Hewlett, S., Hehir, M., & Kirwan, JR. (2007). Measuring fatigue in rheumatoid arthritis: a systematic review of scales in use. *Arthritis Care and Research*, *57*(3), 429-439. https://doi.org/10.1002/art.22611

Measuring fatigue in rheumatoid arthritis: a systematic review of scales in use

S Hewlett,[1] M Hehir,[2] JR Kirwan[2]

[1]Faculty of Health and Social Care, University of the West of England, Bristol, UK

[2]University of Bristol Academic Rheumatology Unit, Bristol, UK

Corresponding author

Sarah Hewlett PhD MA RGN

arc Reader in Clinical Nursing, University of the West of England

Rheumatology Unit

Bristol Royal Infirmary

BRISTOL BS2 8HW

UK

Tel: 44 (0)117 928 2903

Fax: 44 (0)117 928 3841

Email: Sarah.Hewlett@uwe.ac.uk

Running title: Measuring fatigue in rheumatoid arthritis

Key words: Fatigue; rheumatoid arthritis; tiredness; scale; questionnaire; outcome

Word count: 3780

Abstract

Objective: Fatigue is an important outcome for patients with rheumatoid arthritis (RA). The aim of this study was to identify the scales being used to measure RA fatigue, and systematically examine the evidence for their validation.

Methods: Articles measuring fatigue in RA were searched for using the terms RA and fatigue, and RA and tiredness, plus scale, questionnaire, inventory and checklist. Index papers reporting identifiable RA fatigue data were examined for the fatigue scale used. Index and validation papers for each scale were reviewed for evidence supporting scale validation to measure RA fatigue using a standardized checklist of content, face, criterion and construct validity, reliability and sensitivity to change.

Results: 61 index papers used 23 different fatigue scales to measure RA fatigue on 71 occasions. Seventeen scales had either no data on validation in RA, or limited evidence. Reasonable evidence of validation was identified for 6 scales, each also having some evidence of sensitivity to change: Ordinal scales; the SF-36 vitality subscale; the Functional Assessment of Chronic Illness Therapy Fatigue Scale; Visual Analogue Scales, the Profile of Mood States, and the RA-specific Multi-dimensional Assessment of Fatigue scale. However, the 4 generic scales would benefit from further validation in RA patients, the VAS requires standardization, and the MAF would benefit from further sensitivity data.

Conclusion: It was possible to identify evidence of reasonable validation for 6 out of 23 scales being used to measure RA fatigue. Researchers and clinicians should select scales to measure RA fatigue carefully.

Background

Rheumatoid arthritis (RA) is an auto-immune, systemic, inflammatory condition causing pain, disability and psychological distress.[1] Fatigue is experienced by up to 90% of patients with RA and its causality is likely to be multi-dimensional.[2-4] Fatigue has far-ranging consequences on patient's lives and is an important outcome for many patients,[5-8] but is currently not among the 7 internationally agreed core outcome measures in RA clinical trials.[9] However, the OMERACT group (Outcome Measures in Rheumatology Clinical Trials) has been discussing the importance of fatigue as a key outcome and has given a mandate to pursue further work on identifying valid measures.[10]

If treatments are to be developed and tested, fatigue needs to be accurately assessed using scales with adequate validation properties of comprehensiveness, accuracy, biological sense, reliability and sensitivity to change. In common with many other chronic illnesses, there is no agreed definition of fatigue in RA. However, it is widely accepted by clinicians that there is a subjective element to RA fatigue that goes beyond physiological muscle fatigue. Qualitative studies suggest RA fatigue may incorporate not only physical but also cognitive and emotional elements.[2,5] The nature of fatigue as experienced by RA patients may be different to that in other long term conditions,[5] therefore the application of generic scales, or the creation of new scales without patient involvement may have limited validity. Generic scales may contain items that in RA could reflect inflammatory disease or disability rather than fatigue (as seen in some generic depression scales).[11]

The use of inappropriate or unvalidated scales for outcomes can result in unreliable or misleading results. The Association of Rheumatology Health Professionals (ARHP) has recently reviewed a wide range of arthritis measurement scales but this extensive and valuable resource was of necessity limited by space constraints and was not intended to examine the specific validity for scales to measure RA fatigue.[12] The aim of the present

study was therefore to systematically identify scales that have been used to measure fatigue in RA in published studies, and examine the evidence for their validity to measure RA fatigue against recognized criteria.[12-14]

Methods

Systematic identification of scales: A systematic search for articles was undertaken to identify all the fatigue scales that are being used in RA. In the absence of an internationally agreed definition of RA fatigue, the two global descriptors 'fatigue' and 'tiredness' were used in conjunction with 'rheumatoid arthritis'. Searches were performed for both alongside 1) scale, 2) questionnaire, 3) inventory and 4) checklist (keywords, English language). These 8 searches were applied to 5 databases: Medline (Medicine, Dentistry and Nursing) from 1966, EMBASE (Medical) and PsychINFO (Psychology) from 1980, CINHAL (Cumulative Index to Allied Nursing and Health) from 1982, and AMED (Allied and Complementary Medicine Database) from 1985 (final search date 9th February 2004). The 'grey' literature of unpublished work was not searched as it has been shown to add little of value, and the studies may be of lesser quality.[15]

Two researchers (SH, MH) independently reviewed all abstracts for the inclusion criteria of identifiable RA fatigue data. All abstracts fitting these criteria were retained, plus any abstracts where it was unclear. The full papers of these selected abstracts were obtained (index papers) and reviewed to identify the fatigue scales used. Having systematically attempted to identify the fatigue scales being used in RA, information on their validation for use with RA patients was sought by reviewing the data in the index papers, data in the validation papers cited in the methods section of the index papers, and from any relevant references listed in the index and validation papers. For 6 scales that showed reasonable validity for measuring RA fatigue, the databases were later searched again for additional

validation data in RA using the term rheumatoid arthritis and the scale name (final search date 17 January 2005).

Evaluation of validation of scales: Index and validation papers for each scale were reviewed independently (SH, JK, MH) for evidence supporting scale validation to measure fatigue in RA and where further validation papers were referenced these were also obtained and reviewed. Evidence for face, content, criterion, and construct validity, reliability (internal consistency and stability) and sensitivity to change, was assessed using a checklist based on the methods of Katz,[12] Tugwell and Bombardier[13] and the added OMERACT filter of feasibility.[14] The parameters for judging scale validity are presented in Table 1. Construct validity aimed to look for moderate associations with appropriate variables (eg pain, mood) but not strong associations as this might reflect an inability of the scale to differentiate between fatigue and these variables. Criterion validity was assessed when there was comparison with any other RA fatigue measure, as there is no agreed 'gold standard' RA fatigue measure. After independent assessments using this standardized approach (Table 1), the three assessors discussed their findings and, taking all the papers relating to each scale into account, scored each scale as to the strength of the available evidence for each of the validation concepts, to produce validation summaries (0 = not reported or no evidence, 1 = limited evidence, 2 = moderate, 3 = good). Some instruments were well-validated generic fatigue scales, but were assessed specifically for their evidence that the scale was valid for measuring fatigue in RA.

Results

The searches identified abstracts from 166 initial papers containing the search terms, 59 of which it appeared might provide separately identifiable RA fatigue data in the full paper. Two further RA fatigue papers known to the authors were not identified by the searches (one used a scale acronym rather than the word 'scale,' one did not mention fatigue in the

abstract), giving a total of 61 index papers. Upon review of the full articles, 50/61 index papers reported identifiable RA fatigue data. Twenty-three different fatigue scales were used with some studies utilizing more than one scale, giving 71 occasions on which fatigue was measured (Table 2). Validation references were cited on only 35 occasions (49%). Overall, 118 papers were reviewed in the search for validation data, but only papers contributing substantial evidence are reported here.

Fatigue scales where limited evidence of validation for RA could be identified:

On systematic examination of the available validation data, for 7 out of the 23 scales being used to measure fatigue in RA either few validation studies could be identified, or evidence of validation was limited, or the scales had been designed for use in other populations and did not perform well in RA (Table 3).

Seven scales had been created for single RA studies and did not appear to have previously been validated (Table 3). Binary questions can only indicate the presence or absence of fatigue although they may be useful as a screening question.[16,17] A measure of 'Fatigue hours' is sensitive to change in a randomized trial of NSAIDs, but the paper does not describe the question nor provide validity or reliability data.[18] The 'Five items modeled after Tack' scale suggests evidence of sensitivity to change in a cognitive-behavioural therapy intervention (CBT) but no criterion or reliability data are provided.[19] The Morning Fatigue scale[20] may reflect the patient's experience of waking unrefreshed[5] but no validation data is provided. Numerical rating scales have some evidence of construct validity but no data were identified on criterion validity, reliability or sensitivity.[6,21]

Eight generic scales used to measure fatigue had limited data on validation in RA (Table 3). Some scales had been designed to measure cognition or personality. For example, three of the 11 items on the Chalder Fatigue Scale (CFS) address mental clarity,[22] and the scale does not differentiate between RA patients and controls.[23] Three of the 8 items on the Checklist of

Individual Strengths Subjective Fatigue subscale (CIS/SF) may reflect inflammation rather than fatigue (feeling fit, in good shape, in bad condition).[24-26] The CIS/SF demonstrates sensitivity to change in its single study with RA patients undergoing CBT[27] but has not been tested for reliability, construct or criterion validity in RA. The Chronic Fatigue Index (CFI) was developed from interviews in chronic disease and reviewed by women with RA, showing content validity. However, it was not re-tested after item reduction, and little evidence of testing for sensitivity, construct or criterion validity could be identified.[28] The Composite Index of Fatigue Impairment (CIFI) is an 11 point numerical rating scale plus the Nottingham Health Profile energy subscale, but how these are combined is not explained,[29] and the validation reference cited is inappropriate. The Feeling Tone Checklist (FTC) was validated in healthy airforce personnel in 1956, which limits its applicability to RA fatigue, while the single RA study which tested it alongside an activity record, does not report any data (validation reference and scale unobtainable).[30]

Generic fatigue scales can include items that might lead to contamination in RA from outcomes such as disability. The Multi-dimensional Fatigue Inventory (MFI) contains such items ('Physically I feel only able to do a little,' 'Physically I am in bad condition').[31] It does not differentiate between people with RA and ankylosing spondylitis on four of its 5 scales,[32] nor between RA patients and healthy controls on two.[33] The Nottingham Health Profile (NHP) is a well-validated tool for surveying population health[34] and although RA studies show the energy subscale has construct validity, stability and differentiates between RA and other populations, sensitivity to change and internal consistency have not been thoroughly examined in RA.[23,35-39] The 3 item energy subscale contains one item on fatigue, the second concerns energy (an absence of fatigue may not translate to the presence of energy) and the third concerns effort, which in RA may be the result of disability. The Psychasthenia scale is

a personality scale and only 3/10 items assess physical fatigue. Whilst it differentiates between RA and fibromyalgia patients, no other validation data could be identified.[40]

Two RA-specific scales have limited data on validation. The National Institutes of Health Activity Record (NIH ACTRE) was developed to measure outcomes of an energy conservation programme and is reported as having no standard scoring system and did not show sensitivity to change.[30,41] The 'Time to onset of fatigue' question showed that fatigue did not improve in an NSAID trial but little information on construct or criterion validity could be found.[42,43]

Only two of these 17 scales have evidence for stability in RA or have been tested against another fatigue measure and only 5 have been used in RA intervention studies. On the limited evidence available to date, it seems uncertain whether the 17 scales in Table 3 would be the first choice for providing robust, valid, accurate and sensitive measures of fatigue in patients with RA.

Scales with reasonable evidence of validation for measuring fatigue in RA

Multi-dimensional Assessment of Fatigue scale (MAF): The MAF comprises 16 questions concerning the quantity, degree, distress, impact and timing of fatigue.[44] Questions 1-15 form the final score (Global Fatigue Index, 0-50) while question 16 concerns change over the past week. Questions 1-14 are 10 point items, while 15 and 16 are four point items. The MAF is an RA-specific revision of the Piper Fatigue Scale developed in oncology, giving it face and content validity (Table 4) and it has been tested against another fatigue scale.[44,45] Construct validity is shown through moderate convergence with disease activity and mood (0.45-0.54).[44] Higher MAF scores are related to increased depression and reduced sleep, with 61% of variance in fatigue explained by disease activity, gender and psychosocial status.[44,46] Higher MAF scores differentiate between RA patients with and without previous depression, between different levels of disease activity and between patients and controls.[44,47-49]

Reliability is reflected by good internal consistency (inter-item correlations 0.53-0.83; Cronbachs alpha 0.91-0.96).[44,46-48] Sensitivity to change is shown following drug therapy[50] and after exercise, where the MAF showed initial worsening of fatigue followed by improvement beyond baseline, and showed a difference between low and high exercisers.[45] Although designed to be scored as a global fatigue scale, there is some evidence for the validity of the individual sections of the MAF.[44,45]

Functional Assessment of Chronic Illness Therapy fatigue scale (FACIT-F): The index paper using the FACIT in RA did not report fatigue data.[51] The paper reporting RA validation was published shortly after the final search date but is nonetheless reviewed here.[52] The FACIT-F is a 13 item scale originally developed to measure fatigue in cancer patients.[53,54] In RA, the FACIT-F shows convergent validity with disease activity, good internal consistency (cronbach's alpha 0.86-0.87), and evidence of sensitivity to change (effect size 0.19-1.13).[52] However, no information on divergent validity nor stability in RA is reported. The source of the scale items was oncology patients and in RA, several items may be confounded by disability (eg needing help to do usual activities). One item measures energy rather than fatigue and items applicable to cancer patients may hold less relevance for RA patients, for example feeling too tired to eat is not reported in qualitative RA fatigue studies.[2,5]

Ordinal scales: Three studies measured RA fatigue using ordinal scales such as 'none' to 'very severe' with responses ranging from 4-7 points.[6,55,56] Overall, ordinal scales showed reasonable content and construct validity (Table 4), differentiating between RA patients with and without inflammation, and showing association with other symptoms, reduced perceived ability to cope with fatigue, and poor sleep. [6,55,56] When measured 7 times per day over 7 days, the ordinal scale appeared stable, and consistently showed fatigue was least at noon and

worst in the evenings, supporting the ability of ordinal scales to capture variation in fatigue.[55] However, there is no data from intervention studies.

Profile of Mood States (POMS): The POMS[57] fatigue/inertia scale contains 7 items (worn out, listless, fatigued, exhausted, sluggish, weary, bushed). It was designed to measure mood, but may address some of the cognitive elements and overwhelming fatigue raised by RA patients.[5] The POMS has criterion validity in RA (MAF r=0.84),[48] although little information on construct validity could be identified. The POMS differentiates between RA patients with and without a fear of falling and has good internal consistency (cronbach's alpha 0.88).[58] Three dance or exercise-based interventions showed change in POMS, although this only approached significance[45,59,60] while in one of these the MAF did show significant change.[45]

SF-36 Vitality subscale: The vitality subscale of the SF-36 comprises 4 items (full of life, energy, worn out, tired) with 6 responses from 'all' to 'none of the time.[61] Many studies report RA data.[52,61-75] Three studies report RA patients have less vitality than healthy controls, suggesting good construct validity,[23,66,67] but another study reports RA patients have more vitality than controls.[63] Although the vitality scale is associated with measures of disease activity, correlations with ESR vary (r=0.18-0.34) and the vitality scale does not always reflect change in ESR.[62,65,67] The SF-36 vitality subscale is shown to correlate strongly with mood in one study (r=0.67)[65] suggesting it may not differentiate well between fatigue and depression. The SF36 shows good internal consistency (Cronbach's alpha 0.84-0.88) and only a small ceiling and floor effect (Table 4).[63-65,72] Sensitivity to change is shown with biologic agents (effect size 0.25-1.52).[52,68]

Visual Analogue Scales: VAS with identifiable RA fatigue data were used in 22 index papers and 4 times in papers obtained during the search for validation data.[3,23,74-96] Validation references for a fatigue VAS were cited on only 4/26 occasions and referred to

overviews of the ability of VAS to measure pain.[97] Only 10/26 VAS were described sufficiently for them to be fully reproduced and only 3 could be identified as completely identical for descriptors, timescale and length (Table 5). Lack of standardization clearly limits comparison between studies.

Where the scales were described, VAS face and content validity were good (Table 6) except that information was rarely given as to whether patients were the source of the descriptors. Construct validity is reasonable with evidence of convergent validity with pain (r=0.31-0.8), poor sleep (r=0.6) and disability (r=0.33-0.41).[77,79,80,88] VAS fatigue is moderately associated with low mood (r=0.41-0.47).[77,80,84] VAS for fatigue discriminate between patients with RA and fibromyalgia, RA patients and healthy controls, and RA patients with and without pain.[23,76,79,91,92] Overall, available evidence for reliability is limited, with one study showing stability for the VAS but not the SF-36, and another showing the opposite.[89,90] Of 3 intervention studies, one only reported combined OA/RA data,[86] and neither total hip replacement[88] nor an energy conservation program[96] showed change in RA fatigue. Three longitudinal studies show variation in a fatigue VAS over time.[87,89,90]

One paper using a VAS to measure fatigue stated that for scanning purposes, it was formatted as 21 boxes representing a 0-10 scale in 0.5 increments (ClinHAQ).[98] The reduction of a continuous VAS line to a tick-box scale of 21 discrete points may alter its measurement properties. It is not stated in their earlier VAS studies whether it was presented in box or standard VAS format.[3,82,83]

Some studies used several scales simultaneously. A comparison of the VAS, MAF and SF36 vitality subscale in a large RA cohort yielded mean scores of 4.5, 5 and 5.5 (out of 10).[98] The scales correlated well with each other (r=0.71-0.8) and moderately with clinical measures (r=0.5-0.63). The MAF had the smallest floor and ceiling effects at 0.01% and 0.2% (SF36 0.4% and 3.4%; VAS 6.4% and 1.8%). The VAS had the most variability with a SE of 0.032

(MAF and SF36, 0.026). However, the VAS was presented as 21 boxes, the MAF and SF36 were rescaled, and associations with mood were not assessed.

Other studies reporting data from both a fatigue VAS and the SF36 vitality subscale show slightly lower associations (r=-0.58 to –0.71) and suggest the scales have different distributions and behave differently.[80,89,90] The SF36 vitality subscale was more strongly associated with social and mental health, and unlike the fatigue VAS, varied with age. Although one longitudinal study suggested the SF36 and VAS change in similar ways,[87] other studies report that changes in fatigue were sometimes shown on the VAS but not the SF36, and vice versa.[89,90] These differences may reflect the conceptual gap between measuring fatigue and vitality, as a low score on one may not necessarily reflect a high score on the other.

Discussion

The use of poorly validated outcome scales would limit the interpretation of study results, yet this systematic review of scales used to measure RA fatigue shows that for 17 of the 23 scales used, only limited validation could be identified. Many had been created for use in a single study, while for many others, little evidence of attempts at validation in RA could be found. Some of these scales for measuring RA fatigue may be useful and appropriate, but in the absence of identifiable validation data this is difficult to assess. The MAF, SF36, FACIT-F, ordinal scales, POMS and VAS do show evidence of validation for measuring RA fatigue. However, even these 6 scales could benefit from further research, particularly into content validity for RA patients, the inclusion of cognitive items, and sensitivity to change.

This study raises several important issues. First, if authors do not give a validation reference for a scale, or do not fully describe it, it is difficult for readers to assess or replicate their methodology. This is particularly problematic with the VAS where only 10/26 papers fully described the wording, timescale and length. Whilst the combined data from many RA

studies supports reasonable validation for a fatigue VAS, there is no standardized VAS for RA fatigue, making comparison between studies difficult. The most common VAS wording was 'problem' but this infers a combination of severity and consequence rather than pure level of fatigue. There is a pressing need to develop and validate a standardized RA fatigue VAS.

Second, many of these scales were developed ahead of the current systematic, rigorous approach to reporting validation data, making it difficult to assess their validation. In particular, the inclusion of patient opinion into scale development is now considered crucial. The lack of sensitivity to change data is a major drawback, and is complicated by the lack of knowledge about which interventions might reduce RA fatigue. A circular argument develops whereby it is difficult to validate scales for sensitivity until there are effective interventions, and it is difficult to prove efficacy of interventions without validated scales. In RA, it is possible that interventions such as exercise may not improve fatigue, but they may allow patients to do more without increasing fatigue, an additional complication. If future intervention studies could simultaneously measure other aspects of change, such as changing levels of activity or impact of fatigue, this might help clarify a difficult area.

Third, some generic scales could be open to contamination as in RA the wording may reflect problems arising from disability or inflammation rather than fatigue. Some scales include phrases that are not reflected in qualitative studies of the nature of RA fatigue (eg being too tired to eat) and most omit concepts that may be specific to RA fatigue (eg sudden, overwhelming onset, or cognitive elements).[2,5] Several scales measure energy or vitality, but an absence of fatigue does not necessarily mean the presence of energy, therefore such scales may not accurately reflect fatigue. Indeed, the POMS treats fatigue and vigor as different concepts and shows they are only moderately associated (r=0.43) and behave differently in their relationships with mood, suggesting they are not opposite ends of a single concept.[57]

This study has limitations, which include the lack of an internationally agreed definition of RA fatigue, which is outside the scope of this study. Searching databases for RA and fatigue/tiredness produced an unmanageable dataset, therefore as the aim of the study was to identify fatigue measures, the terms scale, questionnaire, inventory and checklist where used. It is therefore possible some RA fatigue studies were missed that might have contributed validation data. However a second search was conducted specifically for the measurement of RA fatigue using the 6 scales identified as having stronger validation. The review scoring system was relatively arbitrary but the aim was to create a broad summary of available evidence, using criteria based on a recognized validation framework,[12-14] systematically applied by three researchers first independently and then in discussion, to arrive at agreement. The treatment and self-management of fatigue in RA is receiving increasing interest as its importance to patients becomes apparent and this requires accurate measurement of fatigue. This study helps clarify a difficult issue by identifying stronger validation in certain fatigue scales. However it also indicates that further work is urgently required to develop and validate a standardized VAS, to consider the addition of cognitive fatigue items to the MAF, to compare questionnaire items to qualitative descriptions of RA fatigue to ensure they capture the essence of the concept in RA, and to provide more sensitivity to change data for several scales. A measure of the impact of fatigue, in addition to the presence or severity of fatigue would be useful.

Acknowledgements

SH designed the study, MH managed the study, and SH, JK and MH each reviewed all papers. SH was principle investigator and all authors commented on and contributed to drafts of the paper. The authors would like to thank the Charitable Trustees of the United Bristol Healthcare Trust for funding MH, and the **arc** for funding SH's academic post.

References

1 Conaghan PG, Green MJ, Emery P. Established rheumatoid arthritis. Baillière's Clinical Rheumatology 1999; 13: 561-575

2 Tack BB. Fatigue in Rheumatoid Arthritis: Conditions, strategies and consequences. Arth Care & Research 1990; 3: 65-70

3 Wolfe F, Hawley DJ, Wilson K. The prevalence and meaning of fatigue in rheumatic disease. J Rheumatol 1996; 23: 1407-17

4 Belza BL, Henke CJ, Yelin EH, Epstein WV, Gilliss CL. Correlates of fatigue in older adults with rheumatoid arthritis. Nursing Research 1993; 42: 2: 93-99

5 Hewlett S, Cockshot Z, Byron M, Kitchen K, Tipler S, Pope D, Hehir M. Patients' perceptions of fatigue in rheumatoid arthritis: Overwhelming, uncontrollable, ignored. Arth Rheum (Arth Care & Research) 2005; 53: 697-702

6 Katz P. The stresses of rheumatoid arthritis: Appraisals of perceived impact and coping efficacy. Arth Care & Research 1998; 11: 9-22

7 Minock P, Bresnihan B. Pain outcome and fatigue levels reported by women with established rheumatoid arthritis. Arth Rheum 2004; 50: 9: 1197

8 Carr A, Hewlett S, Hughes R, Mitchell H, Ryan S, Carr M, Kirwan J. Rheumatology outcomes: the patient's perspective. J Rheumatol 2003, 30: 880-883

9 Felson DT, Anderson JJ, Boers M, Bombardier C, Chernoff M, Fried B, et al. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. Arth Rheum 1993; 36: 729-40

10 Kirwan JR, Hewlett S, Heiberg T, Hughes R, Carr M, Hehir M et al. Incorporating the patient perspective into outcome assessment in rheumatoid arthritis – progress at OMERACT 7. J Rheumatol 2005; 32: 2250-2256

11 Pincus T, Callahan LF. Depression scales in rheumatoid arthritis: criterion contamination in interpretation of patient responses. Patient Ed and Counsel 1993, 20: 133-143

12 Katz PP (Ed). Patient outcomes in rheumatology. Arth Rheum (Arth Care & Research) 2003; 49: 5 supplement

13 Tugwell P, Bombardier C. A methodologic framework for developing and selecting endpoints in clinical trials. J Rheumatol 1982; 9: 5: 758-762

14 Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter in outcome measures for rheumatology. J Rheumatol 1998; 25: 198-199

15 Egger M, Juni P, Bartlett C, Holstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Health Technology Assessment 2003; 7:1

16 Hill J, Bird H, Thorpe R. Effects of rheumatoid arthritis on sexual activity and relationships. Rheumatology 2003; 42: 2: 280-286

17 Drewes AM, Jennum P, Andreasen A, Sjol A, Nielsen K. Self-reported sleep disturbances and daytime complaints in women with fibromyalgia and rheumatoid arthritis. J Musculoskeletal Pain 1994; 2: 4: 15-31

18 Meinicke J, Danneskiold-Samsoe B. Diclofenac sodium (voltaren) and Ibuprofen in rheumatoid arthritis. A randomized double-blind study. Scand J Rheum 1980; 9: Suppl. 35: 3-8

19 Sinclair VG, Wallston KA, Dwyer KA, Blackburn DS, Fuchs H. Effects of a cognitive-behavioral intervention for women with rheumatoid arthritis. Res Nursing & Health 1998; 21: 4: 315-26

20 Choudhury A, Yunus B, Haq S, Alam M, Sebrina F, Adlag J. Clinical features of fibromyalgia syndrome in a Bangladeshi population. J Musculoskeletal Pain 2001; 9: (1): 25-33

21 Katz P, Pasch L, Wong B. Development of an instrument to measure disability in parenting activity among women with rheumatoid arthritis. Arth Rheum 2003; 48: 4: 935-943

22 Chalder T, Berelowitz G, Pawlikowska T, Watts L, Wessley S, Wright D, Wallace E. Development of a Fatigue Scale. J Psychosom Res 1993; 37: 2: 147-153

23 Lwin T-T, Bishay M, Platts R, Booth D, Bowman S. The assessment of fatigue in primary Sjogren's syndrome. Scand J Rheum 2003; 32: 1: 33-37

24 Vercoulen J, Swanink C, Fennis J, Galama J, van der Meer J, Bleijenberg G. Dimensional assessment of chronic fatigue syndrome. J Psychosom Res 1994; 38: 5: 383-392

25 Vercoulen J, Homes O, Swanink C, Jongen P, Fennis J, Galama J et al. The measurement of fatigue in patients with multiple sclerosis. Arch Neurol 1996; 53: 642-9

26 Beurskens A, Bűltmann U, Kant I, Vercoulen J, Bleijenberg G, Swaen G. Fatigue among working people: validity of a questionnaire measure. Occup Environ Med 2000; 57: 353-357

27 Evers A, Kraaimaat F, van Riel P, de Jong A. Tailored cognitive-behavioral therapy

in early rheumatoid arthritis for patients at risk: a randomized controlled trial. Pain 2002; 100: 141-153

28 McNamara A. Cognitive response to symptoms in women with rheumatoid arthritis. Doctoral dissertation, University Of Arizona 1992

29 Waltz M. The disease process and utilization of health services in rheumatoid arthritis: the relative contributions of various markers of disease severity in explaining consumption patterns. Arth Care & Res 2000; 13: 2: 74-88

30 Gerber L, Furst G. Validation of the NIH activity record: a quantitative measure of life activities. Arth Care & Res 1992; 5: 2: 81-86

31 Smets E, Garssen B, Bonke B, De Haes J. The Multidimensional Fatigue Inventory (MFI): Psychometric qualities of an instrument to assess fatigue. J Psychomatic Res 1995; 39: 5: 315-325

32 Chorus A, Miedema H, Boonen A, van der Linden S. Quality of life and work in patients with rheumatoid arthritis and ankylosing spondylitis of working age. Ann Rheum Dis 2003; 62: 1178-84

33 Barendregt P, Visser M, Smets E, Tulen J, van den Meiracker A, Boomsma F, Markusse H. Fatigue in primary Sjogren's syndrome. Ann Rheum Dis 1998; 57: 5: 291-295

34 Hunt SM, McEwen J, McKenna SP. Measuring health status. London: Croom Helm, 1986

35 Houssien D, McKenna S, Scott D. The Nottingham Health Profile as a measure of disease activity and outcome in rheumatoid arthritis. Br J Rheumatology 1997; 36: 69-73

36 Uutela T, Hakala M, Kautiainen. Validity of the Nottingham Health Profile in a Finnish out-patient population with rheumatoid arthritis. Rheumatology 2003; 42: 841-5

37 Sivas F, Erçin O, Tanyolaç Ő, Barça N, Aydoğ S, Őzoran K. The Nottingham Health Profile in rheumatoid arthritis: correlation with other health status measurements and clinical variables. Rheumatol Int 2004: 24: 203-206

38 Fitzpatrick R, Ziebland S, Jenkinson C, Mowat A, Mowat A. A generic health status instrument in the assessment of rheumatoid arthritis. Br J Rheumatology 1992; 31: 87-90

39 Smedstad L, Moum T Vaglum P, Kvien T. The impact of early rheumatoid arthritis on psychological distress. A comparison between 238 patients with RA and 116

matched controls. Scand J Rheumatology 1996; 25: 6: 377-382

40 Ekselius L, Bengtsson A, von Knorring L. Personality traits as determined by means of the Karolinska Scales of Personality in patients with fibromyalgia. J Musculoskeletal Pain 1998; 6: 2: 35-49

41 Furst G, Gerber L, Smith C, Fisher S, Shulman B. A program for improving energy conservation behaviors in adults with rheumatoid arthritis. Am J Occupational Therapy 1987; 41: 2: 102-111

42 McGuire RJ, Wright V. Statistical approach to indices of disease activity in rheumatoid arthritis. Ann Rheum Dis 1971; 30: 574-580

43 Lansbury J, Baier H, McCracken S. Statistical study of variation in systemic and articular indexes. Arth Rheum 1962; 5: 445-456

44 Tack BB. Dimensions and correlates of fatigue in older adults with rheumatoid arthritis. PhD dissertation, University Of California, San Francisco 1991

45 Neuberger GB, Press AN, Lindsley HB, Hinton R, Cagle PE, Carlson K et al. Effects of exercise on fatigue, aerobic fitness, and disease activity measures in persons with rheumatoid arthritis. Res Nursing & Health 1997; 20: 3: 195-204

46 Belza BL, Henke CJ, Yelin EH, Epstein WV, Gilliss CL. Correlates of fatigue in older adults with rheumatoid arthritis. Nursing Research 1993; 42: (2): 93-99

47 Jump RL, Fifield J, Tennen H, Reisine S, Giuliano AJ. History of affective disorder and the experience of fatigue in rheumatoid arthritis. Arth Rheum (Arth Care & Research) 2004; 51: 2: 239-245

48 Belza B. Comparison of self-reported fatigue in rheumatoid arthritis and controls. J Rheumatol 1995; 22: 4: 639-643

49 Gerber LH, Furst G, Yarboro C, El-Gabalawy H. Number of active joints, not diagnosis, is the primary determinant of function and performance in early synovitis. Clin & Exp Rheumatology 2003; 21: Suppl 31: S65-S70

50 Kaltwasser JP, Kessler U, Gottschalk R, Stucki G, Möller B. Effect of recombinant human erythropoietin and intravenous iron on anemia and disease activity in rheumatoid arthritis. J Rheumatol 2001; 28: 2430-2436

51 Weinblatt ME, Keystone EC, Furst DE, Moreland LW, Weisman MH, Birbara C et al. Adalimumab, a fully human anti-tumor necrosis factor α monoclonal antibody, for the treatment of rheumatoid arthritis in patients taking concomitant methotrexate. Arth Rheum 2003; 48: 1: 35-45

52 Cella D, Yount S, Sorensen M, Chartash E, Sengupta N, Grober J. Validation of the Functional Assessment of Chronic Illness Therapy Fatigue Scale relative to other instrumentation in patients with rheumatoid arthritis. J Rheumatol 2005; 32: 811-819

53 Cella D, Webster K. Linking outcomes management to quality of life measurement. Oncology 1997; 11: 11A: 232-235

54 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. J Pain Symptom Manage 1997; 13: 63-74

55 Stone A, Broderick J, Porter L, Kaell A. The experience of rheumatoid arthritis pain and fatigue: examining momentary reports and correlates over one week. Arth Care & Research 1997; 10: 3: 185-93

56 Pinals R, Masi A, Larsen R. Preliminary criteria for clinical remission in rheumatoid arthritis. Arth & Rheum 1981; 24: 10: 1308-1315

57 McNair D, Lorr M, Droppleman L. Profile of Mood States manual. Multi-health Systems Inc, New York, 1992

58 Jamison M, Neuberger G, Miller P. Correlates of falls and fear of falling among adults with rheumatoid arthritis. Arth & Rheum (Arth Care and Res) 2003; 49: 5: 673-680

59 Noreau L, Martineau H, Roy L, Belzile M. Effects of a modified dance-based exercise on cardiorespiratory fitness, psychological state and health status of persons with rheumatoid arthritis. Am J Physical Med & Rehab 1995; 74: 1: 19-27

60 Perlman SG, Connell KJ, Clark A, Robinson MS, Conlon P, Gecht M et al. Dance-based aerobic exercise for rheumatoid arthritis. Arth & Rheum (Arth Care and Res) 1990; 3: 1: 29-35

61 Ware JE, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36). Conceptual framework and item selection. Med Care 1992; 30: 6: 473-483

62 Kosinski M, Keller S, Ware JE, Hatoum HT, Kong SX. The SF-36 health survey as a generic outcome measure in clinical trials of patients with osteoarthritis and rheumatoid arthritis: Relative validity of scales in relation to clinical measures o arthritis severity. Med Care 1999; 37: 5: MS23-MS39

63 Tuttleman M, Pillemer SR, Tilley BC, Fowler SE, Buckley LM, Alarcon SG et al. A cross-sectional assessment of health status instruments in patients with rheumatoid arthritis participating in a clinical trial. J Rheumatol 1997; 24: 10: 1910-1915

64 Kosinski M, Keller S, Hatoum HT, Kong SX, Ware JE. The SF-36 health survey as a generic outcome measure in clinical trials of patients with osteoarthritis and rheumatoid arthritis: Tests of data quality, scaling assumptions and score reliability. Med Care 1999; 37: 5: MS10-MS22

65 Ruta DA, Hurst NP, Kind P, Hunter M, Stubbings A. Measuring health status in British patients with rheumatoid arthritis: reliability, validity and responsiveness of the Short Form 36-item health survey (SF36). Br J Rheumatology 1998; 37: 425-436

66 Bowman SJ, Booth DA, Platts GR. Measurement of fatigue and discomfort in primary Sjörgrens syndrome using a new questionnaire tool. Rheumatology 2004; 43: 758-764

67 Birrell FN, Hassell AB, Jones PW, Dawes PT. How does the Short Form 36 Health Questionnaire (SF36) in Rheumatoid Arthritis (RA) relate to RA outcome measures and SF36 population values? A cross-sectional study. Clin Rheumatol 2000; 19: 195-9

68 Russell AS, Conner-Spady B, Mintz A, Mallon C, Maksymowych WP. The responsiveness of generic health status measures assessed in patients with rheumatoid arthritis receiving infliximab. J Rheumatol 2003; 30: 941-947

69 Keller SD, Ware JE, Hatoum HT, Kong SX. The SF-36 Arthritis-Specific Health Index (ASHI): II Tests of validity in four clinical trials. Med Care 1999; 37: MS51-MS60

70 Loge HJ, Kassa S, Hjermstad MJ, Kvien TK. Translation and performance of the Norwegian SF-36 Health Survey in patients with rheumatoid arthritis I. Data quality, scaling assumptions, reliability, and construct validity. J Clin Epidemiol 1998; 51: 11: 1069-1076

71 Talamo J, Frater A, Gallivan S, Young A. Use of the Short Form 36 (SF36) for health status measurement in rheumatoid arthritis. Br J Rheumatology 1997; 36: 463-469

72 Salaffi F, Stancati A, Carotti M. Responsiveness of health status measures and utility-based methods in patients with rheumatoid arthritis. Clin Rheumatol 2002; 21: 478-487

73 Moss-Morris R, Chalder T. Illness perceptions and levels of disability in patients with chronic fatigue syndrome and rheumatoid arthritis. J Psychosomatic Res 2003; 55: 4: 305-8

74 Gilboe IM, Kvien TK, Husby G. Disease course in systemic lupus erythematosus: Changes in health status, disease activity, and organ damage after 2 years. J

Rheumatol 2001; 28: 2: 266-274

75 Gilboe IM, Kvien TK, Husby G. Health status in systemic lupus erythematosus compared to rheumatoid arthritis and healthy controls. J Rheumatol 1999; 26: 8: 1694-1700

76 Currey SS, Rao JK, Winfield JB, Callahan LF. Performance of a generic health-related quality of life measure in a clinic population with rheumatic disease. Arth Rheum (Arth Care & Research) 2003; 49: 5: 658-664

77 Tack BB. Self-reported fatigue in rheumatoid arthritis. Arth Care & Research 1990; 3: 154-7

78 Scharloo M, Kaptein A, Weinman J, Hazes J, Breedveld F, Rooijmans H. Predicting functional status in patients with rheumatoid arthritis. J Rheumatol 1999; 26: 8: 1686-93

79 Mengshoel AM, Førre Ø. Pain and fatigue in patients with rheumatic disorders. Clin Rheumatology 1993; 12: 4: 515-522

80 Kvien TK, Kaasa S, Smedstad LM. Performance of the Norwegian SF-36 Health Survey in patients with rheumatoid arthritis II. A comparison of the SF-36 with disease-specific measures. J Clin Epidemiol 1998; 51: 11: 1077-1086

81 Crosby L J. Factors which contribute to fatigue associated with rheumatoid arthritis. J Adv Nursing 1991; 16: 8: 974-981

82 Wolfe F, Skevington S. Measuring the epidemiology of distress: the Rheumatology Distress Index. J Rheumatology 2000; 27: 8: 2000-2009

83 Wolfe F, Pincus T, O'Dell J. Evaluation and documentation of rheumatoid arthritis disease status in the clinic: Which variables best predict change in therapy. J Rheumatology 2001; 28: 7: 1712-1717

84 Riemsma RP, Rasker JJ, Taal E, Griep EN, Wouters JMGW, Wiegman O. Fatigue in rheumatoid arthritis: the role of self-efficacy and problematic social support. Br J Rheumatology 1998; 37: 10: 1042-1046

85 Riemsma RP. Taal E, Rasker JJ, Houtman PM, Van Paassens HC, Wiegman O. Evaluation of a Dutch version of the AIMS2 for patients with rheumatoid arthritis. Br J Rheumatology 1996; 35: 8: 755-760

86 Barlow JH, Turner AP, Wright CC. Long-term outcomes of an arthritis self-management programme. Br J Rheumatology 1998; 37: 12: 1315-1319

87 Hagen KB, Smedstad LM, Uhlig T, Kvien TK. The responsiveness of health status measures in patients with rheumatoid arthritis: comparison of disease-specific and

generic instruments. J Rheumatol 1999; 26: 7: 1474-1480

88 Borstlap M, Zant JL, Van Soesbergen RM, Van Der Korst JK. Quality of life assessment: a comparison of four questionnaires for measuring improvements after total hip replacement. Clin Rheumatology 1995; 14: 1:15-20

89 Brekke M, Hjortdahl P, Kvien T. Self-efficacy and health status in rheumatoid arthritis: A two-year longitudinal observational study. Rheumatology. 2001; 40: 4: 387-392

90 Brekke M, Hjortdahl P, Kvien TK. Changes in self-efficacy and health status over 5 years: A longitudinal observational study of 306 patients with rheumatoid arthritis. Arth Rheum (Arth Care & Research) 2003; 49: 3: 342-348

91 Gudbjőrnsson B, Broman JE, Hetta J, Hällgren R. Sleep disturbances in patients with primary Sjögren's syndrome. BJ Rheumatology 1993; 32: 12: 1072-1076

92 Heiberg T, Kvien TK. Preferences for improved health examined in 1,024 patients with rheumatoid arthritis: Pain has highest priority. Arth Rheum (Arth Care & Research) 2002; 47: 4: 391-397

93 Jensen JL, Uhlig T, Kvien TK, Axéll T. Characteristics of rheumatoid arthritis patients with self-reported sicca symptoms: evaluation of medical, salivary and oral parameters. Oral Diseases 1997; 3: 4: 254-61

94 Uhlig T, Kvien TK, Glennås A, Smedstad LM, Forre Ø. The incidence and severity of rheumatoid arthritis. Results from a county register in Oslo, Norway. J Rheumatology 1998; 25: 6: 1078-1084

95 Chiang EP, Bagley P, Selhub J, Nadeau M, Roubenoff R. Abnormal vitamin B_6 status is associated with severity of symptoms in patients with rheumatoid arthritis. Am J Medicine 2003; 114: 4: 283-287

96 Gerber L, Furst G, Shulman B, Smith C, Thornton B, Liang M et al. Patient education program to teach energy conservation behaviors to patients with rheumatoid arthritis: A pilot study. Arch Phys Med Rehab 1987; 68: 442-445

97 Huskisson EC. Visual Analogue Scales. Pain measurement and assessment. Ed Melzack R, Raven Press, New York 1983

98 Wolfe F. Fatigue assessments in rheumatoid arthritis: Comparative performance of visual analog scales and longer fatigue questionnaires in 7760 patients. J Rheumatol 2004; 31: 1896-1902

Table 1: Validation checklist

Concept	Questions	Expectation of a good RA fatigue scale
Validity		
Face validity	Does the method appear sensible?	Language reflects patients' ideas of fatigue in RA[2,5]
Content validity	Is the source of the questions patient-based?	Evidence of patients as source or reviewer of items
	Are all appropriate items included?	Eg physical, emotional, cognitive, consequence, severity
	Are misleading items avoided?	Eg no items that might be confounded/confused with disability
Criterion validity	Is it compared to a 'gold' standard measure?	Tested against another fatigue scale
Construct validity	Does it show convergence with appropriate variables?	Eg moderate correlation with pain, inflammation, mood, anaemia
	Does it show divergence between different groups?	Eg between active/non-active disease; or RA patients/controls
Reliability		
Internal consistency	Is it internally consistent (multi-item scales only)	Inter-item correlation moderate or strong
Stability	Is the scale stable?	Unchanged in stable patients
Sensitivity to change	Is it sensitive to change?	Changes after intervention (eg drugs, surgery, exercise, education)
Feasibility	How long does it take to complete?	10-15 minutes maximum
	Is it self-report or interviewer administered?	Self-report, as appropriate for subjective scale
	How easy is it to score and interpret?	Clear instructions, or accessible computer programme/manual

Table 2 Fatigue scales identified as used in RA

Scale and subscale	Acronym	Description	References reviewed	RA studies Obs[1]	RA studies Int[2]
Binary question: Fatigue		Yes/No	16	1	
Binary question: Tiredness		Yes/No	17	1	
Chalder Fatigue Scale	CFS	11 items, 4 points	22, 23	1	
Checklist Individual Strengths: Subjective Fatigue	CIS/SF	8 items, 7 points	24-27		1
Chronic Fatigue Index	CFI	16 VAS (mean)	28	1	
Composite Index Fatigue Impairment	CIFI	11 point scale + NHP fatigue items	29	1	
Fatigue hours		Unknown	18		1
Feeling Tone Checklist	FTC	10 descriptors, unobtainable	30	1	
Five items modelled after Tack		Mean of 5 items	19		1
Functional Assessment of Chronic Illness Therapy: Fatigue scale	FACIT-F	13 items, rated 0-4	51-54		1
Morning fatigue		Lack vigour / fatigued on waking	20	1	
Multidimensional Assessment of Fatigue	MAF	15 items plus change question	44-50	5	3
Multidimensional Fatigue Inventory	MFI	5 scales of 4 items. 5 points	31-33	2	
National Institutes of Health Activity Record	NIH ACTRE	Log every 30 mins x 48 hours (0-4)	30, 41	1	1
Nottingham Health Profile: Energy subscale	NHP	3 weighted items	23, 34-39	6	
Numerical Rating Scale: Degree		0-5	21	1	
Numerical Rating Scale: Impact		0-10	6	1	
Ordinal degree scales*		Eg None, mild, moderate, severe	6, 55, 56,	3	
Profile of Mood States: Fatigue/inertia subscale	POMS	7 items	44-46, 57-60	3	3
Psychasthenia		10 items	40	1	
Short Form 36 Health Survey: Vitality Subscale	SF36	4 items	23, 52, 61-75, 80, 87, 89, 90	16	5
Time to onset of fatigue		Hours	42, 43	1	1
Visual Analogue Scales	VAS	See table 5	3, 23, 74-97	23	3

[1]Observational studies, [2]Intervention studies

Table 3: Scales currently used to measure fatigue in RA: Little or limited evidence of validity

Feasible	Validity							Reliability			
	Face	Content			Criterion	Construct					
Scale & subscale		Source	Inclusive	No Misleading items	'Gold' standard	Converge	Diverge	Internal consistency	Stability	Sensitivity	Feasibility
Created for single study											
Binary: Fatigue	3	1	3	3	0	0	0	NA	0	0	3
Binary: Tired	1	0	1	1	0	0	1	NA	0	0	3
Fatigue hours	0	0	0	0	0	0	0	NA	0	2	3
5 Items after Tack	1	1	2	2	0	1	0	0	0	2	2
Morning Fatigue	1	0	2	1	0	0	1	NA	0	0	3
Numerical Degree	3	0	3	3	0	1	0	NA	0	0	3
Numerical Impact	3	0	1	3	0	1	0	NA	0	0	3
Scales applied in RA											
CFS: Physical	3	2	3	2	1.5	0	1.5	0	0	0	3
Mental	3	2	3	2	1	0	1	0	0	0	3
CIS/SF	2	0	2	1	0	0	0	0	0	3	2
CFI Total	2	2	3	2	0	1.5	0	1.5	0	0	2
CFI Chronicity	3	2	3	3	0	0	0	1.5	0	0	0
CFI Interference	1	2	1	1	0	0	0	1.5	0	0	0
CIFI	0	0	0	0	0	2.5	0	0	0	0	0
FTC	0	0	0	0	0	0	0	0	0	0	2
MFI General	3	2	3	2	0	1.5	1.5	0	0	0	3
Physical	1	2	1	1	0	1.5	1.5	0	0	0	3
Activity	1	2	1	1	0	1.5	1.5	0	0	0	3
Motivation	1	2	1	1	0	1	1	0	0	0	3
Mental	2	2	2	2	0	1	1	0	0	0	3
NHP: Energy	2	2	1	2	0	2	1.5	1	1.5	0	2
Psychasthenia	2	1	1	1	0	0	1	0	0	0	3
RA-specific scales											
NIH ACTRE	1	0	1	3	0	1	0	0	0	1	1
Time to onset	1	1	1	1	0	0	0	NA	1	1	3

0 = Not reported/No evidence, 1 = Poor or limited evidence, 2 = Moderate evidence, 3 = Good evidence

Table 4: Scales currently used to measure fatigue in RA: Reasonable or good evidence of validity

Scale	Validity							Reliability		
	Face	Content			Criterion	Construct				
		Source	Inclusive	No misleading items	'Gold' standard	Converge	Diverge	Internal consistency	Stability	Sensitivity	Feasibility
MAF: Global Fatigue Index	3	2	2.5	3	2	2.5	2.5	3	1	3	3
FACIT-F	2	1	2.5	1	2	2	0	3	0	2.5	2
Ordinal scales: Best scores	3	0	3	3	0	1	2	NA	2	3	3
Stone: Not at all-extremely	3	0	3	3	0	1	2	NA	2	3	3
Katz: None-v severe *	2	0	3	3	0	1	0	NA	0	0	3
Pinals: None-severe	2	0	3	3	0	1	0	NA	0	0	3
POMS: Fatigue/inertia	2	2	2	2	1.5	0	1	3	0	1.5	1
SF36:											
Vitality (month)	2	1	2	2	2	2	2	3	0	3	3
Vitality (week)	2	1	2	2	0	2	0	2	0	1	3

*Scores dichotomized in paper

0 = Not reported/No evidence, 1 = Poor or limited evidence, 2 = Moderate evidence, 3 = Good evidence

Table 5: Visual analogue scales currently used to measure fatigue in RA patients

VAS	Start descriptor	End descriptor	Timescale	Length	Scoring	Comment in paper
Lwin,[23]Gilboe[74,75]	No fatigue	Fatigue as bad as it could be	Month	100 mm	0-100	
Currey[76]	Fatigue is no problem	Fatigue is a major problem	Week	10 cm	0-10	Unusual fatigue or tiredness
Tack[77]	No fatigue	Fatigue as bad as it could be	Week	100 mm	0-100	
Tack[77]	No distress	Distress as bad as it could be	Week	100 mm	0-100	Fatigue distress
Scharloo[78]	No tiredness	Very severe tiredness	Week	10 cm	0-100	
Mengshoel[79]	No fatigue	Total exhaustion	Week	100 mm	0-100	
Kvien[80]	Fatigue no problem	Fatigue major problem	Month	100 mm	0-100	
Crosby[81]	No fatigue	Extremely fatigued	Evening before	10 cm	0-100%	Vertical line
Wolfe[3,82,83]	Fatigue no problem	Fatigue major problem	Week	-	0-3	Fatigue or tiredness
Riemsma[84,85]	Not tired at all	Very tired	Week	-	0-100	
Barlow[86]	No fatigue	Fatigue as bad as it could be	-	10 cm	0-10	
Hagen[87]	Fatigue is no problem	Fatigue is a major problem	-	100 mm		
Borstlap[88]	-	-	-	-	0-10	Fatigue
Brekke[89,90]	-	-	-	100 mm	0-100	Fatigue
Gudbjornsson[91]	-	-	Time of investigation	-	10 grades	Degree of fatigue
Heiberg[92]	-	-	-	100 mm	0-100	Fatigue
Jensen[93]	-	-	-	100 mm	0-100	Fatigue
Uhlig[94]	-	-	-	100 mm	0-100	Fatigue
Chiang[95]	-	-	-	15 cm	0-150	Degree of fatigue
Gerber[96]	-	-	-	-	0-3	

- Data not supplied in paper

Table 6: Validation of Visual Analogue Scales for measuring fatigue in RA: Best scores for each version

	Validity							Reliability		
	Face	Content			Criterion	Construct				
		Source	Inclusive	No misleading items	'Gold' standard	Converge	Diverge	Stabilitye	Sensitivity	Feasibility
Best scores overall	**3**	**2**	**3**	**3**	**2**	**2.5**	**2.5**	**0**	**1.5**	**3**
Barlow[86]	3	0	2	3	0	0	1	0	0	1
Borstlap[88]	0	0	0	0	0	2	1	0	0	0
Brekke[89,90]	0	0	0	0	1	0	0	0	1.5	0
Chiang[95]	0	0	0	0	0	0	0	0	0	0
Crosby[81]	2	0	1	3	0	2	1	0	0	3
Currey[76]	2	0	3	1.5	0	0	1.5	0	0	3
Gerber[96]	0	0	0	0	0	0	0	0	0	0
Lwin,[23] Gilboe[74,75]	3	0	2	3	1	0	1	0	0	3
Gudbjornsson[91]	2	0	2	3	0	0	2	0	0	2
Hagen[87]	2.5	0	3	2.5	1	1	0	0	0	1
Jensen[93]	0	0	0	0	0	0	1	0	0	0
Heidberg[92]	0	0	0	0	1	0	0	0	0	0
Kvien[80]	2.5	0	3	2.5	2	2	2.5	0	0	3
Mengshoel[79]	3	0	3	3	0	2	1	0	0	3
Riemsma[84,85]	3	0	3	3	0	2.5	0	0	0	3
Scharloo[78]	2.5	0	2	2	0	0	0	0	0	3
Tack[77]	3	2	3	3	2	2	0	0	0	3
Tack[77]	3	2	3	3	1	1.5	0	0	0	3
Uhlig[94]	1	0	0	0	0	0	0	0	0	1
Wolfe[3,82,83]	2	0	3	1.5	1	2	1.5	0	0	2.5

0 = Not reported/No evidence, 1 = Poor or limited evidence, 2 = Moderate evidence, 3 = Good evidence