a scoping review on definitions, measurement instruments, determinants, consequences and interventions. *RMD Open* 2023;9:e003056.
21 Ad Hoc Committee on Systemic Lupus Erythematosus Response Criteria for Fatigue. Measurement of fatigue in systemic lupus erythematosus: a systematic review. *Arthritis Rheum* 2007;57:1348–57.
22 Hewlett S, Hehir M, Kirwan JR. Measuring fatigue in rheumatoid arthritis: a systematic review of scales in use. *Arthritis Rheum* 2007;57:429–39.
23 Hewlett S, Dures E, Almeida C. Measures of fatigue: Bristol Rheumatoid Arthritis Fatigue Multi-Dimensional Questionnaire (BRAF MDQ), Bristol Rheumatoid Arthritis Fatigue Numerical Rating Scales (BRAF NRS) for severity, effect, and coping, Chalder Fatigue Questionnaire (CFQ), Checklist Individual Strength (CIS20R and CIS8R), Fatigue Severity Scale (FSS), Functional Assessment Chronic Illness Therapy (Fatigue) (FACIT-F), Multi-Dimensional Assessment of Fatigue (MAF), Multi-Dimensional Fatigue Inventory (MFI), Pediatric Quality Of Life (PedsQL) Multi-Dimensional Fatigue Scale, Profile of Fatigue (ProF), Short Form 36 Vitality Subscale (SF-36 VT), and Visual Analog Scales (VAS). *Arthritis Care Res (Hoboken)* 2011;63 Suppl 11:S263–86.
24 Hjollund NH, Andersen JH, Bech P. Assessment of fatigue in chronic disease: a bibliographic study of fatigue measurement scales. *Health Qual Life Outcomes* 2007;5:12.
25 Stephenson M, Riitano D, Wilson S, *et al*. Chapter 12: systematic reviews of measurement properties. In: Aromataris E, Munn Z, eds. *JBI manual for evidence synthesis*. 2020. Available: https://synthesismanual.jbi.global
26 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.
27 Page MJ, McKenzie JE, Bossuyt PM, *et al*. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71.
28 Arnett FC, Edworthy SM, Bloch DA, *et al*. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988;31:315–24.

29 Aletaha D, Neogi T, Silman AJ, *et al*. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis Rheum* 2010;62:2569–81.
30 Taylor W, Gladman D, Helliwell P, *et al*. Classification criteria for psoriatic arthritis: development of new criteria from a large international study. *Arthritis Rheum* 2006;54:2665–73.
31 Rudwaleit M, van der Heijde D, Landewé R, *et al*. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection. *Ann Rheum Dis* 2009;68:777–83.
32 Davies K, Dures E, Ng W-F. Fatigue in inflammatory rheumatic diseases: current knowledge and areas for future research. *Nat Rev Rheumatol* 2021;17:651–64.
33 Cella D, Lai J-S, Chang C-H, *et al*. Fatigue in cancer patients compared with fatigue in the general United States population. *Cancer* 2002;94:528–38.
34 Gossec L, Dougados M, Rincheval N, *et al*. Elaboration of the preliminary Rheumatoid Arthritis Impact of Disease (RAID) score: a EULAR initiative. *Ann Rheum Dis* 2009;68:1680–5.
35 Gossec L, Paternotte S, Aanerud GJ, *et al*. Finalisation and validation of the rheumatoid arthritis impact of disease score, a patient-derived composite measure of impact of rheumatoid arthritis: a EULAR initiative. *Ann Rheum Dis* 2011;70:935–42.
36 Wolfe F, Hawley DJ, Wilson K. The prevalence and meaning of fatigue in rheumatic disease. *J Rheumatol* 1996;23:1407–17.
37 Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473–83.
38 Belza BL, Henke CJ, Yelin EH, *et al*. Correlates of fatigue in older adults with rheumatoid arthritis. *Nurs Res* 1993;42:93–9.
39 Belza BL. Comparison of self-reported fatigue in rheumatoid arthritis and controls. *J Rheumatol* 1995;22:639–43.
40 Vercoulen JH, Swanink CM, Fennis JF, *et al*. Dimensional assessment of chronic fatigue syndrome. *J Psychosom Res* 1994;38:383–92.
41 Nicklin J, Cramp F, Kirwan J, *et al*. Collaboration with patients in the design of patient-reported outcome measures: Capturing the experience of fatigue in rheumatoid arthritis. *Arthritis Care Res (Hoboken)* 2010;62:1552–8.
42 Nicklin J, Cramp F, Kirwan J, *et al*. Measuring fatigue in rheumatoid arthritis: a cross-sectional study to evaluate the Bristol Rheumatoid Arthritis Fatigue Multi-Dimensional questionnaire, visual analog scales, and numerical rating scales. *Arthritis Care Res (Hoboken)* 2010;62:1559–68.
43 Varni JW, Burwinkle TM, Szer IS. The PedsQL Multidimensional Fatigue Scale in pediatric rheumatology: reliability and validity. *J Rheumatol* 2004;31:2494–500.
44 Terwee CB, Jansma EP, Riphagen II, *et al*. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res* 2009;18:1115–23.
45 Beaton D, Maxwell L, Grosskleg S, *et al*. Chapter 5: instrument selection for core outcome sets. In: *The OMERACT handbook version 21*. OMERACT, 2021. Available: https://omeract.org/handbook
46 Bartlett SJ, Gutierrez AK, Butanis A, *et al*. Combining online and in-person methods to evaluate the content validity of PROMIS fatigue short forms in rheumatoid arthritis. *Qual Life Res* 2018;27:2443–51.
47 Beaumont JL, Davis ES, Curtis JR, *et al*. Meaningful Change Thresholds for Patient-Reported Outcomes Measurement Information System (PROMIS) Fatigue and Pain Interference Scores in Patients With Rheumatoid Arthritis. *J Rheumatol* 2021;48:1239–42.
48 Bingham III CO, Gutierrez AK, Butanis A, *et al*. PROMIS Fatigue short forms are reliable and valid in adults with rheumatoid arthritis. *J Patient Rep Outcomes* 2019;3:14.
49 Broderick JE, Schneider S, Junghaenel DU, *et al*. Validity and reliability of patient-reported outcomes measurement information system instruments in osteoarthritis. *Arthritis Care Res (Hoboken)* 2013;65:1625–33.
50 Cella D, Lai J-S, Jensen SE, *et al*. PROMIS Fatigue Item Bank had Clinical Validity across Diverse Chronic Conditions. *J Clin Epidemiol* 2016;73:128–34.
51 Christodoulou C, Schneider S, Junghaenel DU, *et al*. Measuring daily fatigue using a brief scale adapted from the Patient-Reported Outcomes Measurement Information System (PROMIS ®). *Qual Life Res* 2014;23:1245–53.
52 DiRenzo D, Saygin D, de Groot I, *et al*. Reliability and validity of PROMIS physical function, pain interference, and fatigue as patient reported outcome measures in adult idiopathic inflammatory myopathies: International study from the OMERACT myositis working group. *Semin Arthritis Rheum* 2023;58:152111.
53 Esfandiary T, Park JK, Alexanderson H, *et al*. Assessing the content validity of patient-reported outcome measures in adult myositis: a report from the OMERACT myositis working group. *Semin Arthritis Rheum* 2020;50:943–8.
54 Jones JT, Carle AC, Wootton J, *et al*. Validation of Patient-Reported Outcomes Measurement Information System Short Forms for Use in Childhood-Onset Systemic Lupus Erythematosus. *Arthritis Care Res (Hoboken)* 2017;69:133–42.
55 Kratz AL, Schilling S, Goesling J, *et al*. The PROMIS FatigueFM Profile: a self-report measure of fatigue for use in fibromyalgia. *Qual Life Res* 2016;25:1803–13.
56 Yost KJ, Waller NG, Lee MK, *et al*. The PROMIS fatigue item bank has good measurement properties in patients with fibromyalgia and severe fatigue. *Qual Life Res* 2017;26:1417–26.
57 Carlozzi NE, Boileau NR, Murphy SL, *et al*. Validation of the Pittsburgh Fatigability Scale in a mixed sample of adults with and without chronic conditions. *J Health Psychol* 2021;26:1455–67.
58 Cella D, Wilson H, Shalhoub H, *et al*. Content validity and psychometric evaluation of Functional Assessment of Chronic Illness Therapy-Fatigue in patients with psoriatic arthritis. *J Patient Rep Outcomes* 2019;3:30.
59 Cella D, Yount S, Sorensen M, *et al*. Validation of the Functional Assessment of Chronic Illness Therapy Fatigue Scale relative to other instrumentation in patients with rheumatoid arthritis. *J Rheumatol* 2005;32:811–9.
60 Cella D, Lenderking WR, Chongpinitchai P, *et al*. Functional Assessment of Chronic Illness Therapy-Fatigue is a reliable and valid measure in patients with active ankylosing spondylitis. *J Patient Rep Outcomes* 2022;6:100.
61 Chandran V, Bhella S, Schentag C, *et al*. Functional assessment of chronic illness therapy-fatigue scale is valid in patients with psoriatic arthritis. *Ann Rheum Dis* 2007;66:936–9.
62 Cleanthous S, Bongardt S, Marquis P, *et al*. Psychometric Analysis from EMBODY1 and 2 Clinical Trials to Help Select Suitable Fatigue PRO Scales for Future Systemic Lupus Erythematosus Studies. *Rheumatol Ther* 2021;8:1287–301.
63 Goligher EC, Pouchot J, Brant R, *et al*. Minimal clinically important difference for 7 measures of fatigue in patients with systemic lupus erythematosus. *J Rheumatol* 2008;35:635–42.
64 Griffiths N, Wratten S, Flynn J, *et al*. Content Validity of Sjögren's Syndrome Symptom Diary and Functional Assessment of Chronic Illness Therapy-Fatigue in Patients with Sjögren's. *Rheumatol Ther* 2022;9:1559–74.
65 Kiss F, Farkas N, Nagy G, *et al*. Minimal Clinically Important Differences (MCID) for the Functional Assessment of Chronic Illness Therapy Fatigue Scale in Patients with Systemic Sclerosis. *IJERPH* 2022;20:771.
66 Lai J-S, Beaumont JL, Ogale S, *et al*. Validation of the Functional Assessment of Chronic Illness Therapy-Fatigue Scale in Patients with Moderately to Severely Active Systemic Lupus Erythematosus, Participating in a Clinical Trial. *J Rheumatol* 2011;38:672–9.
67 Raymond K, Maher S, Saucier CD, *et al*. Validation of the PROFAD-SSI-SF in Patients with Primary Sjögren's Syndrome with Organ Involvement: Results of Qualitative Interviews and Psychometric Analyses. *Rheumatol Ther* 2023;10:95–115.
68 Rendas-Baum R, Baranwal N, Joshi AV, *et al*. Psychometric properties of FACIT-Fatigue in systemic lupus erythematosus: a pooled analysis of three phase 3 randomised, double-blind, parallel-group controlled studies (BLISS-SC, BLISS-52, BLISS-76). *J Patient Rep Outcomes* 2021;5:33.
69 Revicki DA, Rentz AM, Luo MP, *et al*. Psychometric characteristics of the short form 36 health survey and functional assessment of chronic illness Therapy-Fatigue subscale for patients with ankylosing spondylitis. *Health Qual Life Outcomes* 2011;9:36.
70 Rodríguez-Muguruza S, Castro S, Poveda MJ, *et al*. Validation of FACIT-fatigue in Spanish-speaking patients with rheumatoid arthritis. *Revista Colombiana de Reumatología (English Edition)* 2024;31:435–41.
71 Crawford BK, Piault EC, Lai C, *et al*. Assessing fibromyalgia-related fatigue: content validity and psychometric performance of the Fatigue Visual Analog Scale in adult patients with fibromyalgia. *Clin Exp Rheumatol* 2011;29:S34–43.
72 Oude Voshaar MAH, Ten Klooster PM, Bode C, *et al*. Assessment of fatigue in rheumatoid arthritis: a psychometric comparison of single-item, multiitem, and multidimensional measures. *J Rheumatol* 2015;42:413–20.
73 Wolfe F. Fatigue assessments in rheumatoid arthritis: comparative performance of visual analog scales and longer fatigue questionnaires in 7760 patients. *J Rheumatol* 2004;31:1896–902.

74 Khanna D, Pope JE, Khanna PP, *et al*. The Minimally Important Difference for the Fatigue Visual Analog Scale in Patients with Rheumatoid Arthritis Followed in an Academic Clinical Practice. *J Rheumatol* 2008;35:2339–43.
75 Studenic P, Stamm T, Smolen JS, *et al*. Reliability of patient-reported outcomes in rheumatoid arthritis patients: an observational prospective study. *Rheumatology (Oxford)* 2016;55:41–8.
76 Dures EK, Hewlett SE, Cramp FA, *et al*. Reliability and sensitivity to change of the Bristol Rheumatoid Arthritis Fatigue scales. *Rheumatology (Oxford)* 2013;52:1832–9.
77 Hewlett S, Kirwan J, Bode C, *et al*. The revised Bristol Rheumatoid Arthritis Fatigue measures and the Rheumatoid Arthritis Impact of Disease scale: validation in six countries. *Rheumatology (Oxford)* 2018;57:300–8.
78 Primdahl J, Esbensen BA, Pedersen AK, *et al*. Validation of the Danish versions of the Bristol Rheumatoid Arthritis Fatigue Multi-Dimensional Questionnaires (BRAFs). *Scand J Rheumatol* 2021;50:351–9.
79 Sari F, Oskay D, Tufan A. Reliability, validity, and cross-cultural adaptation of the Turkish version of the Bristol Rheumatoid Arthritis Fatigue Multi-Dimensional Questionnaire. *Clin Rheumatol* 2018;37:1465–70.
80 Oude Voshaar MAH, Bode C, Hewlett S, *et al*. Psychometric properties and cross-language equivalence of the revised Bristol Rheumatoid Arthritis Fatigue and the Rheumatoid Arthritis Impact of Disease scales in rheumatoid arthritis. *Qual Life Res* 2019;28:2543–52.
81 Haroon M, Szentpetery A, Ashraf M, *et al*. Bristol rheumatoid arthritis fatigue scale is valid in patients with psoriatic arthritis and is associated with overall severe disease and higher comorbidities. *Clin Rheumatol* 2020;39:1851–8.
82 Ericsson A, Mannerkorpi K. Assessment of fatigue in patients with fibromyalgia and chronic widespread pain. Reliability and validity of the Swedish version of the MFI-20. *Disabil Rehabil* 2007;29:1665–70.
83 Munguía-Izquierdo D, Segura-Jiménez V, Camiletti-Moirón D, *et al*. Multidimensional Fatigue Inventory: Spanish adaptation and psychometric properties for fibromyalgia patients. The Al-Andalus study. *Clin Exp Rheumatol* 2012;30:94–102.
84 Feng C, He Q, Wu Y, *et al*. Psychometric properties of fatigue severity scale in Chinese systemic lupus erythematosus patients. *Health Qual Life Outcomes* 2019;17:71.
85 Gazzali I, Shahin MdA, Islam A, *et al*. Translation, Cross-Cultural Adaptation, and Validation of Fatigue Severity Scale into Bangla in Patients with Systemic Lupus Erythematosus. *Indian Journal of Rheumatology* 2022;17:250–4.
86 Gencay-Can A, Can SS. Validation of the Turkish version of the fatigue severity scale in patients with fibromyalgia. *Rheumatol Int* 2012;32:27–31.
87 Krupp LB, LaRocca NG, Muir-Nash J, *et al*. The fatigue severity scale. Application to patients with multiple sclerosis and systemic lupus erythematosus. *Arch Neurol* 1989;46:1121–3.
88 Lorentzen K, Danielsen MA, Kay SD, *et al*. Validation of the Fatigue Severity Scale in Danish patients with systemic lupus erythematosus. *Dan Med J* 2014;61:A4808.
89 Magro R, Camilleri L, Borg AA. Translation and validation of the Fatigue Severity Scale, Pittsburgh Sleep Quality Index and Modified Health Assessment Questionnaire into the Maltese Language, in a cohort of Maltese Systemic Lupus Erythematosus patients. *Mediterr J Rheumatol* 2017;28:192–200.
90 Mattsson M, Möller B, Lundberg I, *et al*. Reliability and validity of the Fatigue Severity Scale in Swedish for patients with systemic lupus erythematosus. *Scand J Rheumatol* 2008;37:269–77.
91 Gladman D, Nash P, Goto H, *et al*. Fatigue numeric rating scale validity, discrimination and responder definition in patients with psoriatic arthritis. *RMD Open* 2020;6:e000928.
92 Horisberger A, Courvoisier D, Ribi C. The Fatigue Assessment Scale as a simple and reliable tool in systemic lupus erythematosus: a cross-sectional study. *Arthritis Res Ther* 2019;21:80.
93 Michielsen HJ, De Vries J, Drent M, *et al*. Psychometric qualities of the Fatigue Assessment Scale in Croatian sarcoidosis patients. *Sarcoidosis Vasc Diffuse Lung Dis* 2005;22:133–8.
94 Junker CI, Dreyer L, Duch K, *et al*. Validation of the Modified Fatigue Impact Scale in patients with systemic lupus erythematosus. *Lupus (Los Angel)* 2022;31:1572–7.
95 Morris S, Li Y, Smith JAM, *et al*. Multidimensional daily diary of fatigue–fibromyalgia–17 items (MDF-fibro-17). part 1: development and content validity. *BMC Musculoskelet Disord* 2017;18:195.
96 Li Y, Morris S, Cole J, *et al*. Multidimensional daily diary of fatigue-fibromyalgia-17 items (MDF-fibro-17): part 2 psychometric evaluation in fibromyalgia patients. *BMC Musculoskelet Disord* 2017;18:198.
97 Naegeli AN, Flood E, Tucker J, *et al*. The patient experience with fatigue and content validity of a measure to assess fatigue severity: qualitative research in patients with ankylosing spondylitis (AS). *Health Qual Life Outcomes* 2013;11:192.
98 Nikolaus S, Bode C, Taal E, *et al*. Construct Validation of a Multidimensional Computerized Adaptive Test for Fatigue in Rheumatoid Arthritis. *PLoS One* 2015;10:e0145008.
99 Paulo LTSP, Len CA, Hilario MOE, *et al*. Fatigue in children with juvenile idiopathic arthritis: reliability of the "Pediatric Quality of Life Inventory-Multidimensional Fatigue Scale". *Rheumatol Int* 2015;35:133–9.
100 Saraç DC, Metin S, Demir BK, *et al*. Assessing fatigue and related factors in adolescents with familial Mediterranean fever (FMF): psychometric properties of the PedsQL Multidimensional Fatigue Scale. *Turk J Pediatr* 2023;65:630–9.
101 Pearson NA, Tutton E, Martindale J, *et al*. Development of the Warwick Axial Spondyloarthritis faTigue and Energy questionnaire (WASTEd)-a new patient-reported outcome measure. *Rheumatol Adv Pract* 2022;6:rkac027.
102 Strömbeck B, Theander E, Jacobsson LTH. Assessment of fatigue in primary Sjögren's syndrome: the Swedish version of the Profile of Fatigue. *Scand J Rheumatol* 2005;34:455–9.
103 Morel T, Cano S, Bartlett SJ, *et al*. The FATIGUE-PRO: a new patient-reported outcome instrument to quantify fatigue in patients affected by systemic lupus erythematosus. *Rheumatology (Oxford)* 2022;61:3329–40.
104 Prinsen CAC, Vohra S, Rose MR, *et al*. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" - a practical guideline. *Trials* 2016;17:449.
105 Terwee CB, Prinsen CAC, Chiarotto A, *et al*. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res* 2018;27:1159–70.