

Original citation:

Pearson, Nathan, Packham , Jonathan, Tutton, Elizabeth, Parsons, Helen and Haywood, Kirstie L. (2018) Assessing fatigue in adults with Axial Spondyloarthritis : a systematic review of the quality and acceptability of patient-reported outcome measures. Rheumatology Advances in Practice . rky017.

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/102051>

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work of researchers of the University of Warwick available open access under the following conditions.

This article is made available under the Creative Commons Attribution-NonCommercial 4.0 (CC BY-NC 4.0) license and may be reused according to the conditions of the license. For more details see: <http://creativecommons.org/licenses/by-nc/4.0/>

A note on versions:

The version presented in WRAP is the published version, or, version of record, and may be cited as it appears here.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

Manuscript title:

Assessing fatigue in adults with Axial Spondyloarthritis: a systematic review of the quality and acceptability of patient-reported outcome measures

Authors:

Nathan A Pearson^{1*}, Jonathan C Packham^{2,3}, Elizabeth Tutton^{1,4}, Helen Parsons⁵, Kirstie L Haywood¹

Affiliation:

1. Warwick Research in Nursing, Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry, UK
2. Institute of Applied Clinical Science, Keele University, Staffordshire, UK
3. Haywood Academic Rheumatology Centre, Staffordshire, UK
4. Trauma Research, Kadoorie Centre, John Radcliffe Hospital, UK
5. Clinical Trials Unit, Warwick Medical School, University of Warwick, Coventry, UK

Corresponding Author:

Mr Nathan A Pearson. Doctoral Research Student. Warwick Research in Nursing, Department of Health Sciences, Warwick Medical School, University of Warwick, Gibbet Hill, Coventry. CV4 7AL.

Email: n.pearson.1@warwick.ac.uk

Abstract

Objective

Evaluate the quality and acceptability of patient-reported outcome measures used to assess fatigue in patients with Axial Spondyloarthritis.

Methods

A two-stage systematic review of major electronic databases (1980-2017) to: 1) Identify measures; and 2) Identify evaluative studies. Study and measurement quality was evaluated following international standards. Measurement content was appraised against a conceptual model of RA-fatigue.

Results

From 387 reviewed abstracts, 23 articles provided evidence for nine fatigue-specific measures: six multi-item and three single-item. No axSpA-fatigue specific measure was identified.

Evidence of reliability was limited, but acceptable for the Multi-dimensional Fatigue Inventory (internal consistency, test-retest) and Short Form 36-item Health Survey Vitality subscale (SF-36 VT) (internal consistency). Evidence of construct validity was moderate for the Functional Assessment of Chronic Illness Therapy-Fatigue and 10cm visual analogue scale, limited for the SF-36 VT, and not available for the remaining measures. Responsiveness was rarely evaluated.

Evidence of measurement error, content validity or structural validity was not identified.

Most measures provide a limited reflection of fatigue; the most comprehensive were the Multi-dimensional Assessment of Fatigue (MAF), MFI-20, FACIT-fatigue and Fatigue Severity Scale (FSS).

Conclusion

The limited content and often poor quality of the reviewed measures limits any clear recommendation for fatigue-assessment in this population; assessments should be applied with caution until further robust evidence is established. Well developed, patient-derived measures can provide essential evidence of the patient's perspective to inform clinical research and drive tailored healthcare. The collaborative engagement of key stakeholders must seek to ensure that future fatigue assessment is relevant, acceptable and of high quality.

Keywords: fatigue assessment; measurement quality; acceptability; axial spondyloarthritis; systematic review

Key messages:

- 1) Fatigue is important to patients, but the quality and acceptability of assessment is limited.
- 2) Fatigue assessment is limited by methodological quality and limited relevance to patients.
- 3) Future guidance should be co-produced with patients, ensuring both assessment relevance and methodological rigor.

Introduction

Pain, stiffness, reduced mobility and fatigue are cardinal features of axial spondyloarthritis (axSpA), including Ankylosing Spondylitis (AS). (1). However, despite the importance afforded to fatigue by patients (2, 3), fatigue severity was only added to international assessment guidance for axSpA in 2009 (4). Accordingly, fatigue assessment in axSpA clinical trials increased significantly from just 17.1% of trials completed pre-2001 to 84% post-2001 (5), with most trials (84%) using the single fatigue-severity visual analogue scale (VAS) recommended in the assessment guidance (6). A recent conceptualisation of fatigue in rheumatoid arthritis (RA) demonstrated the multi-faceted and often-complex relationships between disease-specific, cognitive/ behavioural (behaviour, cognitive, emotion) and personal (support, health, environment, responsibilities) factors (7) – a complexity that might not be readily captured with a single item of severity (8). Moreover, individuals experiencing significant impairment due to frequent, but not severe (VAS scores <5), fatigue would not be identified if assessment were informed purely by fatigue severity (8). Patients' fatigue experience may, therefore, be better captured with multi-item, multi-domain patient-reported outcome measures (PROMs), providing a structured, patient-reported assessment of health (9, 10). These may be generic – containing items reflecting general health and completed by any population – or specific – to a condition (e.g., axSpA), an aspect of health (e.g., fatigue), or a population (e.g., children). A scoping review of fatigue measures used in rheumatology listed more than twelve multi-item measures, but just one rheumatology-specific, multi-item measure (11) - the Bristol RA Fatigue Multi-Dimensional Questionnaire (BRAf-MDQ) (12, 13). However, the quality, acceptability and relevance of measures was not explored, thus limiting evidence-based recommendations.

This review will systematically appraise, compare and synthesise published evidence of the quality and acceptability of clearly defined single and multi-item PROMs used in fatigue assessment in axSpA to establish the quality and acceptability of fatigue measures. The review will provide a transparent assessment of the evidence with which to inform PROM selection for future application in axSpA research and clinical practice.

Method

Identification of studies and PROMs

Medical Subject Headings (MeSH) and free text searching reflected: 1) population – axSpA/AS; 2) construct – fatigue; 3) assessment type – PROMs; and 4) measurement and practical properties (14). Five databases were searched: MEDLINE (OVID), EMBASE (OVID), PsycINFO (OVID), CINAHL and Web of Science; January 1980 to August 2017. A second search used the names of identified measures: 1) population; 2) construct; 3) named measures; and 4) measurement properties. Reference lists of included studies and existing reviews were reviewed (11, 15).

Eligibility criteria

One author (NAP) assessed all titles and abstracts; agreement was independently checked on a 10% subset by a second author (KLH). A third author (JCP) doubled-assessed all abstracts relating to Psoriatic Arthritis. Any conflicts were resolved through discussion.

Study inclusion

Studies were included if they contained a clearly identifiable and reproducible patient-reported assessment of fatigue, reported evidence of development and/or evaluation following completion by axSpA patients, and were written in English. Studies were excluded if they were only available as abstracts, fatigue assessment was not patient-reported, clearly identifiable or reproducible, or the study described PROM application only.

PROM inclusion

PROMs were included if they were fatigue-specific, assessed fatigue as a separate domain within a multi-domain measure, or were single or multi-item assessments. Clinician-reported assessments were excluded.

Data extraction and appraisal

Data extraction was informed by earlier published reviews (16-19), and the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) checklist (20-22). Study and PROM-specific information was extracted. Evidence of measurement properties included: validity, reliability, responsiveness, and interpretability (Appendix 2). Practical properties included evidence of: feasibility (administration time; scoring) and acceptability (patient relevance). Evidence of fatigue-conceptualisation and information pertaining to patient involvement was extracted and recorded. The RA-fatigue conceptual model (7) informed a comparative appraisal of PROM item content. One reviewer (NAP) completed all data extraction. A 10% subset was independently double-extracted (KLH) and agreement checked.

Assessment of study methodological quality

The COSMIN 4-point checklist informed an assessment of study methodological quality for each reported measurement property: “poor”, “fair”, “good” or “excellent” (20-22). The lowest item rating per measurement property informed the overall score.

Assessment of PROM quality

A synthesis of recommendations described by others (18, 19, 23) facilitated the transparent appraisal of PROM quality. Measurement properties were appraised and rated accordingly: adequate (+); inadequate (-); conflicting (+/-) or unclear (?). (Appendix 2)

Data synthesis and PROM recommendation

Four factors informed the synthesis: 1) study methodological quality (COSMIN); 2) number of studies reporting evidence; 3) ratings for measurement/practical properties per measure; and 4) consistency of results between studies (16, 18). The final synthesis, and hence evidence upon which PROM recommendation will be made, reflects both: 1) the quality of each measurement property: adequate (+), not adequate (-), conflicting (+/-), or unclear (?) (Appendix 2); and 2) overall level of evidence for each measurement property: ‘strong – consistent findings in multiple studies of good methodological quality OR in one study of excellent quality’, ‘moderate - consistent findings in multiple studies of fair methodological quality OR in one study of good quality’, ‘limited – one study of fair methodological quality’, ‘conflicting – conflicting findings’, or ‘unknown – only studies of poor methodological quality’ (18).

PROM recommendations will consider: 1) the extent to which key domains of fatigue identified in the RA-fatigue model are reflected the PROM (content validity); 2) whether there is adequate evidence, minimally, of measurement validity (structural and construct) and reliability (internal consistency and test-retest); and 3) an evidence base that is judged, as a minimum, to be moderate.

Results

Identification of studies and PROMs

A PRISMA flowchart summarises the review process (Figure 1). Twenty-three articles provided evidence for nine fatigue-specific PROMs (Table 1): three multi-dimensional - Multi-dimensional Assessment of Fatigue (MAF) (24); Multi-dimensional Fatigue Inventory (MFI-20) (25); Multi-dimensional Fatigue Symptom Inventory – Short Form (MFSI-SF) (26); three uni-dimensional - Functional Assessment of Chronic Illness Therapy (Fatigue) (FACIT-fatigue) (27), Fatigue Severity Scale (FSS) (28) and the vitality subscale (VT) of the Short-Form 36-item Health Status Survey (SF-36) (29); and three single-item questions: Worst-Fatigue Numeric Rating Scale (WF-NRS) (BFI) (30); the 10cm fatigue severity VAS (from the Bath AS Disease Activity Index (BASDAI)) (31); and a modified 10cm VAS whereby the descriptor ‘none’ was changed to ‘no problem’ (32).

Study and sample characteristics (Appendix 3)

All studies included adults with a primary diagnosis of axSpA, aged between 18 and 72 years old. Sample sizes ranged from 40 to 812. Studies were predominantly cross-sectional, investigating fatigue prevalence and/or its association with other variables.

Measurement properties and methodological quality

Study methodological quality (per PROM) was assessed and recorded (Appendix 4). An evidence synthesis is presented in Table 2. Evidence of measurement error, content or structural validity, criterion-based responsiveness, acceptability or feasibility of completion was not identified.

Fatigue conceptualisation (Table 1) and Patient Involvement

A review of PROM development suggests very limited conceptualisation of fatigue for four PROMs (MFI-20, MFSI-SF, SF-36 and BFI). Item generation or selection was often poorly reported and lacking in transparency. Only the single-item VAS of fatigue severity (taken from the BASDAI) was developed specifically for use with axSpA patients; but a conceptualisation of fatigue is absent. The involvement of patients did not extend beyond participation (that is, simply measurement completion); no study included patients as research partners in measurement evaluation.

Comparative item-content

Whilst similarities of item content exist, all reviewed measures provided a limited reflection of the RA-fatigue model (Table 3). All single-item measures assessed fatigue severity.

Multidimensional fatigue-specific PROMs

Multidimensional Assessment of Fatigue (MAF) (24)

Six poor quality studies provided limited evidence of construct validity (correlations and known-groups validity); including small to moderate associations between the MAF total and AS-specific Bath measures (range 0.23-0.73); and the MAF subscales and SF-36 VT (range 0.3-0.53) and 10cm fatigue-severity VAS (range 0.39-0.53) (33-38); all evaluations lacked a priori hypothesised associations.

Multidimensional Fatigue Inventory (MFI-20) (25)

One poor quality study provided limited evidence of construct validity (40). A fair quality study provided acceptable evidence of internal consistency (Cronbach's alpha from 0.68 (Reduced Motivation (RM) subscale) to 0.86 (Reduced Activity (RA) subscale)) and construct validity (39) (moderate to strong associations between subscales (General Fatigue (GF) with Physical Fatigue (PF) 0.69 / RA 0.52 / RM 0.45 / Mental Fatigue (MF) 0.45; MF with PF 0.40 / RA 0.42 / RM 0.48; RM with PF 0.51 / RA 0.54) supporting assumed a priori hypothesis associations) (39). Limited evidence for 1-week test-retest reliability was also reported for patients following completion of a VAS on a person's overall perceived health, taken from the EuroQoL (EQ-5D) (ICC range: PF 0.57 to 0.75 RM / MF) in a study judged to be of fair quality (40); for three subscales (GF, PF and RA) values less than 0.70 were reported. Distribution-based measures of responsiveness – both Effect Size (ES) statistics and the Standardised Response Mean (SRM) - were calculated from trial data, without any a priori hypotheses, following 3-month completion after the end of spa therapy: small values (<0.3) for domains reflecting reduced activity to large (>0.82) for domains reflecting general fatigue and physical fatigue were reported (ES: GF 0.82 / PF 0.81 / RA 0.28 / RM 0.54 / MF 0.38), SRM (GF 0.70 / PF 0.82 / RA 0.23 / RM 0.51 / MF 0.49; Guyatt statistics GF 0.86 / PF 0.96 / RA 0.30 / RM 0.50 / MF 0.57).

Multidimensional Fatigue Symptom Inventory - Short Form (MFSI-SF) (26)

One poor quality study provided limited evidence of construct validity (41). Weak to strong associations between the MFSI-SF subscales and the BASDAI 10cm VAS were reported (10cm VAS with General Fatigue (GF) 0.71 / Physical Fatigue (PF) 0.74 / Emotional Fatigue (EF) 0.56 / Mental Fatigue (MF) 0.45 / Vigor -0.32) following completion by 62 AS patients. Whilst

association between variables could be assumed, a priori hypothesised associations were not stated.

Unidimensional fatigue PROMs

Functional Assessment of Chronic Illness Therapy (Fatigue) (FACIT-fatigue) (27)

One poor quality study provided acceptable evidence of internal consistency (Cronbach's alpha 0.82/ 0.86), item-level performance (corrected Item-total correlation (ITC): 0.56 / 0.88) and construct validity (42). Good quality evidence of construct validity was available from the same article. Strong associations were reported between the FACIT-fatigue and SF-36 VT (range $r=0.74$ to 0.82) and the 10cm VAS ($r=-0.69$), with moderate associations with the BASDAI index score ($r=-0.47$) and Bath AS Functional Index (BASFI) ($r=-0.56$) (42). These findings confirmed a priori hypothesised associations between variables.

Fatigue Severity Scale (FSS) (28)

Both strong (0.77) (44) and moderate (0.53) (45) associations between the FFS and the 10cm fatigue-severity VAS have been reported in two studies judged to be of poor quality. Small effect sizes (ES) were reported at 28-days for participants in both arms of a placebo-controlled trial of subcutaneous etanercept (ES 0.15 / -0.23 ; SRM 0.22 / 0.22) (46).

Short Form Health Survey (SF-36) vitality subscale (VT) (29)

One fair quality study provided acceptable evidence of construct validity (42): a strong association between the VT subscale and the FACIT-fatigue was reported ($r=0.74$; $r=0.82$), a moderate association with the 10cm VAS ($r=-0.49$), and a weak association with the BASFI ($r=-$

0.33). One good quality study provided acceptable evidence of internal consistency and item-level performance (Cronbach's alpha 0.78/ 0.88; ITC 0.57/ 0.64) (42). Moderate to large ES statistics were reported at both 28-days (ES 0.54; SRM 0.83), and 112 days (ES 0.69; SRM 0.75) in patients receiving 25mg of etanercept subcutaneously, twice weekly (46).

Single-item fatigue PROMs

10cm Fatigue-Severity VAS (31)

One good quality study provided acceptable evidence of construct validity (42). A strong association between the item and the FACIT-fatigue ($r = -0.69$), and a moderate association with the SF-36 VT ($r = -0.49$) was reported following completion by AS patients participating in a double-blind, placebo-controlled clinical trial – supporting a priori hypothesised associations. A level of test-retest reliability judged to be below accepted standards for group analysis (ICC 0.60) was reported following a 6-week test-retest period in patients defined as stable on the EuroQoL EQ-VAS (general health); the study was judged to be of fair quality (40). However, estimates for test-retest reliability are below accepted thresholds for use with groups (0.70) or individuals (0.90) (54). In comparison to participants who received placebo or NSAIDs (small ES -0.35) (47), large effect size statistics (ES 0.89; SRM 0.89; Guyatt statistics 0.92) were reported at 6-weeks for participants receiving the active, spa therapy intervention (40).

Modified 10cm VAS (32)

The 10cm fatigue-severity VAS descriptor 'none' was modified to 'no problem' – changing the response scale. One poor quality study provided a limited, poor quality interpretative guidance (32).

Brief Fatigue Inventory (BFI) – Worst Fatigue Numeric Rating Scale (WF-NRS) (30)

One qualitative study explored the relevance and acceptability of the WF-NRS single item taken from the BFI (55). Although the item was judged to be relevant, the phraseology was confusing ('what best describes your *worst* fatigue'). A longer recall period than 24-hours was also recommended – to better express fatigue variability.

Discussion

Greater understanding of the impact of fatigue has been identified as a priority by axSpA patients (3). However, current assessment guidance is limited to a single-item measure of fatigue severity (4), which underestimates the often profound and wide-ranging impact of fatigue on an individual's life. Of the nine reviewed measures, just three were multi-dimensional, containing items reflecting different aspects of fatigue. However, no measure was specific to the experience of axSpA-fatigue and none had been evaluated for their relevance to axSpA patients. There was limited and often poor-quality evidence of reliability and construct validity; and an absence of interpretative guidance and evidence of measurement error, content validity or structural validity for any of the reviewed measures. Evidence of responsiveness was limited to the reporting of effective size statistics, which fail to provide an accurate evaluation of a measure's ability to detect meaningful change in health (19). Consequently, the lack of minimal measurement evidence for validity and reliability means it is not possible to make any assessment recommendations.

This is the first review of the quality and acceptability of measures of fatigue following completion by patients with axSpA. The results are strengthened by an evaluation of both study (20, 21) and PROM quality (16, 18, 19, 23); paired with a detailed comparative appraisal of item-content. However, much of the extracted data came from studies where PROM evaluation was not the primary focus of the study. As such, the rigour of the COSMIN criteria meant that these studies typically scored poorly. Although a single reviewer (NAP) assessed all titles and abstracts for review eligibility, a sub-set of titles and abstracts were reviewed by a second reviewer (KLH) and reliability checked.

Adoption of the RA-fatigue conceptual model in the current review highlighted the limited content validity of the reviewed measures: no PROM fully reflects the RA-model of fatigue. Both the MAF and FSS include the assessment of fatigue frequency *and* severity, two important components of the fatigue experience for axSpA patients (8). However, only two PROMs – the MFI-20 (10/20 (total) items) and FACIT-fatigue (6/13 items) – include items which seek to assess the cognitive/ behavioural (and emotional) impact of fatigue. Other PROMs (MAF, MFSI-SF, FSS) include items limited to only two of the cognitive/ behavioural domains. Whilst adequate evidence of internal consistency and reliability were reported for the MFI-20, it is unclear whether the PROM can detect change, or if it measures components of fatigue important to axSpA patients. Acceptable, but limited, evidence of a strong association between the FACIT-fatigue and SF-36 VT, enhances confidence in the ability of the FACIT-fatigue to measure fatigue in this population. However, evidence of measurement reliability and responsiveness is lacking in the axSpA population. Consequently, whilst demonstrating acceptable item content, both measures lack acceptable evidence of essential psychometric properties to currently support their use in axSpA-fatigue assessment. A robust fatigue assessment is necessary to detect and detail the nuances of fatigue experience that are essential to providing individualised and tailored healthcare to axSpA patients.

Qualitative research has detailed a similar experience of fatigue in axSpA, highlighting the significant impact of fatigue on social life patient mental health and relationships with others, their ability to engage with usual activities of daily living (56), and their reliance on self-management strategies (2). This demonstrates the importance of considering these aspects in the assessment of fatigue impact, and the insufficient information available from using only a single-item VAS of fatigue severity (56). Similarities between RA and axSpA-fatigue experience

supports the appropriateness of the RA-fatigue model as a framework against which PROM content and relevance can be appraised for use with the axSpA population (56, 57). However, growing evidence demonstrates that fatigue experience is a dynamic, complex and multifaceted experience that is, to a large extent, disease-specific. For example, evidence has shown both similarities and differences in fatigue experience between related and unrelated conditions (fibromyalgia, multiple sclerosis, AS, stroke) (58) and between different stages of illness, too – such as patients with active cancer compared to cancer survivors (59). Therefore, whilst this review has used the RA-fatigue model to appraise PROM item content, it is essential that a conceptual model is developed to reflect the nuances specific to the experience and impact of axSpA-fatigue.

A review of the quality of fatigue measures used in a range of chronic illnesses also highlighted the lack of evidence of essential measurement properties, thus limiting recommendations (61). However, the judgement of measurement quality lacked transparency and study methodological quality was not determined. International guidance promotes the importance of greater transparency in the assessment of measurement quality and acceptability (19, 23, 62). Adoption of the COSMIN checklist as per the current review, facilitates the incorporation of study methodological quality in the final judgement of PROM quality (20-22).

Well-developed, patient-derived PROMs are both robust and relevant to the experience of patients - capturing the outcomes that really matter (10, 63). However, numerous 'legacy' measures, where content was largely driven by the perspective of clinicians, may lack relevance to patients (10, 63). The failure of PROMs to capture the outcomes that really matter to patients (9, 64-66), undermines the potential contribution to patient-centred care and shared decision-making, and was the driver for the co-development of a new, patient-derived measure of fatigue

for RA – the Bristol RA Fatigue – Multi-dimensional Questionnaire (BRAf-MDQ) (12, 13). Of the 9 PROMs identified in this review, only four provided a limited conceptualisation of fatigue, which was mostly derived from literature reviews and clinical experts. Only one PROM (FACIT-Fatigue) was developed following a qualitative method – semi-structured interviews – but did not provide a conceptualisation of fatigue. Qualitative research offers greater insight into key health issues affecting patients, improving the relevance and acceptability of PROM content. This can highlight the unmet needs of patients, supporting targeted healthcare efforts to address what really matters to the patient.

The MFI-20 and FACIT-fatigue provide the most comprehensive assessment of fatigue (7), but evidence of their psychometric qualities in the axSpA population is limited.

A limited number of fatigue-specific PROMs have been evaluated for their quality and acceptability for use in axSpA fatigue assessment. However, recommendations are limited by the poor methodological quality of most studies coupled with the limited evidence of robust measurement or practical properties. These limitations also suggest that data generated from the application of these measures in routine practice or clinical research settings should be interpreted with caution. A comparative appraisal of PROM content suggests that the MFI and FACIT-fatigue provide the most comprehensive assessment of fatigue, including the impact on both cognition and behaviour. However, further exploration of the relevance and acceptability of the reviewed measures to patients with axSpA-fatigue is warranted. Moreover, comparative evaluations of those measures that have acceptable content validity, are urgently required to establish robust evidence essential measurement properties – specifically, reliability, validity, responsiveness and interpretation.

Declarations

Conflict of interest

The authors declare no conflicts of interest.

Funding statement

This work was supported by the National Ankylosing Spondylitis Society (NASS) [grant number: WAR1].

References

1. Raine C, Keat A. Axial spondyloarthritis. *Medicine*. 2014;42(5):251-6.
2. Davies H, Brophy S, Dennis M, Cooksey R, Irvine E, Siebert S. Patient perspectives of managing fatigue in Ankylosing Spondylitis, and views on potential interventions: a qualitative study. *BMC Musculoskeletal Disorders*. 2013;14(1):163.
3. National Ankylosing Spondylitis Society. NASS Research Priorities 2013-2018; 2013 [Available from: <http://nass.co.uk/research/nass-research-priorities/>].
4. Sieper J, Rudwaleit M, Baraliakos X, Brandt J, Braun J, Burgos-Vargas R, et al. The Assessment of SpondyloArthritis international Society (ASAS) handbook: a guide to assess spondyloarthritis. *Annals of the Rheumatic Diseases*. 2009;68(Suppl 2):ii1.
5. Bautista-Molano W, Navarro-Compán V, Landewé RBM, Boers M, Kirkham JJ, van der Heijde D. How well are the ASAS/OMERACT Core Outcome Sets for Ankylosing Spondylitis implemented in randomized clinical trials? A systematic literature review. *Clinical Rheumatology*. 2014;33(9):1313-22.
6. van der Heijde D, Calin A, Dougados M, Khan MA, van der Linden S, Bellamy N. Selection of instruments in the core set for DC-ART, SMARD, physical therapy, and clinical record keeping in ankylosing spondylitis. Progress report of the ASAS Working Group. *Assessments in Ankylosing Spondylitis. The Journal of rheumatology*. 1999;26(4):951-4.
7. Hewlett S, Chalder T, Choy E, Cramp F, Davis B, Dures E, et al. Fatigue in rheumatoid arthritis: time for a conceptual model. *Rheumatology*. 2011;50(6):1004-6.
8. Haywood KL, Packham JC, Jordan KP. Assessing fatigue in ankylosing spondylitis: the importance of frequency and severity. *Rheumatology*. 2014;53(3):552-6.
9. Haywood KL. Patient-reported outcome: Measuring what matters or just another paper exercise? *Musculoskeletal Care*. 2006;4(2):63-6.
10. Haywood KL, Wilson R, Stanisiewska S, Salek S. Using PROMs in Healthcare: Who Should Be in the Driving Seat-Policy Makers, Health Professionals, Methodologists or Patients? *The patient*. 2016;9(6):495-8.
11. Hewlett S, Dures E, Almeida C. Measures of fatigue: Bristol Rheumatoid Arthritis Fatigue Multi-Dimensional Questionnaire (BRAFMQ), Bristol Rheumatoid Arthritis Fatigue Numerical Rating Scales (BRAFNRS) 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 & research*. 2011;63 Suppl 11:S263-86.
12. Nicklin J, Cramp F, Kirwan J, Urban M, Hewlett S. Collaboration with patients in the design of patient-reported outcome measures: capturing the experience of fatigue in rheumatoid arthritis. *Arthritis care & research*. 2010;62(11):1552-8.
13. Nicklin J, Cramp F, Kirwan J, Greenwood R, Urban M, Hewlett S. 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 & research*. 2010;62(11):1559-68.
14. Terwee CB, Jansma EP, Riphagen, II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments.

Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2009;18(8):1115-23.

15. Haywood KL, Garratt AM, Dawes PT. Patient-assessed health in ankylosing spondylitis: a structured review. *Rheumatology*. 2005;44(5):577-86.
16. Haywood KL, Collin SM, Crawley E. Assessing severity of illness and outcomes of treatment in children with Chronic Fatigue Syndrome/Myalgic Encephalomyelitis (CFS/ME): a systematic review of patient-reported outcome measures (PROMs). *Child: care, health and development*. 2014;40(6):806-24.
17. Haywood KL, Staniszewska S, Chapman S. Quality and acceptability of patient-reported outcome measures used in chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME): a systematic review. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*. 2012;21(1):35-52.
18. Elbers RG, Rietberg MB, van Wegen EEH, Verhoef J, Kramer SF, Terwee CB, et al. Self-report fatigue questionnaires in multiple sclerosis, Parkinson's disease and stroke: a systematic review of measurement properties. *Quality of Life Research*. 2012;21(6):925-44.
19. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of clinical epidemiology*. 2007;60(1):34-42.
20. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of clinical epidemiology*. 2010;63(7):737-45.
21. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of Life Research*. 2010;19(4):539-49.
22. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Quality of Life Research*. 2012;21(4):651-7.
23. Conijn AP, Jens S, Terwee CB, Breek JC, Koelemay MJ. Assessing the quality of available patient reported outcome measures for intermittent claudication: a systematic review using the COSMIN checklist. *European journal of vascular and endovascular surgery : the official journal of the European Society for Vascular Surgery*. 2015;49(3):316-34.
24. Tack BB. Dimensions and correlates of fatigue in the older adult with rheumatoid arthritis. San Francisco: University of California; 1991.
25. Smets EM, Garsen B, Bonke B, De Haes JC. The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. *Journal of psychosomatic research*. 1995;39(3):315-25.
26. Stein KD, Jacobsen PB, Blanchard CM, Thors C. Further validation of the multidimensional fatigue symptom inventory-short form. *Journal of pain and symptom management*. 2004;27(1):14-23.
27. Yellen SB, Cella DF, Webster K, Blendowski C, Kaplan E. Measuring fatigue and other anemia-related symptoms with the Functional Assessment of Cancer Therapy (FACT) measurement system. *Journal of pain and symptom management*. 1997;13(2):63-74.

28. Krupp LB, LaRocca NG, Muir-Nash J, Steinberg AD. The fatigue severity scale. Application to patients with multiple sclerosis and systemic lupus erythematosus. *Archives of neurology*. 1989;46(10):1121-3.
29. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical care*. 1992;30(6):473-83.
30. Mendoza TR, Wang XS, Cleeland CS, Morrissey M, Johnson BA, Wendt JK, et al. The rapid assessment of fatigue severity in cancer patients. *Cancer*. 1999;85(5):1186-96.
31. Garrett S, Jenkinson T, Kennedy LG, Whitelock H, Gaisford P, Calin A. A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. *The Journal of rheumatology*. 1994;21(12):2286-91.
32. Wheaton L, Pope J. The Minimally Important Difference for Patient-reported Outcomes in Spondyloarthropathies including Pain, Fatigue, Sleep, and Health Assessment Questionnaire. *Journal of Rheumatology*. 2010;37(4):816-22.
33. Aissaoui N, Rostom S, Hakkou J, Berrada Ghziouel K, Bahiri R, Abouqal R, et al. Fatigue in patients with ankylosing spondylitis: prevalence and relationships with disease-specific variables, psychological status, and sleep disturbance. *Rheumatology International*. 2012;32(7):2117-24.
34. Durmus D, Alayli G, Cil E, Canturk F. Effects of a home-based exercise program on quality of life, fatigue, and depression in patients with ankylosing spondylitis. *Rheumatology International*. 2009;29(6):673-7.
35. Ibn Yacoub Y, Amine B, Laatiris A, Abouqal R, Hajjaj-Hassouni N. Assessment of fatigue in Moroccan patients with ankylosing spondylitis. *Clinical Rheumatology*. 2010;29(11):1295-9.
36. Stebbings SM, Treharne GJ, Jenks K, Highton J. Fatigue in patients with spondyloarthritis associates with disease activity, quality of life and inflammatory bowel symptoms. *Clinical Rheumatology*. 2014;33(10):1467-74.
37. Turan Y, Duruoaz MT, Bal S, Guvenc A, Cerrahoglu L, Gurgan A. Assessment of fatigue in patients with ankylosing spondylitis. *Rheumatology International*. 2007;27(9):847-52.
38. Alkan BM, Fidan F, Erten S, Aksekili H, Alemdar A, Eroglu E, et al. Fatigue and correlation with disease-specific variables, spinal mobility measures, and health-related quality of life in ankylosing spondylitis. *Modern Rheumatology*. 2013;23(6):1101-7.
39. Da Costa D, Zummer M, Fitzcharles MA. Biopsychosocial determinants of physical and mental fatigue in patients with spondyloarthropathy. *Rheumatology International*. 2011;31(4):473-80.
40. Van Tubergen A, Coenen J, Landewé R, Spoorenberg A, Chorus A, Boonen A, et al. Assessment of Fatigue in Patients With Ankylosing Spondylitis: A Psychometric Analysis. *Arthritis & Rheumatism: Arthritis Care & Research*. 2002;47(1):8-16.
41. Gunaydin R, Karatepe AG, Cesmeli N, Kaya T. Fatigue in patients with ankylosing spondylitis: relationships with disease-specific variables, depression, and sleep disturbance. *Clinical Rheumatology*. 2009;28(9):1045-51.
42. Revicki DA, Rentz AM, Luo MP, Wong RL. 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.
43. Maksymowych WP, Richardson R, Mallon C, van der Heijde D, Boonen A. Evaluation and validation of the patient acceptable symptom state (PASS) in patients with ankylosing-spondylitis. *Arthritis & Rheumatism-Arthritis Care & Research*. 2007;57(1):133-9.

44. Bedaiwi M, Sari I, Thavaneswaran A, Ayearst R, Haroon N, Inman RD. Fatigue in Ankylosing Spondylitis and Nonradiographic Axial Spondyloarthritis: Analysis from a Longitudinal Observation Cohort. *Journal of Rheumatology*. 2015;42(12):2354-60.
45. Schneeberger EE, Marengo MF, Dal Pra F, Maldonado Cocco JA, Citera G. Fatigue assessment and its impact in the quality of life of patients with ankylosing spondylitis. *Clinical Rheumatology*. 2015;34(3):497-501.
46. Wanders AJB, Gorman JD, Davis JC, Landewe RBM, Van Der Heijde DMFM. Responsiveness and Discriminative Capacity of the Assessments in Ankylosing Spondylitis Disease-Controlling Antirheumatic Therapy Core Set and Other Outcome Measures in a Trial of Etanercept in Ankylosing Spondylitis. *Arthritis Care and Research*. 2004;51(1):1-8.
47. Dernis-Labous E, Messow M, Dougados M. Assessment of fatigue in the management of patients with ankylosing spondylitis. *Rheumatology*. 2003;42(12):1523-8.
48. Fallahi S, Jamshidi AR, Bidad K, Qorbani M, Mahmoudi M. Evaluating the reliability of Persian version of ankylosing spondylitis quality of life (ASQoL) questionnaire and related clinical and demographic parameters in patients with ankylosing spondylitis. *Rheumatology International*. 2014;34(6):803-9.
49. Park HJ, Kim S, Lee JE, Jun JB, Bae SC. The reliability and validity of a Korean translation of the BASDAI in Korean patients with Ankylosing spondylitis. *Value in Health*. 2008;11:S99-S104.
50. Yilmaz O, Tutoglu A, Garip Y, Ozcan E, Bodur H. Health-related quality of life in Turkish patients with ankylosing spondylitis: impact of peripheral involvement on quality of life in terms of disease activity, functional status, severity of pain, and social and emotional functioning. *Rheumatology International*. 2013;33(5):1159-63.
51. Dagfinrud H, Vollestad NK, Loge JH, Kvien TK, Mengshoel AM. Fatigue in patients with ankylosing spondylitis: A comparison with the general population and associations with clinical and self-reported measures. *Arthritis & Rheumatism-Arthritis Care & Research*. 2005;53(1):5-11.
52. Fernandez-Sueiro JL, Willisch A, Pertega-Diaz S, Tasende JAP, Fernandez-Lopez JC, Villar NO, et al. Validity of the Bath Ankylosing Spondylitis Disease Activity Index for the evaluation of disease activity in axial psoriatic arthritis. *Arthritis Care and Research*. 2010;62(1):78-85.
53. Bodur H, Ataman Ş, Rezvani A, Buğdaycı DS, Çevik R, Birtane M, et al. Quality of life and related variables in patients with ankylosing spondylitis. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation*. 2011;20(4):543-9.
54. D.L. Streiner and G.R. Norman *Health Measurement Scales—A Practical Guide to Their Development and Use* 4th ed. Oxford, UK: Oxford University Press; 2008.
55. Naegeli AN, Flood E, Tucker J, Devlen J, Edson-Heredia E. The patient experience with fatigue and content validity of a measure to assess fatigue severity: Qualitative research in patients with ankylosing spondylitis (AS). *Health and Quality of Life Outcomes*. 2013;11 (1) (no pagination)(192).
56. Farren W, Goodacre L, Stigant M. Fatigue in ankylosing spondylitis: causes, consequences and self-management. *Musculoskeletal Care*. 2013;11(1):39-50.
57. Hewlett S, Cockshott Z, Byron M, Kitchen K, Tipler S, Pope D, et al. Patients' perceptions of fatigue in rheumatoid arthritis: overwhelming, uncontrollable, ignored. *Arthritis and rheumatism*. 2005;53(5):697-702.

58. Eilertsen G, Ormstad H, Kirkevold M, Mengshoel AM, Söderberg S, Olsson M. Similarities and differences in the experience of fatigue among people living with fibromyalgia, multiple sclerosis, ankylosing spondylitis and stroke. *Journal of Clinical Nursing*. 2015;24(13-14):2023-34.
59. de Raaf PJ, de Klerk C, Timman R, Hinz A, van der Rijt CC. Differences in fatigue experiences among patients with advanced cancer, cancer survivors, and the general population. *Journal of pain and symptom management*. 2012;44(6):823-30.
60. Neuberger GB. Measures of fatigue: The Fatigue Questionnaire, Fatigue Severity Scale, Multidimensional Assessment of Fatigue Scale, and Short Form-36 Vitality (Energy/Fatigue) Subscale of the Short Form Health Survey. *Arthritis care & research*. 2003;49(S5):S175-S83.
61. Whitehead L. The measurement of fatigue in chronic illness: a systematic review of unidimensional and multidimensional fatigue measures. *Journal of pain and symptom management*. 2009;37(1):107-28.
62. Boers M, Kirwan JR, Wells G, Beaton D, Gossec L, d'Agostino M-A, et al. Developing Core Outcome Measurement Sets for Clinical Trials: OMERACT Filter 2.0. *Journal of clinical epidemiology*. 2014;67(7):745-53.
63. Staniszewska S, Haywood KL, Brett J, Tutton L. Patient and public involvement in patient-reported outcome measures: evolution not revolution. *The patient*. 2012;5(2):79-87.
64. US Food and Drug Administration Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Rockville: Department of Health and Human Services, Food and Drug Administration, Centre for Drug Evaluation and Research; 2009.
65. Devlin N, Appleby J. Getting the Most Out of PROMs. Putting health outcomes at the heart of NHS decision-making.: The Kings Fund, London; 2010 [Available from: <https://www.kingsfund.org.uk/sites/files/kf/Getting-the-most-out-of-PROMs-Nancy-Devlin-John-Appleby-Kings-Fund-March-2010.pdf>]
66. Patrick DL, Burke LB, Powers JH, Scott JA, Rock EP, Dawisha S, et al. Patient-reported outcomes to support medical product labeling claims: FDA perspective. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*. 2007;10 Suppl 2:S125-37.

Figure 1: PRISMA flow-chart of study inclusion

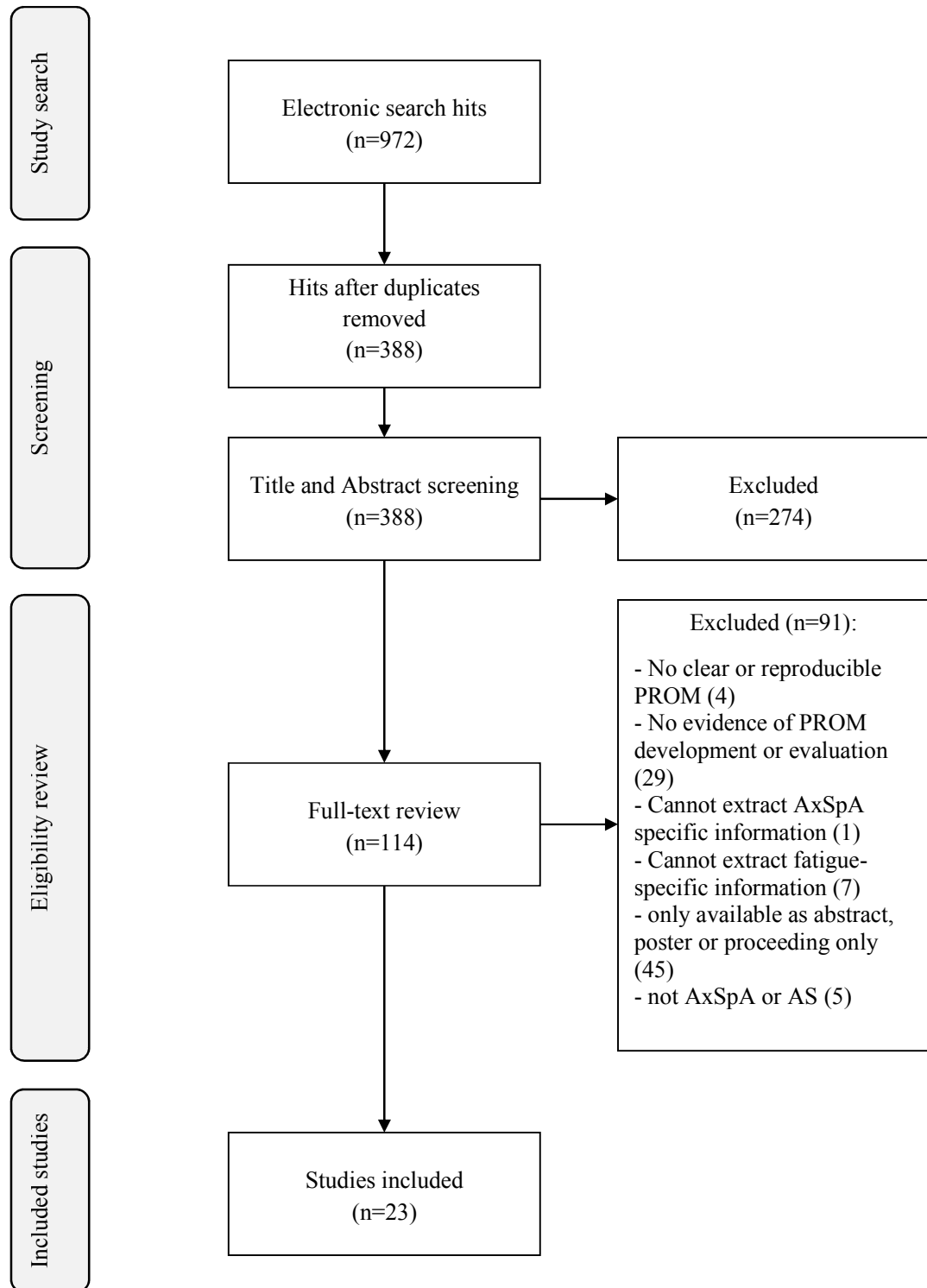


Table 1: Summary of the reviewed single- and multi-item fatigue PROMs (n=9).

PROM First author (year) Completion format Completion time	Conceptual focus Response options; Recall period	How to score
Multi-dimensional fatigue measures (3/9)		
Multi-dimensional Assessment of Fatigue (MAF) Tack (1991) (24) <i>Completion format:</i> self-completed <i>Completion time:</i> approx.. <10 mins.	A revision of the Piper Fatigue Scale (PFS) which was developed and tested with oncology patients which was designed to measure: temporal, severity and sensory fatigue dimensions. The 41-item PFS was reduced to form the 16-item MAF. The MAF was developed to measure 4 dimensions: fatigue severity, distress, impact and timing. No information provided about item selection, retention or generation. No evidence of patient involvement as research partners. MAF developed to measure multiple dimensions of fatigue in adult rheumatoid arthritis patients. 15 questions which contribute to a global fatigue index (GFI). A 16 th question does not contribute to the GFI. 4 subscales explored: distress, severity, interference in daily living activities and frequency/change during the last week. Response options NRS from 1-10. Anchors vary by items. Items 1 and 4-14 are anchored with “not at all” and “a great deal”. Item 2 is anchored with “mild” and “severe”. Item 3 is anchored with “no distress” and “a great deal of distress”. Items 15 and 16 require an ordinal response with 4 options scored 1-4. Item 15 is anchored with “hardly any days” to “every day”. Item 16 is anchored with “decreased” to “increased”. Recall period One week	If item 1 is scored as 0 then the remaining items should be scored as 0. Calculating the GFI should be done using the following four steps: i) sum items 1-3 ii) average items 4-14 (do not include activities marked as “do not do” in the average). iii) multiple item 15 by 2.5 to create a score from 0-10. iv) sum the values from i, ii and iii together to get a GFI. Item 16 does not contribute to the GFI but is scored from 1-4.
Multi-dimensional Fatigue Inventory (MFI-20) Smets (1995) (25) <i>Completion format:</i> self-completed <i>Completion time:</i> approx.. 5 mins.	Developed to measure fatigue in cancer patients without somatic items. Initial development was informed by the authors and previous research, resulting in 5 proposed domains: general, physical sensations and cognitive symptoms which were theoretically supported for inclusion in the MFI based on factor analyses conducted in other fatigue studies. Reduced motivation and reduced activity formed the final 2 components, but unclear whether reduced activity was considered a consequence of fatigue. No information about the process of item generation. Patients were included as participants but not research partners. Initially evaluated in patients with cancer, chronic fatigue syndrome and healthy individuals whom may experience physical fatigue (military personnel) or mental fatigue (newly qualified doctors). 5 subscales, each made up of 4 items (total = 20 items). Half of the items are positively phrased, thus requiring reverse scoring. The following 5 subscales were explored: general fatigue, physical fatigue, reduced activity, reduced motivation and mental fatigue. Response options A total of 5 check boxes with anchors “yes that is true” to “no that is not true”. Recall period No specific timescale stated. Instructions state “... how you have been feeling <u>lately</u> ...”	Each item is scored from 1-5. Positively phrased questions (n=10) must be reverse scored. Subscales can be individually scored by summing their respective items.
Multi-dimensional Fatigue Symptom Inventory – Short Form (MFSI-SF)	MFSI-SF is derived from the MFSI – an 83-item measure informed by evidence from fatigue literature, discussions with experts who treat fatigue patients and a review of ‘current’ fatigue measures. Subscales were empirically derived from factor analysis. 5 domains were identified: 1) global fatigue experience; 2) somatic symptoms; 3) cognitive symptoms; 4) affective symptoms; and 5) behavioural symptoms. Patients were included as participants but not as research partners. The MFSI-SF is made up of the empirically derived subscales only. Developed for use	Total score is a summation of 4 subscales (general + physical + emotional + mental) minus vigour.

<p>Stein (2004) (26)</p> <p><i>Completion format:</i> self-completed</p> <p><i>Completion time:</i> approx.. 5 mins.</p>	<p>in cancer patients.</p> <p>5 subscales, each made up of 6 items (total = 30 items). The 5 subscales explore: general fatigue, physical fatigue, emotional fatigue, mental fatigue and vigour.</p> <p>Response options 5 response options ranging from 0-4 with anchors “not at all” to “extremely”.</p> <p>Recall period Past 7 days</p>	
<p>Uni-dimensional fatigue measures (2/9)</p>		
<p>Functional Assessment Chronic Illness Therapy (FACIT-fatigue)</p> <p>Yellen (1997) (27)</p> <p><i>Completion format:</i> self-completed</p> <p><i>Completion time:</i> approx.. <5 mins.</p>	<p>FACIT-fatigue was developed to measure cancer-related fatigue. Development was two phase: phase 1 (item generation) used semi-structured interviews with cancer patients and medical experts; phase 2 (item reduction) was a presentation of questions to a second group of medical experts and their review. No conceptual model of fatigue was reported. Patients were included as participants but not research partners. Initial psychometric evaluation (with cancer patients) indicates that the measure has good validity and reliability.</p> <p>Measure is made up of 13 items which ask about fatigue, and how it impacts on daily activities.</p> <p>Response options 5 response options with anchors “not at all” to “very much”.</p> <p>Recall period Past 7 days</p>	<p>Global score calculated by summing item scores.</p>
<p>Fatigue Severity Scale (FSS)</p> <p>Krupp (1989) (28)</p> <p><i>Completion format:</i> self-completed</p> <p><i>Completion time:</i> approx.. <5 mins.</p>	<p>Developed to assess fatigue in patients with multiple sclerosis and systemic lupus erythematosus. Initial 28-item fatigue questionnaire was reduced based on the results of a factor analysis, item analysis and theoretical considerations – unclear what these considerations were. Five judges sorted items, without labels, into domains. No conceptual model of fatigue was reported. Patients were included in the study as research participants only, and not as research partners.</p> <p>Measure made up of 9 items and an additional ‘global fatigue’ VAS.</p> <p>Response options 7 response options ranging from 1-7 with anchors “strongly disagree” to “strongly agree”.</p> <p>Recall period Past week</p>	<p>Global score calculated by summing item scores and then averaging.</p>
<p>Single-item fatigue measures (3/9)</p>		
<p>Brief Fatigue Inventory (BFI)</p> <p>Worst Fatigue Numeric Rating Scale (WF-NRS)</p> <p>Mendoza (1999) (30)</p> <p><i>Completion format:</i> self-completed</p>	<p>The BFI was developed to assess fatigue in cancer patients and the impact over the past day. The BFI was based on the Brief Pain Inventory (BPI). Questionnaire items were modified following revision of “a fatigue questionnaire” completed by cancer patients and healthy controls in a previous study. Partial conceptualisation was provided in the explanation of item revision – but the information was limited. Questions seek to investigate: fatigue severity, interference with function; factors that worsen fatigue; and contributing factors to fatigue. It consists of ten items across two subscales: Fatigue Severity (4 items); and Fatigue Impact (6 items). Patients were involved as participants but there was no evidence reported that they were involved as research partners.</p> <p>The WF-NRS is a single-item on fatigue severity.</p> <p>Response options WF-NRS (item 3) ranges from 0-10 with anchors ranging from “no fatigue” to “as bad as you can imagine”.</p>	<p>Score taken from the value marked on the VAS.</p>

<p><i>Completion time:</i> approx.. <2 mins.</p>	<p>Recall period Past 24 hours</p>	
<p>Bath AS Disease Activity Index (BASDAI)</p> <p>10cm VAS</p> <p>Garrett (1994) (31)</p> <p><i>Completion format:</i> self-completed</p> <p><i>Completion time:</i> approx.. <2 mins.</p>	<p>The BASDAI is an AS-specific measure of disease activity. Development was driven by a team of physiotherapists, research associates and rheumatologists. Patients were used as participants to complete a pilot version of the questionnaire, but not as research partners. No conceptual model of fatigue was reported. The six items assess pain (severity), fatigue/tiredness (severity), stiffness (duration and severity) and tenderness (severity).</p> <p>Response options 10cm visual analogue scale with anchors “none” and “very severe” at either end of the scale.</p> <p>Recall period 1 week</p>	<p>Score taken from the value marked on the VAS.</p>
<p>Modified 10cm VAS</p> <p>Wheaton (2010) (32)</p> <p><i>Completion format:</i> self-completed</p> <p><i>Completion time:</i> approx.. <2 mins.</p>	<p><i>See BASDAI 10cm VAS</i></p> <p>Response options 10cm visual analogue scale with anchors “no problem” and “very severe” at either end of the scale.</p> <p>Recall period Not specified – assumed 1 week (as per original BASDAI single-item on fatigue severity).</p>	<p>Score taken from the value marked on the VAS.</p>
<p>Fatigue-specific PROM subscale (1/9)</p>		
<p>Short Form 36-item Health Survey (SF-36)</p> <p>SF-36 vitality subscale</p> <p>Ware (1992) (29)</p> <p><i>Completion format:</i> self-completed</p> <p><i>Completion time:</i> approx.. <5 mins.</p>	<p>The SF-36 is a generic measure of health status, containing 36-items across eight health domains: physical functioning, role limitations due to physical health or emotional problems, energy, pain, emotional wellbeing, social role functioning and general health. Most of the items included were taken from established measures. The content of these were reviewed with a view to assign the content to the pre-defined domains – these were informed by data from a previous iteration of the short form health survey (SF-20). Patients were not involved as research partners in the development process.</p> <p>Vitality subscale made up of 4 items (2 positively, 2 negatively phrased).</p> <p>Response options 6 response options with anchors “none of the time” and “all of the time”.</p> <p>Recall period Last 4 weeks</p>	<p>Positively phrased questions must be reverse scored. Scores then summed for all 4 items.</p>

Table 2: Data synthesis ^a, levels of evidence and overall quality ^b of reviewed PROMs (n=9)

PROM / Study	Number of evaluations	Reliability		Construct Validity		Responsiveness		Interpretation
		Internal consistency	Reliability	Hypothesis testing	Known-groups	Responsiveness	Other	
Multidimensional fatigue measures (3/9)								
MAF	7				? Unknown			
MFI-20	2	+ Limited	+ Limited	? Unknown	? Unknown		<i>ES, SRM, Guyatt</i>	
MFSI-SF	1			? Unknown	? Unknown			
Unidimensional fatigue measures (2/9)								
FACIT-fatigue	2	? Unknown		+ Moderate		? Unknown		
FSS	3			? Unknown	? Unknown		<i>ES, SRM</i>	
Single-item fatigue measures (3/9)								
BFI WF NRS	1	<i>Review of practical properties only</i>						
10cm VAS (BASDAI)	18		- Limited	+ Moderate	? Unknown	+ Limited	<i>ES, SRM, Guyatt</i>	? Unknown
Modified 10cm VAS	1							? Unknown
Fatigue-specific PROM subscale (1/9)								
SF-36 (vitality domain)	10	+ Moderate		+ Limited	? Unknown		<i>ES, SRM</i>	

Footnote:

^a Data synthesis: Qualitative synthesis of the data determined the quality and acceptability of each reviewed PROM. The synthesis considered four factors:

- 1) Study methodological quality (COSMIN scores)
- 2) Number of studies reporting evidence of measurement properties (per PROM)
- 3) Results for each measurement property (per PROM)
- 4) Consistency of results between studies.

The results of data synthesis include two ratings:

- 1) Overall measurement property quality: adequate (+), not adequate (-), conflicting (+/-), or unclear (?).
- 2) Levels of evidence for the overall quality per measurement property. Five outcomes were defined: 'strong' – consistent findings in multiple studies of good methodological quality OR in one study of excellent quality; 'moderate' – consistent findings in multiple studies of fair methodological quality OR in one study of good methodological quality; 'limited' – one study of fair methodological quality; 'conflicting' – conflicting findings; or 'unknown' evidence – only studies of poor methodological quality (18, 23).

b Overall quality: There was no measurement evidence available for the following measurement properties and they are therefore not referred to in the synthesis table: measurement error, content validity and structural validity.

Table 3: Item content of the reviewed single- and multi-item fatigue measures (n=9): item content distribution as per the Hewlett conceptual model of RA fatigue (7)

PROM	Concepts of RA Fatigue								
	Disease-specific	Cognitive, behavioural			Personal				Symptoms
		Behaviour	Cognition	Emotion	Support	Health	Environment	Responsibilities	
Multi-dimensional fatigue measures (3/9)									
MAF		Physical impact (2) – ability to do chores in house, exercise. Physical leisure activities (2) – sex, recreational activities. Physical effect on daily activities (5) – cook, bathe, dress, walk, shop/errands.		Emotional impact (1) – distress.				Interference with social life (1) and work (1).	Severity (1) Frequency (1) General (1) – extent of experiencing fatigue.
MFI-20		Limitation (1) – physically feel I can only do a little. Capability (1) – physically I feel I can take a lot on. Activity level - (1) - I get little done	Cognition (4) – focus, concentration. Forethinking (1) – lots of plans. Motivation (1) – don't feel like doing anything.	Anxiety (1) – dread.		Self-perception (10) - I am rested, physically I am in bad condition, I tire easily, physically I am in excellent condition.			
MFSI-SF			Memory (2) – trouble remembering, forgetful. Concentration/ focus (4) – confused, paying attention, unable to concentrate, making mistake, forgetful.	Mood (9) – upset, cheerful, nervous, relaxed, sad, depressed, tense, calm, distressed.		Symptom manifestation (12) – aching, weak, tired, heavy, pooped, fatigued. Self-perception (3) – feel: lively, refreshed, energetic.			
Uni-dimensional fatigue measures (2/9)									
FACIT-fatigue		Ability (1) – able to do usual activities. Impairment (1) – I need help to do my usual activities. Social activity (1) –	Motivation (2) – trouble starting, trouble finishing.	Mood (1) – frustrated by being too tired.		Impact (2) – need to sleep, too tired to eat. Symptom manifestation (4) – fatigued, weak, listless, tired.			

		have to limit because tired.				Energy (1) – I have energy.			
FSS		Impact (2) – interferes with functioning, prevents sustained physical functioning.	Motivation (1) – reduced motivation. Reflection (1) – most disabling symptom.					Interference (1) – work, family or social life.	Generic (1) – easily fatigued? Global perception (1) – what number best reflects global fatigue Frequency (1) – causes frequency problems Causes of fatigue (1) - exercise brings on fatigue
Single-item fatigue measures (3/9)									
BFI WF-NRS									Severity (1)
10cm VAS									Severity (1)
Modified 10cm VAS									Severity (1)
Fatigue-specific PROM subscale (1/9)									
SF-36 VT (vitality subscale)						Sense of energy/fatigue (4) – full of pep? A lot of energy? Feel worn out? Feel tired			