

Measures of Fatigue

Bristol Rheumatoid Arthritis Fatigue Multi-Dimensional Questionnaire (BRAf MDQ), Bristol Rheumatoid Arthritis Fatigue Numerical Rating Scales (BRAf NRS) for Severity, Effect, and Coping, Chalder Fatigue Questionnaire (CFQ), Checklist Individual Strength (CIS20R and CIS8R), Fatigue Severity Scale (FSS), Functional Assessment Chronic Illness Therapy (Fatigue) (FACIT-F), Multi-Dimensional Assessment of Fatigue (MAF), Multi-Dimensional Fatigue Inventory (MFI), Pediatric Quality Of Life (PedsQL) Multi-Dimensional Fatigue Scale, Profile of Fatigue (ProF), Short Form 36 Vitality Subscale (SF-36 VT), and Visual Analog Scales (VAS)

SARAH HEWLETT, EMMA DURES, AND CELIA ALMEIDA

INTRODUCTION

Fatigue is common to all the rheumatic conditions, in varying degrees, and is a frequent, often severe problem that has major consequences on patients' lives (1–4). In response to these concerns, a body of research subsequently led to international consensus that fatigue must be evaluated in all clinical trials of rheumatoid arthritis and potentially all fibromyalgia syndrome trials (5,6). The 12 fatigue patient-reported outcome measures (PROMs) reviewed in this section have been selected because they are currently or have recently been used in rheumatology populations. Fatigue PROMs in rheumatology were identified from previous reviews (7), then Medline, Cumulative Index to Nursing and Allied Health Literature, and PsycINFO searched for each PROM name plus each major rheumatologic condition. Not all articles could be evaluated and reported in this overview; therefore, those that evidenced strengths and weaknesses were included where possible. However, a full systematic review with meta-analysis would be welcome, as a limitation of this overview is that some articles contributing useful data may have been omitted. The fatigue PROMs are reviewed in

alphabetical order. Three additional scales with fatigue components are reviewed elsewhere in this edition: the Bath Ankylosing Spondylitis Disease Activity Index in the Measures of Ankylosing Spondylitis article, the Fibromyalgia Impact Questionnaire in the Measures of Fibromyalgia article, and the Nottingham Health Profile in the Adult Measures of General Health and Health-Related Quality of Life article.

When selecting a fatigue PROM, researchers and clinicians should consider whether their needs are best served by a single-item PROM as a screening tool, by multi-item PROMs that explore broader fatigue issues to create a global score, or by multidimensional PROMs that produce subscale scores for a range of different facets or domains of fatigue (e.g., cognitive and physical fatigue). Multidimensional PROMs with subscales may be useful for informing or evaluating interventions or exploring fatigue causality. Some fatigue PROMs relate to severity only, while others include items of both severity and consequence or impact.

Fatigue PROMs should differentiate between rheumatology populations and healthy controls. Many studies have shown that association between fatigue PROMs and inflammatory markers is not strong, and that fatigue is likely to have multicausal pathways of clinical variables (e.g., pain, disability) and psychosocial variables (e.g., mood, beliefs) combined in varying amounts (1,8–10). Fatigue PROMs should therefore show moderate correlation ($r = 0.3–0.49$) or large correlation ($r > 0.5$) with these variables (11). Very strong associations (e.g., >0.75) might be expected when examining criterion validity with other fatigue scales. Fatigue in rheumatologic conditions can be constant and persistent, but can also appear without warning as an overwhelming event (2–4). Reliability of fatigue PROMs can therefore be problematic to evaluate due to the fluctuating and unpredictable nature of fatigue itself. Some

Dr. Hewlett was in receipt of an unrestricted educational grant from GlaxoSmithKline, which partly-funded a PhD student to develop a rheumatoid arthritis fatigue questionnaire.

Sarah Hewlett, FRCN, PhD, MA, RN, Emma Dures, PhD, MSc, BSc, C Psychol, Celia Almeida, MSc, BSc, BA: University of the West of England, Bristol, UK.

Address correspondence to Sarah Hewlett, FRCN, PhD, MA, RN, Arthritis Research UK Professor of Rheumatology Nursing, University of the West of England, Academic Rheumatology, Bristol Royal Infirmary, Bristol BS2 8HW UK. E-mail: Sarah.Hewlett@uwe.ac.uk.

Submitted for publication April 6, 2011; accepted in revised form July 20, 2011.

fatigue PROMs have therefore been tested for stability over several weeks, and some over a matter of hours, both attempting to capture patients during a stable episode. Test–retest correlations of ≥ 0.7 are considered acceptable (12). Evaluation data are presented for rheumatology populations, but where these could not be found, data are presented from the original condition in which the PROM was developed.

BRISTOL RHEUMATOID ARTHRITIS (RA) FATIGUE MULTI-DIMENSIONAL QUESTIONNAIRE (BRAFF MDQ)

Description

Purpose. The BRAFF MDQ was developed to assess the overall experience and impact of RA fatigue, and its different dimensions. It was published in 2010 (13,14).

Content. The BRAFF MDQ covers domains of physical fatigue (e.g., average fatigue level over last 7 days), living with fatigue (e.g., has fatigue made it difficult to bathe or shower?), cognitive fatigue (e.g., has fatigue made it difficult to concentrate?), and emotional fatigue (e.g., has being fatigued upset you?).

Number of items. 20 items, providing a total fatigue score, including 4 subscale scores for physical fatigue (4 items), living with fatigue (7 items), cognitive fatigue (5 items), and emotional fatigue (4 items).

Response options. Four options from “Not at all,” “A little,” “Quite a bit,” to “Very much,” except for the first 3 items, which are numerical or categorical as appropriate (e.g., how many days did you experience fatigue in the past 7 days? 0–7).

Recall period for items. The past 7 days.

Endorsements. None found for rheumatologic conditions.

Examples of use. As this is a recently developed patient-reported outcome measure (PROM), there are no additional published studies yet. The BRAFF MDQ is currently being used in up to 11 clinical or research studies internationally (information available from the developers).

Practical Application

How to obtain. Available from the developers by e-mail (Sarah.Hewlett@uwe.ac.uk), the web site (available at URL: <http://hls.uwe.ac.uk/research/Default.aspx?pageid=312>), or by postal mail (Sarah Hewlett, Academic Rheumatology, Bristol Royal Infirmary, Bristol BS2 8HW, UK). The BRAFF MDQ is free to use.

Method of administration. Patient self-report, pen and paper.

Scoring. Items scored 0–3, except for items 1 (scored 0–10), 2 (scored 0–7), and 3 (scored 0–2). A total fatigue score is obtained by summing the 20 item scores. Subscale items are summed to produce scores for physical fatigue, living with fatigue, cognitive fatigue, and emotional fatigue. Instructions for missing data are that only 3 questions may be omitted in total, questions 1 and 2 must be completed, and only 1 question may be omitted from each

subscale (replaced with patient’s average score for that subscale). Scoring instructions and template can be downloaded from developers’ web site.

Score interpretation. Higher scores reflect greater severity. Total fatigue score is 0–70; subscale scores are physical fatigue 0–22, living with fatigue 0–21, cognitive fatigue 0–15, and emotional fatigue 0–12. In terms of normative data, in the developmental study, 229 people with RA recruited with fatigue visual analog scale (VAS) ≥ 5 out of 10 had a mean \pm SD total fatigue score of 38.4 ± 13.7 of 70, physical fatigue 16.7 ± 4 of 22, living with fatigue 9.6 ± 5.5 of 21, cognitive fatigue 6.1 ± 3.7 of 15, and emotional fatigue 5.8 ± 3.4 of 12 (14). No data for healthy controls could be found.

Respondent burden. Time to complete not reported but probably 4–5 minutes. Items do not appear difficult and have undergone cognitive interviewing.

Administrative burden. Time to score not reported, probably 2–3 minutes using template.

Translations/adaptations. Translated using appropriate linguistic methodology of forward translation, independent back translation by several native speakers, consolidation, then independent back translation to consolidate the final version (information available from the developers). English (UK), French (Belgium), Dutch (Belgium), Spanish (US), German, English (US), Japanese, South Korean, and Taiwan (Chinese) versions can be downloaded from developers’ web site. Over 20 further translations are in progress.

Psychometric Information

Method of development. Items were generated from qualitative research with patients (2), in collaboration with a patient research partner, refined through focus groups, then 45 draft items tested for clarity by cognitive interviewing (13). The 20-item MDQ and its subscales evolved from iterative rounds of Cronbach’s alpha (internal consistency) and factor analysis, with the weakest item removed each time. The resulting 20-item, 4-factor structure was confirmed by a second set of factor analysis on 20 separate, random samples of 50% of the data (bootstrapping) and showed no overlapping items. Subscale labels were discussed and agreed with patient collaborators (14).

Acceptability. The BRAFF MDQ appears easily readable; items have undergone cognitive interviewing, with fatigue specifically mentioned in every item. Levels of missing data are not reported. Floor effects (patients unable to report fatigue deterioration) and ceiling effects (patients unable to report improvement) are unlikely to be significant: in 229 patients with RA, <1% scored the maximum possible score for total fatigue, 2.7% for cognitive and for living with fatigue, 4.5% for physical fatigue, and 7% for emotional fatigue; no patients scored the minimum possible fatigue score for total and physical fatigue, <1% for living with fatigue, and 5% and 6% for cognitive and emotional fatigue (in patients recruited with a fatigue VAS >5 out of 10) (15).

Reliability. *Internal consistency.* Cronbach’s alpha for total fatigue was 0.93, physical fatigue 0.71, living with fatigue 0.91, cognitive fatigue 0.92, and emotional fatigue

0.89 (229 patients with RA) (14). Correlations between total fatigue and the 4 subscales range from $r = 0.75$ – 0.88 (14).

Test–retest. As fatigue onset is unpredictable and sudden (2), test–retest was conducted 1–2 hours apart ($n = 50$ patients with RA before and after clinic visits); total fatigue correlated $r = 0.95$, physical fatigue $r = 0.94$, living with fatigue $r = 0.89$, cognitive fatigue $r = 0.89$, and emotional fatigue $r = 0.92$ (16).

Validity. Content validity. Items and their wording cover a range of fatigue severity and impact and were derived from patient interviews, then refined with focus groups (13).

Construct validity. In 229 patients with RA, total fatigue correlated positively with depression, anxiety, disability, and helplessness (0.50–0.63); subscale physical fatigue (severity) correlated moderately with disability, depression, and helplessness (0.37–0.45), and weakly with anxiety (0.26); living with fatigue correlated positively with depression, anxiety, disability, and helplessness (0.45–0.61); cognitive fatigue correlated moderately with depression, anxiety, and helplessness (0.33–0.49), and weakly with disability (0.21); emotional fatigue correlated positively with depression and anxiety (0.54 and 0.57, respectively) and moderately with helplessness and disability (0.45 and 0.35, respectively); neither total fatigue nor the subscales are strongly associated with pain (0.14–0.38) (14).

Criterion validity. In 229 patients with RA, total fatigue correlated very strongly with the Multi-Dimensional Assessment of Fatigue (RA specific) at 0.82, and the Functional Assessment of Chronic Illness Therapy fatigue subscale at -0.81 , and positively with the Short Form 36 vitality subscale (SF-36 VT) at -0.64 (14). A range of moderate to strong correlations were also seen between these measures and the BRAF subscales: physical fatigue (severity) -0.68 to 0.83 , living with fatigue -0.54 to -0.74 , emotional fatigue -0.50 to 0.66 , and cognitive fatigue -0.40 to 0.55 (14). Lower levels of association seen in cognitive fatigue reflect the lack of cognitive fatigue items in other fatigue measures. For total fatigue and all subscales, correlation with SF-36 VT is weaker than with other fatigue scales (see section on SF-36 VT).

Ability to detect change. In patients with RA in flare receiving an intramuscular injection of glucocorticoids ($n = 42$), effect sizes of 0.33 – 0.56 for the total BRAF MDQ and subscales were seen at 2 weeks (all $P < 0.04$) (17).

Critical Appraisal of Overall Value to the Rheumatology Community

Strengths. The BRAF MDQ is RA specific, and was developed in collaboration with patients, with cognitive interviewing of draft items, and includes the word “fatigue” in every item. Factor analysis shows novel subscales of emotional, cognitive, and living with fatigue, which may help elucidate different causal or perpetuating mechanisms, or highlight individual patient dimensions that require targeted interventions. Internal consistency, test–retest reliability, and construct validity are good, and the BRAF MDQ shows criterion validity with other fatigue scales.

Caveats and cautions. Sensitivity data are still under peer review, and the full article on reliability and sensitivity is awaited. As a recent PROM, it has not yet been widely used; therefore, all evidence is from the developers’ article only.

Clinical usability. The available data suggest the BRAF MDQ may be a useful tool in identifying different types of RA fatigue, which might inform individualized self-management interventions. There is no significant administrative or respondent burden.

Research usability. The available data suggest the BRAF MDQ may be a useful research tool in identifying the overall fatigue experience and different types of RA fatigue, and potentially how these might have different causal factors or treatment responses.

BRISTOL RHEUMATOID ARTHRITIS (RA) FATIGUE NUMERICAL RATING SCALES (BRAFF NRS) FOR SEVERITY, EFFECT, AND COPING

Description

Purpose. Lack of standardized NRS and visual analog scales (VAS) for fatigue limits the interpretation of data and researchers often create individual items for individual studies (7); therefore, the aim of the BRAFF NRS was to develop standardized NRS for measuring a range of RA fatigue domains: severity, effect on life, and coping ability. The BRAFF NRS were published in 2010 (13,14).

Content. 3 single-item NRS on fatigue severity (average level of fatigue), effect (effect fatigue has had on your life), and coping (how well you have coped with fatigue).

Number of items. 3, 1 for each concept.

Response options. Patients circle the NRS from 0–10. Anchors are: for severity, “No fatigue” to “Totally exhausted”; for effect, “No effect” to “A great deal of effect”; and for coping, “Not at all well” to “Very well.” Initial test–retest data suggested the lack of specificity in the anchors of the coping NRS caused confusion; therefore, anchors were rephrased as “Not coped at all” to “Coped very well” and are being retested (16).

Recall period for items. During the past 7 days.

Endorsements. None found for rheumatologic conditions.

Examples of use. As a recently developed patient-reported outcome measure (PROM), there are no additional published studies yet. The BRAFF NRS are currently being used in up to 11 clinical or research studies internationally (information from the developers).

Practical Application

How to obtain. Available from the developers by e-mail (Sarah.Hewlett@uwe.ac.uk), the web site (available at URL: <http://hls.uwe.ac.uk/research/Default.aspx?pageid=312>), or by postal mail (Sarah Hewlett, Academic Rheumatology, Bristol Royal Infirmary, Bristol BS2 8HW, UK). The BRAFF NRS is free to use.

Method of administration. Patient self-report, pen and paper.

Scoring. Each NRS is scored 0–10.

Score interpretation. Scores range from 0–10 with higher scores reflecting greater problems for severity and effect NRS, but lower scores reflecting greater problems for coping NRS. In terms of normative data, 229 people with RA recruited with a screening fatigue VAS ≥ 5 out of 10 had a mean \pm SD BRAF severity NRS of 6.8 ± 1.8 , effect NRS of 6.5 ± 2.2 , and coping NRS of 5.7 ± 2.3 (14). No data for healthy controls could be found.

Respondent burden. Time to complete not reported but probably 1 minute for the trio. Items do not appear difficult and have undergone cognitive interviewing.

Administrative burden. Time to score not reported but probably 1 minute.

Translations/adaptations. Translated using appropriate linguistic methodology of forward translation, independent back translation by several native speakers, consolidation, then independent back translation to consolidate the final version (information available from the developers). English (UK), French (Belgium), Dutch (Belgium), Spanish (US), German, English (US), Japanese, South Korean, and Taiwan (Chinese) versions can be downloaded from developers' web site. Over 20 further translations are in progress.

Psychometric Information

Method of development. The topics and wording were generated from qualitative research with patients (2), refined by patient research partner, focus groups, and cognitive interviewing (13).

Acceptability. The BRAF NRS appear easily readable, having been developed with patients and undergone cognitive interviewing. Floor effects and ceiling effects are unlikely to be significant: in 229 patients with RA, 4% scored the worst possible score for severity and effect, and 3% for coping; no patients scored the minimum possible score for severity, 0.4% for effect, and 2.6% for coping (patients were recruited with a fatigue VAS > 5 out of 10) (15).

Reliability. *Test–retest.* As fatigue onset is unpredictable and sudden (2), test–retest was conducted 1–2 hours apart ($n = 50$ patients with RA before and after clinic attendance); severity NRS correlated at $r = 0.92$, effect $r = 0.85$, and coping $r = 0.62$ (16). Coping NRS anchors were subsequently reworded to enhance clarity and are currently being retested (16).

Validity. *Content validity.* The single-item NRS cover aspects of fatigue generated from patient interviews, refined with focus groups (13); fatigue coping is not available as a separate domain in other PROMs (7).

Construct validity. In 229 patients with RA, severity NRS correlated moderately with helplessness, depression, disability, anxiety, and pain (0.31–0.45); effect NRS also correlated moderately with these (0.34–0.49); coping NRS correlated moderately with depression (–0.32 to –0.42), weakly with disability and anxiety (–0.21 to –0.29), and not with pain (–0.08); and there was no association be-

tween the NRS and raised plasma viscosity (–0.01 to –0.25) (14).

Criterion validity. In 229 patients with RA, the NRS correlated with fatigue measures Multi-Dimensional Assessment of Fatigue, Functional Assessment Chronic Illness Therapy (Fatigue), and Short Form 36 vitality subscale (SF-36 VT): severity 0.65–0.80, effect 0.65–0.75, and coping 0.37–0.38 (14). Lower levels of association seen in coping NRS reflect the lack of coping items in other fatigue measures. For all NRS, correlations with SF-36 VT are weaker than other fatigue scales (see later SF-36 VT section). Correlation between severity and effect ($r = 0.71$) is strong, while associations between perceived coping and both severity and effect are weak to moderate ($r = -0.235$ and -0.352 , respectively), suggesting coping is a different concept (14).

Ability to detect change. In patients with RA in flare receiving an intramuscular injection of glucocorticoids ($n = 42$), effect sizes of 0.47 and 0.46 for the BRAF severity and effect short scales were seen, but no significant change in BRAF coping (17).

Critical Appraisal of Overall Value to the Rheumatology Community

Strengths. The BRAF NRS are RA specific, and were developed in collaboration with patients, including cognitive interviewing of items. They differentiate fatigue severity from effect and perceived coping ability. They show good construct and criterion validity, and severity and effect show good test–retest reliability. Identically phrased VAS were tested alongside the NRS (13–16). However, the developers recommend the NRS versions as they can be telephone administered, may be conceptually easier to understand and therefore more accurate (13–16,18), and the BRAF NRS showed stronger construct and criterion validity than VAS versions (14–16).

Caveats and cautions. The full article on reliability and sensitivity is awaited. Reverse scoring of the coping NRS may mean that interpretation of coping scores is not immediately obvious, and it has weaker reliability. As a recent PROM, the NRS have not yet been widely used; therefore, all evidence is from developers' articles only.

Clinical usability. The available data suggest the BRAF NRS may be a useful, quick tool to identify 3 different concepts of RA fatigue, which might inform individualized self-management interventions, with no significant administrative or respondent burden.

Research usability. The available data suggest the BRAF MDQ may be a useful research tool to screen for entry criteria, and to identify different facets of fatigue that might be changed differentially by interventions (e.g., fatigue not reduced but perceived coping and impact improved).

CHALDER FATIGUE QUESTIONNAIRE (CFQ)

Description

Purpose. Sometimes referred to as the Chalder Fatigue Scale, or simply the Fatigue Questionnaire or Scale, the

CFQ was developed to assess disabling fatigue severity in hospital and community populations and was originally published in 1993 with further psychometric evaluation in 2010 (19,20).

Content. Covers physical fatigue (e.g., lack energy, feel weak, less muscle strength, need to rest), and mental fatigue (e.g., concentration, memory).

Number of items. 11 items to produce a global score and 2 domains of physical and mental fatigue.

Response options. 4 options, slightly reworded in the latest evaluation: “Less than usual,” “No more than usual,” “More than usual,” and “Much more than usual” (20).

Recall period for items. In the last month.

Endorsements. None found for rheumatologic conditions.

Examples of use. The CFQ has been used in systemic lupus erythematosus (SLE), primary Sjögren’s syndrome (PSS), rheumatoid arthritis (RA), psoriatic arthritis (PsA), fibromyalgia syndrome (FMS), and upper-extremity or carpal tunnel disorder (21–28), as well as chronic fatigue syndrome (CFS) and chronic widespread pain.

Practical Application

How to obtain. From the developer by e-mail: trudie.chalder@kcl.ac.uk. The CFQ is free to use.

Method of administration. Patient self-report, pen and paper.

Scoring. Items can be scored in 2 ways. The first is on a scale of 0–3, giving a global score range of 0–33, a physical fatigue domain range of 0–21 (items 1–7), and a mental fatigue domain range of 0–12 (items 8–11). Second, the CFQ can be scored in a binary fashion (0,0,1,1), then summed to produce a global score of 0–11. No information could be found on handling missing items.

Score interpretation. Higher scores reflect greater fatigue. For Likert scoring, a score of 29 of 33 discriminates clinically relevant fatigue from nonclinically relevant fatigue (20), and for binary scoring, a global score of ≥ 4 of 11 designates a “case” of fatigue (19). In terms of normative data, mean \pm SD Likert scores in a community population ($n = 1,615$) were 14.2 ± 4.6 of 33 versus 24.4 ± 5.8 in patients with CFS ($n = 361$; $P < 0.0001$) (20). In FMS, mean \pm SD score was 18.2 ± 6.1 ($n = 30$), and in PsA 6.8 ± 2.8 ($n = 9$) (21). Mean \pm SD binary scores in a community population ($n = 1,615$) were 3.27 ± 3.21 of 11 versus 9.14 ± 2.73 in 274 patients with CFS (20). Many studies use the draft 14-item CFQ (score range 0–42), giving a mean global fatigue score in 120 patients with SLE of 22 (interquartile range [IQR] 16–28) (22). Using the draft 14-item CFQ with a 5th response option added (score range 0–56), median global fatigue in 51 patients with PSS was 37 (IQR 32–42) versus 28 (IQR 28–32) in 51 controls ($P = 0.000$) (23).

Respondent burden. Time to complete not reported but probably 2–3 minutes. Items appear easy to interpret.

Administrative burden. Time to score not reported, probably 2–3 minutes.

Translations/adaptations. The CFQ comprises 11 items scored 0–33 (19) and underwent minor wording change in 2010 (20). However, while the 8 rheumatology studies

identified here quote the original validation article (19), only 2 use this version (21,28). Three articles use the 14-item draft CFQ (22,24,25) giving scores of 0–42; it is not clear from 2 articles which version has been used (26,27), and a Swedish version combined the 14-item draft CFQ with an additional 5th response option (“Much better than usual”), giving a global score of 0–56 (23).

Psychometric Information

Method of development. Fourteen draft items were generated by professionals to represent physical and mental fatigue, and evaluated in new registrants at a general practice (GP; family doctor; $n = 274$, ages 18–45 years) (19). Factor analysis identified 3 items for removal (19). The resultant 11-item scale includes 2 clear domains on factor analysis (physical fatigue, mental fatigue) with slight overlap between factors for 1 item (concentration) (19,20).

Acceptability. Items appear easy to read. No data on missing item rate or floor/ceiling effects in rheumatology could be found.

Reliability. *Internal consistency.* Cronbach’s alpha was calculated in 274 GP patients for all 14 draft items and by taking out different items one at a time (0.88–0.90) and for the 2 domain scores (physical 0.84, mental 0.82) (19). For the final 11-item version, Cronbach’s alpha was 0.89 in GP patients ($n = 274$), 0.92 in patients with CFS ($n = 361$), and 0.88 in a survey of GP attenders ($n = 1,615$) (19,20). No internal consistency data for rheumatology could be found.

Test-retest. No test-retest data could be found.

Validity. *Content validity.* Items were generated by experts and the final 11-item CFQ covers a range of physical and mental fatigue issues and produces a domain score for each (19).

Construct validity. A CFQ score of 29 of 33 discriminates patients with CFS from the general population in 96% of cases (20). In SLE, using the 14-item draft CFQ, mean fatigue was significantly different between patients (23.5, SEM 0.9; $n = 93$) and controls (15.0, SEM 0.6; $n = 41$) (24). However, no difference in total CFQ, physical or mental domains was found between controls and patients with SLE or PSS, and patients with RA only differed from controls in physical fatigue ($P < 0.05$) (28). In SLE, the draft 14-item CFQ scores were moderately associated with each of 4 disease activity measures ($r = 0.36$ – 0.4), and with aerobic capacity ($r = -0.33$) (22). In chronic upper-extremity pain ($n = 73$), CFQ was moderately associated with pain disability ($r = 0.44$) and pain intensity ($r = 0.32$) (27).

Criterion validity. Using a validated psychiatric fatigue interview schedule as a comparator, the cut off for a “case” of fatigue was identified as ≥ 4 of 11 on the 14-item draft CFQ with 75.5% sensitivity and 74.5% specificity (100 consecutive GP attenders) (19). In SLE ($n = 120$), the 14-item draft CFQ was strongly associated with the Fatigue Severity Scale and a fatigue visual analog scale (VAS; both $r = 0.6$) (22). In a group of patients with PSS, RA, and SLE, CFQ total and physical fatigue scores correlated moderately with a fatigue VAS ($r = 0.42$ and $r = 0.46$, respectively), but neither the total, physical, nor mental CFQ

scores correlated significantly with Short Form vitality subscale (which also did not correlate with the fatigue VAS) (28).

Ability to detect change. In 93 patients with SLE randomized to exercise, relaxation, or control, CFQ improved significantly at exit (22 to 15 versus 24 to 21), and was significantly different between the 8 patients who continued to exercise at 3 months and the 25 who had stopped (11, SEM 5–17 versus 17, SEM 12–26) (26).

Critical Appraisal of Overall Value to the Rheumatology Community

Strengths. The CFQ is a fatigue severity scale rather than a measure of impact or consequence, and has physical and mental domains. CFQ has good internal consistency in CFS populations, and good sensitivity to change in rheumatology.

Caveats and cautions. Users need to obtain the correct version (20) from the developer. Many researchers continue to use the draft 14-item version, which makes interpretation across studies difficult. The response options comprise 1 positive, 1 neutral, and 2 negative responses, which might bias Likert scoring (0–3), although using binary scoring (0/1) to define “cases” resolves this issue. There are few rheumatology data on the 2 domains, nor on internal consistency or test–retest. In one study, CFQ did not differentiate between people with rheumatologic conditions and controls. Construct and criterion validity were only moderate in rheumatology.

Clinical usability. A short scale, potentially useful in clinical situations to measure fatigue severity. No significant administrative or respondent burden.

Research usability. Potentially a short, useful patient-reported outcome measure to measure fatigue severity. No significant administrative or respondent burden.

CHECKLIST INDIVIDUAL STRENGTH (CIS20R AND CIS8R)

Description

Purpose. The CIS was developed to measure several aspects of fatigue in chronic fatigue syndrome (CFS) in 1994 (29).

Content. The CIS covers domains of the subjective fatigue experience (e.g., Physically I feel exhausted), concentration (e.g., Thinking requires effort), motivation (e.g., I don't feel like doing anything), and physical activity levels (e.g., I think I do very little in a day).

Number of items. 20 items providing a total CIS20R score, including 4 subscale scores for subjective fatigue experience (8 items), concentration (5 items), motivation (4 items), and physical activity levels (3 items). Although the entire CIS20R assesses fatigue, the 8-item subjective fatigue subscale is commonly the only subscale reported and is often referred to as CIS8R, CIS-Fatigue, or Fatigue Severity.

Response options. 7 boxes ranging from “Yes that is true,” to “No that is not true.”

Recall period for items. The past 2 weeks.

Endorsements. None found for rheumatologic conditions.

Examples of use. Mainly used in CFS, multiple sclerosis, neurologic disorders, and healthy working adults, but has been used in rheumatoid arthritis (RA) and fibromyalgia syndrome (FMS) (30–39).

Practical Application

How to obtain. From the developer by e-mail: j.vercoulén@mps.umcn.nl. The CIS is free to use.

Method of administration. Patient-self-report, pen and paper.

Scoring. Items scored 1–7, with 11 positively phrased items reverse scored. A total CIS20R score is obtained by summing the 20-item scores. Subscale items are summed to produce scores for subjective fatigue (CIS8R), concentration, motivation, and physical activity. No information is given on handling missing items.

Score interpretation. Higher scores reflect greater severity. Overall CIS20R score is 20–140; subscale scores are subjective fatigue (CIS8R) 8–56, concentration 5–35, motivation 4–28, and physical activity 3–21. In terms of normative data, in healthy controls ($n = 60$), mean \pm SD subjective fatigue was 2.4 ± 1.4 , concentration 2.2 ± 1.2 , motivation 2.0 ± 1.0 , and physical activity 2.0 ± 1.3 (29). Cut offs on the subjective fatigue CIS8R scale for patients with RA are based on the mean score for healthy adults plus 1 or 2 SDs, i.e., 27–35 for heightened fatigue, and ≥ 35 for severe fatigue (31), with ≥ 35 reported as similar to fatigue levels in CFS (38). In patients with RA ($n = 228$), mean \pm SD subjective fatigue CIS8R was 31.5 ± 12.8 , with 20% reporting heightened fatigue and 42% reporting severe fatigue (31).

Respondent burden. Time to complete not reported but probably 4–5 minutes.

Administration burden. Time to score not reported, probably 4–5 minutes to reverse score some items, then identify and sum subscale items.

Translations/adaptations. The CIS originates from The Netherlands. Dutch, English, Swedish, and Korean versions are available from the developer.

Psychometric Information

The subjective fatigue subscale (CIS8R) is the most commonly and often the only reported data in studies using the CIS.

Method of development. No information could be found on how items were generated; 20 of the original 24 draft items were retained as they performed best in factor analysis (29). Subscales were generated through principal components analysis and Cronbach's alpha (internal consistency) (29). Evaluation in RA is reported in an abstract (37).

Acceptability. In a rheumatology population, 3 items might be interpreted in relation to RA or disability (“I feel fit,” “Physically I feel I am in bad form,” “Physically I feel I am in an excellent condition”) and thus may not be sensitive to RA fatigue. Levels of missing data and floor/ceiling effects are not reported.

Reliability. *Internal consistency.* In CFS, total CIS20R score Cronbach's alpha was 0.90 and Gutman split-half reliability coefficient 0.92; Cronbach's alpha for subscales ranged from 0.83–0.92 (29). In patients with RA (n = 227), Cronbach's alpha for subjective fatigue CIS8R was 0.92 (37), and 0.89 in patients with FMS (n = 78) (36). In patients with RA, factor analysis is reported as confirming the 4 subscales (no data are provided) (n = 227) (37).

Test-retest. In 227 patients with RA, intraclass correlation coefficient for subjective fatigue CIS8R over 1 month was 0.81 (37).

Validity. *Content validity.* No information is provided on how items were generated (29) but the CIS20R covers a range of fatigue issues likely to be common in rheumatology populations (2–4).

Construct validity. In patients with RA (n = 228), subjective fatigue CIS8R correlated strongly with pain (0.55), moderately with disability, sleep disturbance, helplessness, anxiety, and depression (0.32–0.40), weakly with rheumatoid factor, Disease Activity Score in 28 joints, and tender or swollen joints (0.18–0.3), and not with disease duration or inflammatory indices (31). The total CIS20R score discriminates between healthy workers and workers with health reasons for being fatigued (39).

Criterion validity. In patients with RA (n = 227), subjective fatigue CIS8R correlated very strongly with Short Form 36 vitality subscale and with a fatigue numerical rating scale (both 0.81) (37). In patients with FMS (n = 224), subjective fatigue CIS8R correlated with a fatigue visual analog scale (VAS) at 0.61 (35).

Ability to detect change. In patients with FMS (n = 78) receiving cognitive-behavioral therapy (CBT), subjective fatigue CIS8R improved by a mean \pm SD -10.6 ± 10.7 (36). In patients with RA started on anti-tumor necrosis factor therapy (n = 126), total CIS20R score improved from a mean 85 (65–97) to 69 (48–90) over 6 months (30), while in a subset of 59 working-age patients, CIS20R score improvement was 11.8% at 6 months (33). In early, distressed patients with RA (n = 30), CBT gave an effect size of 0.55 posttreatment for subjective fatigue CIS8R (0.48 at 6 months) (34). No minimum clinically important difference is reported, but in FMS (n = 78), change in subjective fatigue CIS8R correlated with a transition question on perceived change (0.53), and with a VAS for usefulness of and satisfaction with the level of change (0.42 and 0.33, respectively) (36).

Critical Appraisal of Overall Value to the Rheumatology Community

Strengths. The CIS was developed in CFS but has been evaluated in many long-term conditions, suggesting it is a useful generic scale. The CIS has good internal consistency and reliability, construct and criterion validity, and sensitivity to change. Subscales differentiate between cognitive and physical fatigue.

Caveats and cautions. The full article evaluating use of CIS in RA is awaited. Most evidence is reported only for the subjective fatigue subscale. Three items may be confounded by disability or disease activity in rheumatology populations.

Clinical usability. The available data suggest the CIS may be a useful tool in identifying cognitive and physical fatigue, which might inform individualized self-management interventions. No significant respondent or administrative burden.

Research usability. The available data suggest the CIS may be a useful research tool in identifying both the overall fatigue experience and different types of fatigue, within the caveats above.

FATIGUE SEVERITY SCALE (FSS)

Description

Purpose. The FSS was developed to assess disabling fatigue in multiple sclerosis (MS) and systemic lupus erythematosus (SLE), and was published in 1989 (40).

Content. The FSS covers physical, social, or cognitive effects of fatigue (e.g., function, work, motivation).

Number of items. 9 items to produce a global score.

Response options. 7 options from “Strongly disagree” to “Strongly agree” (1–7).

Recall period for items. The past week.

Endorsements. After systematic review of 15 fatigue scales used in SLE, the Ad Hoc Committee recommended the FSS for use in SLE (41).

Examples of use. Has been used extensively in SLE studies, and also in rheumatoid arthritis (RA), osteoarthritis (OA), and ankylosing spondylitis (AS), with a modified version in psoriatic arthritis (PsA) (41–54), as well as many long-term conditions (e.g., MS, cancer, neurologic disorders).

Practical Application

How to obtain. From the developer by e-mail: lkrupp@notes.cc.sunysb.edu. The FSS is free to use.

Method of administration. Patient self-report, pen and paper.

Scoring. Items are scored 1–7, summed, then averaged to produce a global score.

Score interpretation. Scores range from 1–7 with higher scores reflecting greater fatigue. In terms of normative data, mean \pm SD score in healthy adults (n = 20) was 2.3 ± 0.7 , compared to 4.7 ± 1.5 in patients with SLE, and 4.2 ± 1.2 in patients with RA (n = 29 and 122, respectively) (40,44). In patients with PsA (n = 135) using a modified FSS scaled from 0–10 (mFSS; see adaptations below), mean score was 5.7 (95% confidence interval [95% CI] 5.1–6.3) (45). In another PsA study (n = 75) using the mFSS, patients reporting fatigue in a clinical assessment had a mean \pm SD mFSS score of 6.9 ± 2.4 compared to 3.8 ± 2.8 in those reporting no fatigue (42). In OA (n = 137), mean \pm SD FSS was 3.63 ± 1.55 (54).

Respondent burden. Time to complete not reported but probably 2–3 minutes. Items appear easy to interpret.

Administrative burden. Time to score not reported, probably 2–3 minutes.

Translations/adaptations. Translated into multiple languages, including Spanish, French, Chinese, and Portuguese (41) with a Swedish translation describing

appropriate linguistic methodology, then evaluation of reliability, and construct and criterion validity in SLE (46). Adaptations include a multidimensional, 29-item Fatigue Assessment Instrument in German (55); a US adaptation for telephone administration in RA, which reduced the response options from 1–7 to 1–5 and states FSS has 10 rather than 9 items (47); and an mFSS used in PsA that increased the response options from 1–7 to 0–10 (“Not at all” to “Entirely”), although no rationale for this was presented (45).

Psychometric Information

Method of development. Factor analysis was performed on 28 draft items, and identified 9 items common to both SLE and MS ($n = 29$ and 25 , respectively); it is not stated where the 28 items originated from or whether patients were involved in their development (40).

Acceptability. Items appear easy to read, the Swedish version underwent cognitive debriefing (46) and fatigue is mentioned in every question. In one study, none of the 22 patients with SLE omitted any questions; the study reported no ceiling effects, but a possible floor effect for 1 item, where the median score was also the maximum possible score (46).

Reliability. *Internal consistency.* Cronbach's alpha was 0.89 – 0.94 in SLE ($n = 22$ – 29) (40,46). In the mFSS (0–10 response option), Cronbach's alpha was 0.95 in both PsA ($n = 91$) and SLE ($n = 113$) (45).

Test-retest. No significant difference was seen in FSS in stable patients with SLE over 1 week (46).

Validity. *Content validity.* FSS covers a range of fatigue issues. It is not stated how items were generated (40) but the FSS later underwent cognitive testing in Swedish patients with SLE (46).

Construct validity. FSS correctly discriminated 90% of 29 patients with SLE from healthy controls (40). A systematic review reports evaluation of construct validity of the FSS in a number of SLE studies, demonstrating a range of correlations with disease activity (0.16 – 0.53), depression (0.22 – 0.59), and pain (0.35 – 0.54) (41). In patients with SLE ($n = 22$), FSS correlated strongly with pain, general health, and physical and social roles (-0.59 to -0.60), and moderately with function, emotional role, and mental health (-0.41 to -0.44) (46); no association was found with inflammatory indices ($n = 57$ SLE) (48). In an RA working population ($n = 122$), FSS correlated with anxiety and depression (0.55 and 0.53 , respectively), and disability, pain, and stress (0.33 – 0.48) (43). In PsA ($n = 135$), the mFSS (10-point response) correlated with number of active joints (0.37), but not swollen or damaged joints (49).

Criterion validity. mFSS correlated strongly with a fatigue visual analog scale at 0.81 (SLE, $n = 29$) (40) and with Functional Assessment Chronic Illness Therapy (Fatigue) at -0.79 in patients with PsA ($n = 135$) (49). Correlation with Short Form 36 vitality subscale (SF-36 VT) was -0.56 to -0.63 in SLE, OA, and RA ($n = 32$, 137 , and 52 , respectively) (46,54).

Ability to detect change. In patients with AS randomized to etanercept or placebo ($n = 40$), FSS showed an effect size of 0.43 for treatment at 4 months (SF-36 VT

effect size 0.69); FSS was not responsive at 1 month, unlike SF-36 VT (effect size 0.15 versus 0.54) (50). In SLE ($n = 58$), effect sizes of 0.55 and 0.44 were shown from telephone interventions for fatigue (modified 10-item, 5-response option FSS) (47). Based on linear regression analysis on comparative fatigue ratings from patients after paired interviews, the effect size (mean change/SD at baseline) required for an average patient to move to a different fatigue category (i.e., much, somewhat or a little, less or more fatigued) is calculated as 0.74 in RA (52) and 0.41 (95% CI 0.2 – 0.57) in SLE, where the authors also present this as an FSS minimum clinically important difference score of 0.6 (95% CI 0.3 – 0.9) (53). Based on a systematic review of earlier SLE studies, a recommendation for important improvement in FSS for patients with SLE was 15% (41).

Critical Appraisal of Overall Value to the Rheumatology Community

Strengths. The FSS has good internal consistency, reliability, and construct and criterion validity, and is sensitive to change. It has been evaluated in several rheumatologic conditions, particularly SLE, where it is the recommended fatigue scale (41).

Caveats and cautions. Differences in sensitivity compared to SF-36 VT were shown in one study, but without a third fatigue comparator patient-reported outcome measure (PROM) for that study, it is not possible to conclude whether FSS or SF-36 VT is more accurate. Comparison between rheumatologic groups may be difficult if some groups use the differently scaled mFSS rather than the FSS.

Clinical usability. A short scale, potentially useful in clinical situations.

Research usability. Potentially a short, useful PROM for research, although researchers should be aware items suggest FSS may measure fatigue impact rather than severity.

FUNCTIONAL ASSESSMENT CHRONIC ILLNESS THERAPY (FATIGUE) (FACIT-F)

Description

Purpose. The FACIT-F was developed in 1997 to measure fatigue in oncology patients with anemia and is a stand-alone (or add-on) questionnaire in the Functional Assessment in Cancer Therapy measurement system (56). This has since been widened to include assessment of chronic illnesses (FACIT measurement system). The current version of FACIT-F is number 4.

Content. The FACIT-F covers physical fatigue (e.g., I feel tired), functional fatigue (e.g., trouble finishing things), emotional fatigue (e.g., frustration), and social consequences of fatigue (e.g., limits social activity).

Number of items. 13 to produce a global score.

Response options. 5 responses from “Not at all” to “Very much.”

Recall period for items. Past 7 days.

Endorsements. None found for rheumatologic conditions.

Examples of use. Has been evaluated in rheumatoid arthritis (RA) and psoriatic arthritis (PsA), and used in primary Sjögren's syndrome (PSS), osteoarthritis (OA), and systemic lupus erythematosus (SLE) (49,52,53,57–63), as well as many long-term conditions (e.g., multiple sclerosis, cancer, neurologic disorders).

Practical Application

How to obtain. From the FACIT web site after free registration at URL: <http://www.facit.org/>. English versions are free to use, a fee is payable for non-English versions used in commercial studies.

Method of administration. Patient self-report, interviewer or telephone administered.

Scoring. Items scored 0–4, with 2 positively phrased items reverse scored. Items are summed, multiplied by 13, then divided by the number of items actually answered, therefore allowing for missing items. However, more than 50% of items must be answered (i.e., at least 7 items). Scoring instructions can be downloaded from the developers' web site, including computerized versions.

Score interpretation. Scores range from 0–52, with higher scores reflecting less fatigue. In terms of normative data, mean \pm SD score for 1,010 healthy adults was 43.6 ± 9.4 (64); this compares to 29.17 ± 11.06 in patients with RA, 35.8 ± 12.4 in patients with PsA, 25.7 ± 12.0 in patients with SLE, and 30.1 in patients with PSS ($n = 631$, 135, 80, and 277, respectively) (49,53,57,63).

Respondent burden. 3–4 minutes. Items appear easy to interpret.

Administrative burden. Time to score not reported but probably 3–4 minutes.

Translations/adaptations. Available in over 50 languages.

Psychometric Information

Method of development. Items were generated in semi-structured interviews with 14 anemic oncology patients and 8 clinicians, followed by item reduction by 5 medical experts, then evaluation in 49 oncology patients (56). In patients with RA ($n = 271$), analysis using item response theory suggested the FACIT-F covers a wider range of fatigue (with the exception of those with a very low level) than either the Short Form 36 vitality subscale (SF-36 VT) or Multi-Dimensional Assessment of Fatigue (MAF) (57).

Acceptability. The items are brief and easy to understand. However, in arthritis populations, some items have the potential for misinterpretation. Two items could potentially be interpreted as relating to disability rather than fatigue as fatigue is not stipulated in the wording ("Ability" and "Needing help to do usual activities"), 1 item measures energy, which may be a positive health state that is not necessarily the opposite end of a fatigue continuum (i.e., people who are not feeling energized may not necessarily feel fatigued), and 1 item that is applicable to patients with cancer may hold less relevance for patients with RA ("Feeling too tired to eat" is not reported in qualitative RA fatigue studies) (2,3,65). Floor/ceiling effect data could not be found in rheumatology.

Reliability. *Internal consistency.* Cronbach's alpha was 0.86–0.87 at 3 time points in RA ($n = 631$) and 0.96 in PsA ($n = 135$) (49,57).

Test–retest. Intraclass correlation coefficient over 1 week was 0.95 in patients with PsA ($n = 73$) (49).

Validity. *Content validity.* Items were generated by patients with cancer (56) but cover a range of fatigue issues likely to be common to arthritis (2–4).

Construct validity. In PsA ($n = 135$), FACIT-F correlated with inflamed joint count ($r = -0.43$, 95% confidence interval [95% CI] -0.56 to -0.28) but not with damaged joint count ($r = 0.06$, 95% CI -0.23 to 0.11), age, or disease duration (49). In RA ($n = 505$), FACIT-F correlated with disability (Health Assessment Questionnaire) and inflammation (Disease Activity Score in 28 joints) at $r = -0.42$ to -0.44 (60). FACIT-F scores were not statistically significantly different between patients with OA ($n = 43$) and PSS ($n = 71$), but sleepiness was more strongly associated with FACIT-F in PSS than in patients with OA (0.53 versus 0.27) (58).

Criterion validity. In RA, FACIT-F correlated strongly with MAF at 0, 12, and 24 weeks of antirheumatic treatment (-0.84 to -0.88), and with SF-36 VT (0.73–0.84) ($n = 567$ –631) (57). In PsA ($n = 135$), correlation with modified Fatigue Severity Scale was -0.79 (95% CI -0.85 to -0.72) while those patients responding positively to an anchor question on overwhelming fatigue had lower FACIT-F scores (i.e., more severe fatigue) than those responding negatively (mean \pm SD 24.8 ± 13.9 versus 38.5 ± 10.4 ; $n = 135$) (49).

Ability to detect change. After 24 weeks of antirheumatic treatments in patients with RA ($n = 631$), FACIT-F showed a mean change of 2.1 in patients who did not achieve American College of Rheumatology 20% criteria for improvement in disease activity (ACR20; effect size 0.19), compared to 12.4 in those who achieved ACR70 (effect size 1.13) (57). Sensitivity has also been shown in other anti-tumor necrosis factor trials in RA (61,62), and in PsA where changes in FACIT-F were similar to changes in SF-36 VT ($n = 313$) (59). A minimum clinically important difference (MCID) of 3–4 points is generally used, which was calculated using 0.2 and 0.5 effect size cut offs for 5 groups ("Major worsening" to "Major improvement") in 631 patients with RA receiving antirheumatic treatments, then confirmed in a second study ($n = 271$) (57). On a normalized scale of 0–100 (rather than 0–52), others have proposed an MCID for RA of 15.9 points (52). In SLE ($n = 80$), based on linear regression analysis on comparative fatigue ratings from patients after paired interviews, the effect size required (mean change/SD at baseline) for an average patient to move to a different fatigue category (i.e., much, somewhat or a little, less or more fatigued) is calculated as 0.5 (95% CI 0.31–0.65), which the authors also present as a FACIT-F MCID score of -5.9 (95% CI -8.1 to -3.6) (53).

Critical Appraisal of Overall Value to the Rheumatology Community

Strengths. FACIT-F is used across many rheumatologic conditions, particularly in pharmacologic trials. It covers a

range of fatigue concepts in easy to understand language. FACIT-F has good internal consistency and reliability, construct and criterion validity, and sensitivity to change.

Caveats and cautions. FACIT-F might potentially be limited for use in rheumatology by the phrasing of 4 of the 13 items.

Clinical usability. Would be easy to use in clinical practice, giving a global fatigue score.

Research usability. Would be easy to use in research where a global fatigue score is required.

MULTI-DIMENSIONAL ASSESSMENT OF FATIGUE (MAF)

Description

Purpose. The MAF was developed in 1991 to measure multiple dimensions of fatigue in adults with rheumatoid arthritis (RA) (66). It was a revision of the Piper Fatigue Scale, which had been developed and tested with oncology patients (67).

Content. The MAF covers 4 dimensions of fatigue: severity, distress, interference in activities of daily living (doing chores, cooking, bathing, dressing, working, visiting, sexual activity, leisure, shopping, walking and exercising), and frequency and change during the previous week.

Number of items. 15 items provide a global score (Global Fatigue Index [GFI]). The 16th question ("To what degree has your fatigue changed during the past week?") does not contribute to the GFI.

Response options. The number of response options depends on the nature of each item. The original version used visual analog scales (VAS) for items 1 and 4–14, but based on feedback from respondents, these were changed to numerical rating scales (NRS) ranging from 1–10 in 1995 (68). Items 1 (degree) and 4–14 (interference) have anchors of "Not at all" to "A great deal," item 2 (severity) has anchors of "Mild" to "Severe," and item 3 (distress) has anchors "No distress" to "A great deal of distress." Items 4–14 (interference with activities) provide an opportunity for respondents to indicate if they do not carry out the activity because of reasons other than fatigue, and the item is then not completed. Items 15 and 16 have 4 ordinal response options scored 1–4, with item 15 (frequency) ranging from "Hardly any days" to "Every day," and item 16 (change) ranging from "Decreased" to "Increased."

Recall period for items. The past week.

Endorsements. None found for rheumatologic conditions.

Examples of use. Although developed for use in RA, the MAF has also been used in other rheumatologic conditions, including osteoarthritis (OA), ankylosing spondylitis (AS), systemic lupus erythematosus (SLE), and fibromyalgia syndrome (FMS) (52,53,55,57,68–79), as well as other long-term conditions such as human immunodeficiency virus, multiple sclerosis, and cancer.

Practical Application

How to obtain. The MAF is copyrighted by the developer, Basia Belza, and may be downloaded after free reg-

istration from the web site available at URL: www.son.washington.edu/research/maf/, or obtained by postal mail at the following address: Basia Belza, PhD, RN, Department of Biobehavioral Nursing and Health Systems, Box 357266, University of Washington, Seattle, WA 98195-7266. There is no charge for individual use of the MAF, although a nominal fee may be charged for commercial use.

Method of administration. Patient self-report, pen and paper.

Scoring. The MAF was developed to provide an aggregated score, the GFI. If the respondent indicates "No fatigue at all" for item 1, all remaining items should be scored as 0. Items 1–3 are summed, items 4–14 are averaged but should not be scored where the respondent indicates that they do not do an activity "For reasons other than fatigue," and item 15 is transformed into a 0–10 scale by multiplying the score by 2.5. The GFI is then calculated by adding these 3 components (sum of items 1–3, average of items 4–14, and transformed item 15). Item 16 (change) does not contribute to the GFI, and is scored 1–4. No information is given on handling missing data, but the developer has suggested that nonresponse to ≥ 3 of the 16 items would mean the GFI could not be calculated (Tack BB: unpublished observations).

Score interpretation. The GFI ranges from 1 (no fatigue) to 50 (severe fatigue). A higher score represents greater fatigue severity, distress, or interference with activities of daily living. Item 16 (change) is scored from 1 (fatigue decreased) to 4 (fatigue increased). In terms of normative data, in healthy controls ($n = 46$), mean \pm SD GFI was 17.0 ± 11.3 (68). In rheumatologic conditions, mean \pm SD GFI was 29.2 ± 9.9 in RA, 32 ± 20 in AS, 36.4 ± 8.1 in FMS, 31.1 ± 11.4 in SLE, and 27.7 ± 10.8 in OA ($n = 51-1,636$) (53,68,71,74,76).

Respondent burden. Time to complete is not reported but is likely to be ~5–8 minutes.

Administrative burden. Time to score is not reported but is likely to be ~4–5 minutes to transform scores, sum and average dimensions, and create the GFI.

Translations/adaptations. The MAF was originally developed in US English. The MAPI Research Institute has versions in Spanish, Dutch, French, Mandarin, Croatian, Danish, Finnish, Czech, German, Turkish, Swedish, Afrikaans, Russian, Portuguese, Polish, Italian, Hungarian, Hebrew, and Norwegian. Translation was undertaken using both forward and backward translations. Translations can be obtained through the MAPI web site at URL: <http://www.mapi-institute.com/home>.

Psychometric Information

Method of development. Items from a 41-item cancer fatigue scale (67) that were considered to describe activities often affected in RA were selected for the MAF (66). No patients were involved in selecting items. Following patient feedback from early studies, the VAS format was changed to NRS in 1995 (68).

Acceptability. Items appear easy to understand, but some may contain overlapping concepts; walking (item 13) and exercise (item 14) might both be considered leisure

activities (item 11). Response options for frequency over the past week (item 15) might be subject to different interpretations; for example, when wishing to report 2 days of fatigue, some patients might consider that “Occasionally” and others might consider it “Hardly any days.” At the start of items 4–14, respondents are clearly instructed to consider to what degree fatigue has interfered with activities, but fatigue is subsequently not mentioned in the question stems; thus, respondents might inadvertently consider interference due to disability rather than fatigue when scoring these 11 items. Missing data have been reported as a problem; based on the MAF not being able to be scored if they have ≥ 3 missing items, 2 studies report 21.5% of questionnaires (49 of 229) and 13.9% of questionnaires (1,077 of 7,760) to be unusable (14,70). In RA ($n = 271$), item response theory suggests the MAF covers the middle range of fatigue severity, broader than Short Form 36 vitality subscale (SF-36 VT) but slightly less than the Functional Assessment Chronic Illness Therapy (Fatigue) (FACIT-F) (57). Floor/ceiling effect data could not be found.

Reliability. *Internal consistency.* Cronbach's alpha for internal consistency was 0.93 in the original VAS version ($n = 133$ patients with RA), 0.92 for the final NRS version ($n = 122$ patients with RA), and 0.92 in knee OA ($n = 44$) (69,77,79).

Test-retest. No significant change in MAF over 3 time points (6–8 week intervals) is reported for patients with RA ($n = 51$) (68); in cancer ($n = 37$), test-retest was $r = 0.87$ over 48 hours (79).

Validity. *Content validity.* The MAF covers a range of fatigue issues (severity, distress, interference with activities, frequency, and change) to create a single, composite score (GFI). The original factor analysis in RA ($n = 35$) showed that the 15 items comprising the GFI load on a single factor (all >0.55) (66). A later analysis in RA ($n = 7,760$) indicated 3 factors: interference with leisure-type activities; interference with bathing/dressing; and fatigue frequency, degree, severity, and distress, with 4 further items loading across all 3 factors equally (70).

Construct validity. In RA ($n = 51$), MAF correlated with depression, pain, disability, and sleep ($r = 0.47$ – 0.58) and very weakly with inflammatory markers (0.12) (68). MAF discriminated between people with RA ($n = 48$) with and without prior history of depression (34.3; SD 10.0 versus 28.8; SD 9.5) (77). In knee OA ($n = 44$), MAF correlated with female sex, pain, depression, anxiety, and cardio-respiratory stamina ($r = 0.52$ – 0.62), but not with muscle (quadriceps) fatigue ($r = 0.01$) (78). In AS ($n = 68$), MAF correlated moderately with pain and hemoglobin (0.39 and -0.38 , respectively), weakly with SF-36 mental health (-0.27), and weakly but not significantly with SF-36 emotional role (-0.22) (72).

Criterion validity. In RA, MAF correlated strongly with the Profile of Mood States fatigue and vigor subscales at 0.84 and -0.62 , respectively ($n = 51$) (68), with a fatigue VAS at 0.8 ($n = 7,760$) (70), and with an NRS of “bothersome fatigue” at 0.69 ($n = 48$) (77). Correlation with SF-36 VT was variable, ranging from -0.79 in RA ($n = 7,760$) to

-0.54 in OA ($n = 137$) and -0.37 in AS ($n = 68$) (55,70,72).

Ability to detect change. In RA ($n = 631$), after 24 weeks of antirheumatic treatments, MAF showed a mean change of -2.1 in patients who did not achieve American College of Rheumatology criteria for 20% improvement in disease activity (ACR20; effect size -0.18), compared to mean change of -14.9 in those who achieved ACR70 (effect size -1.25) similar to findings for FACIT-F and SF-36 VT (57). In FMS ($n = 267$), after 8 weeks of esreboxetine, MAF improved by -6.39 (SE 0.75) compared to -2.82 (SE 1.74) on placebo (75). Based on linear regression analysis on comparative fatigue ratings from patients after paired interviews, the effect size required (mean change/SD at baseline) for an average patient to move to a different fatigue category (i.e., much, somewhat or a little, less or more fatigued) is calculated as 0.75 in RA ($n = 61$) (52) and 0.45 (95% confidence interval [95% CI] 0.25–0.61) in SLE ($n = 80$), where the authors also present this as MAF minimum clinically important difference score of 5.0 (95% CI 2.8–7.2) (53).

Critical Appraisal of Overall Value to the Rheumatology Community

Strengths. The MAF is RA specific and covers numerous aspects of fatigue in order to produce a global score. It has good internal consistency, construct and criterion validity, reliability, and it is sensitive to change.

Caveats and cautions. High levels of missing data are reported, making a substantial proportion of questionnaires unusable. The lack of reference to fatigue on the 11 items asking about interference with activities may reduce clarity for patients who may respond with regards to disability interference.

Clinical usability. The MAF might be useful in clinical practice in providing a global score, while the multiple questions might help identify target areas for therapeutic intervention.

Research usability. The MAF produces a global score based on a range of fatigue impacts. It has reasonable participant and administrative burden, although problems in scoring may arise where there are a large amount of missing items.

MULTI-DIMENSIONAL FATIGUE INVENTORY (MFI)

Description

Purpose. The MFI was originally developed to measure cancer fatigue using a multidimensional, short questionnaire, specifically without any somatic items (80,81). Published in 1995, it was evaluated initially in cancer and chronic fatigue syndrome (CFS) patients and in healthy volunteers who might be physically tired (army recruits) or cognitively tired (junior doctors) (80).

Content. The MFI covers domains of general fatigue (e.g., I feel tired), physical fatigue (e.g., physically I feel only able to do a little), activity (e.g., I feel very active),

motivation (e.g., I dread having to do things), and mental fatigue (e.g., my thoughts easily wander).

Number of items. 20 items, yielding 5 subscales of 4 items each (general fatigue, physical fatigue, reduced activity, reduced motivation, and mental fatigue). Creating a total score is discouraged by the developers.

Response options. 5 check boxes ranging from “Yes that is true,” to “No that is not true.”

Recall period for items. This is stated as “Lately.”

Endorsements. None found for rheumatologic conditions.

Examples of use. In addition to cancer and several long-term conditions (e.g., Parkinson’s disease, liver disease), MFI has been used in a number of studies in rheumatoid arthritis (RA), fibromyalgia syndrome (FMS), ankylosing spondylitis (AS), primary Sjögren’s syndrome (PSS), and systemic lupus erythematosus (SLE) (41,52,53,82–88).

Practical Application

How to obtain. From the developers by e-mail: e.m.smets@amc.uva.nl. Also available by postal mail at the following address: E. M. A. Smets, PhD, Medical Psychology J3-220, Academic Medical Center, University of Amsterdam, PO Box 22660, 1100 DD Amsterdam, The Netherlands. The MFI is free for academic use, charges apply for commercial use.

Method of administration. Patient self-report, pen and paper.

Scoring. Items scored 1–5, with 10 positively phrased items reverse scored. Subscale items summed to produce scores for general fatigue, physical fatigue, reduced activity, reduced motivation, and mental fatigue.

Score interpretation. Scores range from 4–20 with higher scores reflecting greater severity. In terms of normative data, in healthy women ($n = 32$), general fatigue mean \pm SD score was 8.16 ± 3.8 compared to 15.57 ± 4.3 in PSS ($n = 49$), and 12.93 ± 4.5 in RA ($n = 44$), physical fatigue 6.47 ± 3.2 versus 14.06 ± 4.4 and 12.45 ± 5.0 , reduced activity 6.72 ± 3.0 versus 11.32 ± 4.6 and 11.48 ± 4.7 , reduced motivation 6.66 ± 2.4 versus 9.96 ± 4.0 and 9.27 ± 4.1 , and mental fatigue 6.53 ± 3.0 versus 10.31 ± 5.4 and 8.34 ± 4.0 (82). In 53 women with FMS, scores were more severe than RA or PSS with subscale mean \pm SD of 17.9 ± 1.9 , 16.2 ± 2.7 , 15.1 ± 3.9 , 12.9 ± 3.0 , and 14.4 ± 3.5 , respectively (84).

Respondent burden. Time to complete not reported but is likely to be 4–5 minutes. Items appear easy to understand.

Administrative burden. Time to score not reported but is likely to be 4–5 minutes to reverse score some items, then identify and sum subscale items.

Translations/adaptations. Authorized translations in most European languages can be obtained from the developers.

Psychometric Information

Method of development. 24 draft items were generated based on existing literature (80) and pilot in-depth in-

terviews with patients with cancer (81), from which the developers postulated the 5 fatigue domains, for each of which they tried to create brief, positively and negatively phrased items that exclude somatic issues (80). Factor analysis on the 24 items supported the 5 subscales with adjusted goodness of fit index (AGFI) ranging from 0.95–0.98 (111 patients with cancer and 357 patients with CFS) (80); the 4 items with the weakest correlations were later removed, leaving a 20-item scale with 4 items per subscale, which had AGFI properties >0.9 ($n = 97$, 116 patients with cancer) (81). The original MFI-20 had 7 response options (80) but this was revised to the current version with 5 response options following evaluation (81).

Acceptability. The items are brief and easy to understand. Missing item levels are low with 98.2–99.4% completion rates in FMS ($n = 166$) (87). Items do not contain the word fatigue and thus could potentially be interpreted by rheumatology patients as relating to disability (e.g., I think I do very little in a day) or disease activity (e.g., physically I feel I am in a bad condition). In patients with cancer ($n = 116$), 10.4–33.6% scored the best possible score for the different subscales (mental fatigue 33.6%), suggesting a potentially substantial ceiling effect; 4.5–15.7% scored the worst possible score (reduced activity 15.7%), suggesting a lesser, but still potentially important floor effect (81). No data could be found for rheumatology populations.

Reliability. *Internal consistency.* Cronbach’s alpha for most subscales ranged from 0.85–0.89 in 82 patients with RA or PSS, with reduced motivation at 0.68 (86).

Test-retest. In AS and PSS ($n = 40$ and 28, respectively), repeat administrations at between 2 and 42 days gave intraclass correlation coefficients (ICCs) of 0.57–0.85 across the subscales (83,84). The ICC in patients with chronic widespread pain or FMS ($n = 36$) ranged from 0.75–0.92 (87).

Validity. *Content validity.* The MFI covers 5 domains of fatigue, which resonate with qualitative studies in rheumatology (2–4). In PSS, 29 patients scored the coverage of fatigue by the MFI as a mean \pm SD 2.96 ± 0.6 on a scale of 1–4 (“Poorly” to “Very well”) (84).

Construct validity. All subscales differentiated between fatigued and nonfatigued patients with AS ($n = 415$ and 361, respectively) based on a cut off of 5 out of 10 on a fatigue visual analog scale (VAS) (83). All subscales differentiated between healthy women ($n = 32$) and women with RA ($n = 44$), but after controlling for depression, reduced motivation and mental fatigue no longer differentiated patients from controls (82). Subscales correlated strongly with depression at $r = 0.58$ – 0.74 (reduced motivation 0.74) in RA ($n = 44$) (82). Inflammatory indices (erythrocyte sedimentation rate) were not associated with fatigue subscales in PSS, but in RA, Disease Activity Score scores were moderately associated with general fatigue, physical fatigue, and reduced activity at 0.42–0.47 ($n = 49$ and 44, respectively) (82). In RA, associations with Short Form 36 (SF-36) pain were stronger for general fatigue, physical fatigue, and reduced activity (-0.51 to -0.61)

than for mental fatigue and reduced motivation (-0.23 and -0.40 , respectively) ($n = 490$) (85).

Criterion validity. In AS and RA ($n = 812$ and 490 , respectively), 4 subscales correlated with SF-36 vitality subscale at -0.53 to -0.74 , while mental fatigue correlated less strongly (-0.42 and -0.4 , respectively), supporting it as a distinct fatigue concept (83,85). Correlations with a fatigue VAS in RA and PSS were strong for general fatigue (0.7 and 0.77 , respectively), physical fatigue (0.67 and 0.72 , respectively), and reduced activity (0.54 and 0.58 , respectively), but moderate for reduced motivation (0.31 and 0.53 , respectively) and mental fatigue (0.34 and 0.39 , respectively; $n = 48$ and 490 , respectively) (84,85). In FMS, correlations with a fatigue VAS were 0.62 for general fatigue, but 0.32 – 0.36 for the remaining subscales ($n = 165$) (87).

Ability to detect change. Three studies report effect sizes (mean change/SD at baseline). In 40 patients with AS randomized to spa therapy, effect sizes were general fatigue 0.82 , physical fatigue 0.81 , reduced activity 0.28 , reduced motivation 0.52 , and mental fatigue 0.38 , compared to 0.89 in a fatigue VAS (83). In FMS ($n = 1,196$), a significant improvement was seen in a 20-item–totaled MFI score after milnacipran (88). Also using the 20-item–totaled MFI score (not recommended by the developers), and based on linear regression analysis on comparative fatigue ratings from patients after paired interviews, the effect size required for an average patient to move to a different fatigue category (i.e., much, somewhat or a little, less or more fatigued) is calculated as 0.76 in RA ($n = 61$) (52). In SLE, again using the totaled 20 items with a range of 20–100, the effect size was 0.59 (95% confidence interval [95% CI] 0.42 – 0.72), which the authors also present as MFI minimum clinically important difference score of 11.5 (95% CI 8.0 – 15.0) (53).

Critical Appraisal of Overall Value to the Rheumatology Community

Strengths. The MFI provides a profile of 5 domains of fatigue, and has been used in many long-term and rheumatologic conditions. Internal consistency and test-retest show a range of results, while construct and criterion validity are good. Sensitivity to change was good for general and physical fatigue.

Caveats and cautions. A proportion of patients with cancer had minimum or maximum scores, suggesting there may potentially be significant ceiling and floor effects. Criterion validity was variable across subscales. In rheumatology, the wording of some items may be interpreted as relating to disability or disease activity, and sensitivity to change was weak for some subscales.

Clinical usability. An easy scale to complete in clinic, giving information about fatigue profiles.

Research usability. An easy scale to include in an outcome package. However, potential floor/ceiling effects, and interpretation of some phraseology as relating to broader RA issues rather than fatigue, should be considered.

PEDIATRIC QUALITY OF LIFE (PEDSQL) MULTI-DIMENSIONAL FATIGUE SCALE

Description

Purpose. The PedsQL was developed to measure child and parent perceptions of fatigue in pediatric patients and was published in 2002 (89). It was developed in patients with cancer but is intended as a generic measure for pediatric patients. Versions are available for young adults (ages 18–25), teenagers (ages 13–18), and children (ages 8–12) using developmentally appropriate language, with mirror versions for their parents. A “smiley-face” response version is available for young children (ages 5–7), with a written version for parents, and a parent version for toddlers (ages 2–4).

Content. Covers domains of general fatigue (e.g., I feel tired), fatigue related to sleep/rest (e.g., I feel tired when I wake up in the morning), and cognitive fatigue (e.g., it is hard for me to keep my attention on things).

Number of items. 18 items, giving a total fatigue score and including 3 subscales, each of 6 items (general fatigue, sleep/rest fatigue, and cognitive fatigue).

Response options. 5 response options from “Never a problem” to “Almost always a problem.”

Recall period for items. Acute version 7 days, standard version 1 month.

Endorsements. None found for rheumatologic conditions.

Examples of use. The PedsQL Multi-Dimensional Fatigue Scale is a module from PedsQL Measurement model, a modular approach to measuring pediatric health-related quality of life (90). It has been used in studies of mixed rheumatologic disorders, fibromyalgia syndrome (FMS), and juvenile idiopathic arthritis (JIA) (91–93) as well as patients with cancer, cerebral palsy, obesity, cerebral tumours, chronic pain, and multiple sclerosis.

Practical Application

How to obtain. From the web site at URL: <http://www.pedsq.org>. The PedsQL is free to use in unfunded/internally funded research, otherwise a scale of charges apply depending on funding source (see web site for details).

Method of administration. Child and/or parent self-report, pen and paper. Questionnaires should be read aloud to any children unable to read them. For children unable to understand their age-appropriate version, the preceding version should be offered, or the parent proxy used. For the young child (age 5–7), read questions aloud and show smiley faces response choice page for them to select responses.

Scoring. Raw scores (0–4) are reverse scored and transformed to 0–100 (i.e., $0 = 100$, $1 = 75$, $2 = 50$, $3 = 25$, $4 = 0$), so that higher scores reflect better health. All 18 items summed and averaged for a total fatigue score, and the 6 items in each subscale summed and averaged for the 3 subscales' scores (general fatigue, sleep/rest fatigue, and cognitive fatigue), all of which range from 0–100. If $>50\%$ of the items missing, the scale cannot be scored. Scoring instructions can be downloaded from developers' web site.

Score interpretation. Scores range from 0–100 with higher scores reflecting less fatigue. In terms of normative data, in 52 healthy children (ages 5–18), mean \pm SD total fatigue was 80.49 ± 13.33 compared to 76.68 ± 20.523 in children with a range of rheumatologic conditions ($n = 152$) and 55.48 ± 21.19 in FMS ($n = 29$), general fatigue in healthy controls was 85.34 ± 14.95 versus 76.82 ± 23.19 and 48.97 ± 25.14 , sleep/rest fatigue in healthy controls was 75 ± 18.76 versus 71.77 ± 24.27 and 52.36 ± 21.08 , and cognitive fatigue in healthy controls 81.14 ± 17.43 versus 81.30 ± 22.65 and 65.17 ± 24.30 (92). Thus, according to this measure, children with FMS have greater fatigue than children with other rheumatologic conditions, and both groups are worse than healthy controls (although this did not always reach significance in children with broad rheumatologic conditions).

Respondent burden. Estimated at <5 minutes to complete. Items appear easy to read in age-appropriate versions, having undergone cognitive testing.

Administrative burden. Detailed administration instructions need to be read first and suggest some training or practice is required (administration and scoring instructions available on developers' web site). Time to score is not reported, but likely to be ~4–5 minutes to reverse score, transform, sum, and average.

Translations/adaptations. Available in 25 languages (see web site).

Psychometric Information

Method of development. Items and subscales were generated through literature review of adult and pediatric cancer fatigue, patient and parent focus groups, and individual interviews, followed by cognitive interviewing, pretesting, and field testing in cancer (89). Factor analysis appears to have been performed later and data are available for the young adult version (432 university students), where general fatigue and cognitive fatigue loaded on factors 1 and 2, but subscale sleep/rest fatigue loaded across both factors 2 and 3 (94).

Acceptability. In rheumatology, missing item rates of 0.4% and 0.53% are reported for children and 0.7% and 0.8% for parents (91,92). One item might be interpreted in relation to disability or disease activity from rheumatologic conditions rather than fatigue ("I spend a lot of time in bed") and may not be sensitive to rheumatoid arthritis fatigue. Data on floor/ceiling effects could not be located.

Reliability. *Internal consistency.* In rheumatology, Cronbach's alpha ranged from 0.88–0.95 for the total scale and 3 subscales for all age-appropriate versions ($n = 163$) (91); in FMS, Cronbach's alpha ranged from 0.76–0.94 ($n = 29$) (92).

Test-retest. No data could be found for the PedsQL.

Interrater reliability. Child and parent (proxy) fatigue scores correlated in a rheumatology population ($n = 163$) at 0.85–0.93 for total fatigue and all subscales (91).

Validity. *Content validity.* Items were generated through literature review, and patient and parent focus groups and individual interviews in cancer populations (89).

Construct validity. In 175 children with a range of rheumatologic conditions, total fatigue and the 3 subscales

correlated strongly with quality of life, pain, physical and psychosocial health, and emotional, social, and school functioning at 0.53–0.91, while all scales had slightly lower, but still positive, associations with daily activities (0.48–0.58); the highest association for total fatigue was with psychological health (0.84), for general fatigue was physical health (0.80), for sleep/rest fatigue was psychological health (0.73), and for cognitive fatigue was poor school functioning (0.77) (91). In 29 children with FMS, the highest correlation for total fatigue and the subscales was always with quality of life (0.69–0.81) (92). Correlation with physician global opinion of disease activity was moderate (–0.30 to –0.39) in a broad rheumatology population (91). Children with inactive JIA ($n = 29$) showed less fatigue on all subscales than children with active disease ($n = 18$) (93).

Criterion validity. No data could be found in rheumatology populations. In 432 university students, PedsQL young adult version correlated moderately to strongly with the single item Short Form 8 vitality subscale at 0.56 (general fatigue), 0.54 (total fatigue), 0.4 (cognitive fatigue) and 0.36 (sleep/rest fatigue) (94). In Chinese pediatric patients with cancer, correlation with the fatigue scale-children was –0.45 to –0.61 ($n = 108$) (95).

Ability to detect change. No sensitivity to intervention data found for any population.

Critical Appraisal of Overall Value to the Rheumatology Community

Strengths. The PedsQL is a module from a well-established quality of life measurement system. It reports total fatigue and a range of subscales and has been evaluated in many pediatric long-term conditions. In rheumatology, internal consistency is good, and cognitive fatigue correlates with poor school functioning.

Caveats and cautions. Criterion validity data could not be found for rheumatologic populations, while stability and sensitivity data could not be found for any population. Three subscales appeared to have been generated through a literature review (89) but on later factor analysis (94), the sleep/rest subscale loaded equally across 2 factors, not on a single factor. Total fatigue correlates strongly with psychological status in rheumatology.

Clinical usability. Appears to be a useful tool for clinical use, which is quick to complete.

Research usability. A relatively easy tool to use, but criterion, stability, and sensitivity data are required.

PROFILE OF FATIGUE (PROF)

Description

Purpose. The ProF was developed to characterize patterns of fatigue associated with primary Sjögren's syndrome (PSS) and published in 2003 (96).

Content. Contains somatic fatigue items for needing to rest (e.g., feeling exhausted), difficulty getting started (e.g., hard to get going), low stamina (e.g., hard to keep going), and weak muscles (e.g., feeling weak), and contains mental

fatigue items for concentration (e.g., not thinking clearly) and memory (e.g., forgetting things).

Number of items. 16 items giving a total fatigue score, including 6 facets: need rest (4 items), poor starting (4 items), low stamina (2 items), and weak muscles (2 items) can be combined to form the somatic domain. Facets poor concentration (2 items) and poor memory (2) can be combined to form the mental domain.

Response options. 8 response options asking about how patients felt when they were at their worst, ranging from “Not at all” to “As bad as imaginable” (0–7).

Recall period for items. Last 2 weeks.

Endorsements. Developed by the UK Sjögren’s Interest Group (96).

Examples of use. Used in PSS, systemic lupus erythematosus (SLE), and rheumatoid arthritis (RA) studies (63,86,96–102).

Practical Application

How to obtain. Obtained from the developers by e-mail: Simon.Bowman@uhb.nhs.uk. The ProF is free to use.

Method of administration. Patient self-report, pen and paper.

Scoring. The 6 facet scores can be reported alone, or combined to form 2 domain scores, or a total score. Facet scores (0–7) are formed by summing and averaging the items in each facet: need rest (items 1–4), poor starting (items 5–8), low stamina (items 9 and 10), weak muscles (11 and 12), poor concentration (13 and 14), and poor memory (15 and 16). Domain scores (0–7) are formed by summing and averaging items 1–12 for somatic fatigue, and items 13–16 for mental fatigue. A total fatigue score (0–7) is created by summing and averaging all 16 items.

Score interpretation. Scores for facets, domains, and total score all range from 0–7 with higher scores reflecting greater fatigue severity. In terms of normative data, in the somatic fatigue domain, the 4 facets had mean values of 1.4–2.2 in 103 healthy controls, and all were significantly different to patients with PSS, RA, and SLE who had mean scores of 2.7–4.4 ($n = 18, 18,$ and 11 , respectively); in the mental fatigue domain, the 2 facets both had mean values of 1.5 in healthy controls, which were significantly different to patients with PSS and SLE (mean 2.3–2.5) but not patients with RA (mean 1.9–2.1) (96). The developers used the difference between controls and patients to identify cut points for a “case” of fatigue; a “case” for a facet is someone who scores >2 out of 7 in that facet except for the need rest facet, where ≥ 3 out of 7 is required (96). A fatigue case for the somatic fatigue domain is a patient who is a “case” in at least 2 of the 4 related facets, while a fatigue case for the mental fatigue domain is a patient who is a “case” in at least 1 of the 2 related facets (96).

Respondent burden. Time to complete not reported but likely to be 4–5 minutes. Items were developed with patients and do not appear difficult, with the possible exception of one (“It’s a battle”), which might not be clear to interpret.

Administrative burden. Time to score not reported, likely to be 3–4 minutes to calculate facet, domain, and total scores.

Translations/adaptations. Translated into Swedish using appropriate linguistic methodology (98). A shorter, 6-item ProF was published in 2009 and contains 1 item for each of the 6 facets (97). However, the long version is the most commonly used. A state version (“Right now” rather than “Over the past 2 weeks”) has been used (86).

Psychometric Information

Method of development. The ProF contains 16 items from the 64-item Profile of Fatigue and Discomfort–Sicca Symptoms Inventory (96). Draft items were generated using the words of patients with PSS, collected in diaries that were later discussed in focus groups, then subsequently piloted with patients with PSS, RA, and SLE ($n = 18, 18,$ and 11 , respectively) who generated 5 clusters of similar statements concerning 4 somatic and 1 mental facet of fatigue (96). The mental component was then split into 2 facets, and all 6 were evaluated in patients with PSS, RA, and SLE ($n = 137, 174,$ and 66 , respectively) and controls ($n = 103$) (96). Factor structure in 82 patients with PSS or RA showed 5 rather than 6 clear factors, with facets need rest and low stamina not being well differentiated, although the somatic and mental fatigue domains were well differentiated (86). The 6 items of the short ProF load on 2 factors, somatic and mental fatigue (97).

Acceptability. Items were developed with patients with PSS and appear easy to read, with missing data reported as only 0.6% (96). The item “It’s a battle” might not be answered specifically about fatigue. Authors of one study reported that no floor or ceiling effects were observed (98).

Reliability. *Internal consistency.* Cronbach’s alpha for the total fatigue score was 0.97, ranged from 0.91–0.93 for the 2 domains, and from 0.9–0.97 for the 6 facets in patients with PSS (98).

Test-retest. In patients with PSS ($n = 12$), over a median 3 days (range 0–7 days), the weighted kappa coefficient for total fatigue was 0.63 (interquartile range [IQR] 0.48–0.75); over a median 12 days (range 0–71 days) it was 0.51 (IQR 0.48–0.55; $n = 37$) (98).

Validity. *Content validity.* The ProF was derived from focus group discussion of PSS patient diaries (96). In the Swedish translation, 19 of 20 patients and both rheumatologists considered the items covered adequate content for PSS fatigue (98).

Construct validity. In terms of somatic fatigue facets, in PSS ($n = 18$) these correlated with the World Health Organization Quality of Life (WHOQoL) physical domain (–0.62 to –0.69), weak muscles and low stamina correlated with WHOQoL energy at –0.6, and needs rest and poor starting correlated with anxiety and depression at –0.36 to –0.52 (96). In terms of mental fatigue facets, in PSS ($n = 18$) these correlated with Short Form 36 (SF-36) mental health domain (–0.27 to –0.44), WHOQoL psychological domain (–0.32 to –0.47), and anxiety and depression (–0.34 to –0.48) (96). Sensitivity of facets to classify PSS correctly ranged from 67% (poor memory) to 88% (low stamina), with specificity from 66% (poor starting) to 73% (poor concentration; $n = 18$) (96). Somatic and mental domains had generally weak associations with accumulative systemic damage in PSS (0.02–0.34) (99). ProF dem-

onstrated that somatic and mental fatigue deteriorate during the day (39 women with PSS or RA) (101).

Criterion validity. In patients with PSS, somatic and mental fatigue domains correlated strongly with SF-36 vitality subscale -0.84 and -0.63 , respectively), and with a fatigue visual analog scale (VAS; 0.73 and 0.64 , respectively; $n = 50$) (98). In 82 patients with PSS or RA, all ProF facets correlated strongly with relevant domains from the Multi-Dimensional Fatigue Inventory (0.65 – 0.86) (86). Patients with PSS, classified as fatigued by a score of >4 on the Fatigue Severity Scale (FSS), had significantly higher mean \pm SD ProF scores than those not classified as fatigued on the FSS: somatic fatigue 4.1 ± 1.5 compared to 2.2 ± 1.6 , mental fatigue 3.3 ± 1.7 compared to 0.9 ± 1.6 ($n = 94$) (102). The 6 items of the short ProF correlated with the original 16-item long version domains (0.78 to >0.9) and the short ProF somatic and mental fatigue domains correlated with a fatigue VAS (0.77 and 0.55 , respectively; $n = 43$ PSS) (97).

Ability to detect change. In 17 patients with PSS, somatic fatigue improved significantly in patients randomized to rituximab ($P = 0.009$) but not to placebo ($P = 0.087$; actual data not provided) (100).

Critical Appraisal of Overall Value to the Rheumatology Community

Strengths. The ProF was developed specifically in and for patients with PSS and measures a range of fatigue concepts. Internal consistency is strong and construct and criterion validity are good.

Caveats and cautions. Test–retest reliability is only moderate, and actual data on sensitivity would be helpful. Data on factor structure support 5 rather than 6 factors (facets), and some studies contain relatively few numbers on which to evaluate 6 facets.

Clinical usability. Appears to be a useful, easy tool for clinic use.

Research usability. Appears appropriate for research use, but the caveats above should be considered. There appears to be stronger evidence for the domain structure (somatic, mental fatigue) than the 6-facet structure.

SHORT FORM 36 VITALITY SUBSCALE (SF-36 VT)

The SF-36 is a multidimensional, general health status patient-reported outcome measure (PROM) containing subscales for 8 domains. The detailed review of the entire instrument is presented in the article “Adult Measures of General Health and Health-Related Quality of Life” elsewhere in this issue. This section reports only additional data specific to the SF-36 VT.

Description

Purpose. The SF-36 VT was developed to measure vitality, conceptualized as a single continuum from energy to fatigue, in general and clinical populations, and the complete SF-36 was first published in 1992 (103). The second version was published in 2000 (SF-36v2, see article

on Adult Measures of General Health), and in SF-36v2, 1 vitality question has been reworded (from “full of pep” to “full of life”). The SF-12v2, a shorter version published at the same time, also includes a vitality subscale. Most articles do not state whether they have used SF-36 VT or the reworded SF-36v2 VT.

Content. The SF-36 VT covers energy (e.g., feeling full of pep) and fatigue (e.g., feeling worn out), while SF-12 VT contains 1 item on energy.

Number of items. Original and revised versions have 4 items in the SF-36 VT (2 on energy and 2 on fatigue) to produce a single score; SF-12 VT has 1 item (energy).

Response options. In the original SF-36, the vitality subscale had 6 response options ranging from “All of the time” to “None of the time.” In SF-36v2 and SF-12v2, these have been reduced to 5 options to improve psychometric performance (see developer’s web site at URL: <http://www.sf-36.org/tools/sf36.shtml>).

Recall period for items. 4 weeks, plus a 1-week acute version.

Endorsements. None found for rheumatologic conditions.

Examples of use. SF-36 VT can be aggregated with other subscales to form the mental component score and in earlier literature, SF-36 VT data were not always reported separately. However, with the recent evidence that fatigue is a rheumatology patient priority and part of core data in several conditions (5,6), SF-36 VT data are increasingly being provided. SF-36 VT reports in musculoskeletal studies include data from rheumatoid arthritis (RA), psoriatic arthritis (PsA), ankylosing spondylitis (AS), primary Sjögren’s syndrome (PSS), systemic lupus erythematosus (SLE), fibromyalgia syndrome (FMS), and osteoarthritis (OA) (14,26,52,53,57,70,83,98,104–117). Only a few studies using the SF-12 could be found (all OA), and these did not report the single vitality item separately. Overall, the SF-36 has been used in 14,000 articles, and the revised SF-36v2 in 260, as reported on the developer’s web site at the following URL: <http://www.qualitymetric.com/WhatWeDo/GenericHealthSurveys/tabid/184/Default.aspx>.

Practical Application

How to obtain. See article on Adult Measures of General Health for web site information on access and cost.

Method of administration. Patient self-report. A range of administration modalities is described in the article on Adult Measures of General Health.

Scoring. As energy items are positive and fatigue items are negative, some items need to be recoded before scoring, then they are summed and transformed to a 0–100 scale (see article on Adult Measures of General Health for details of computerized scoring systems, norm-based algorithms, and handling missing data). The only difference in scoring between the original SF-36 and the SF-36v2 is in the contribution of vitality to the mental and physical component scores. In the original scoring system, the SF-36 VT subscale only contributes to the mental component score but factor analysis in RA ($n = 1,030$) suggests that vitality correlates equally with both the mental and physical components (0.61 and 0.53 , respectively) (113). Therefore, in

the revised SF-36v2 scoring system, vitality is now included in both the physical and the mental component scores, but still contributes a larger weighting to the mental component score.

Score interpretation. Scores range from 0–100 with higher scores representing less fatigue. In terms of normative data, age- and sex-based norms are available for many countries (see article on Adult Measures of General Health). Rheumatology studies report mean SF-36 VT scores for healthy controls of 57.4 and 62.2 ($n = 77$ – 606) (63,105). This compares to SF-36 VT mean \pm SD of 43.4 ± 23.4 in RA, 43.0 ± 24 in AS, 38.9 in PSS, 35.9 ± 23.1 in SLE, 27.1 ± 21.1 in FMS, and 25.7 ± 20.1 in PsA, although SDs are wide ($n = 152$ – $13,722$) (63,106,109). However, 2 studies report higher vitality than healthy controls in RA, and in patients with OA 2–10 years after arthroplasty (107,108).

Respondent burden. SF-36 VT has only 4 items and therefore would take only 1 minute for respondents to complete, but it is administered with the whole SF-36 questionnaire, which may take up to 10 minutes to complete. The format is not difficult to understand.

Administrative burden. Scoring the SF-36 VT is relatively quick, but it is rarely administered in isolation and scoring the whole SF-36 is more complex and takes longer. Computerized systems are available for purchase from Quality Metric (see article on Adult Measures of General Health).

Translations/adaptations. Available in over 120 languages (see article on Adult Measures of General Health).

Psychometric Information

Method of development. A detailed review of the development of the entire SF-36 is found in the Adult Measures of General Health article. The SF-36 VT items were generated from a review of existing instruments, with the aim of including a balance of favorably and unfavorably worded items (103). There was no patient involvement.

Acceptability. Most items appear acceptable to patients and clearly relate to fatigue or energy. However, in the original SF-36, the item “Full of pep” has the potential to cause confusion in countries where it is not a common term, and has been replaced in SF-36v2 by “Full of life.” In RA ($n = 1,030$), 2.3–5.8% of respondents omitted answers in each of the 4 SF-36 VT items (113). In patients with OA 2–10 years after arthroplasty ($n = 58$), no floor effects were found but problematic ceiling effects (defined as $\geq 15\%$) were found, with 18% of respondents recording best possible scores (107). In contrast, no floor or ceiling effects were found in another report of patients with OA up to 5 years after arthroplasty ($n = 59$ – 135) (110). In RA ($n = 271$), item response theory suggests SF-36 VT covers mainly the less severe range of fatigue severity, in comparison to the Multi-Dimensional Assessment of Fatigue (MAF) and Functional Assessment Chronic Illness Therapy (Fatigue) (FACIT-F), which cover a broader range (57).

Reliability. See article on Adult Measures of General Health for detailed reliability review of SF-36.

Internal consistency. For SF-36 VT, in RA ($n = 631$), Cronbach's alpha was 0.84–0.88 over 3 time points (57),

and in OA ($n = 62$), all SF-36 domains, including SF-36 VT, had a Cronbach's alpha of 0.75–0.94 (107).

Test-retest. In one OA study ($n = 62$, mean age 58 years, 2–10 years postarthroplasty), 4-week test-retest reliability of the SF-36 VT was $r = 0.92$ (107); in contrast, another OA study found very poor 1-week stability at $r = 0.03$ ($n = 21$, mean age 70 years) (111). In RA ($n = 150$), intraclass correlation coefficient for SF-36 VT was 0.91 (95% confidence interval [95% CI] 0.86, 0.94) over 2 weeks (114).

Validity. See article on Adult Measures of General Health for detailed validity review of SF-36.

Content validity. SF-36 VT covers both energy and fatigue, but these may not be opposite ends of a single continuum, as feeling energized is a positive health state rather than the absence of fatigue. Thus, while a person who is not fatigued would score 0 out of 100 on a scale containing 4 fatigue items, they would potentially score 50 on the SF-36 VT by answering “no” to both the energy and fatigue items, due to a lack of energy rather than the presence of fatigue. This is further supported by data from an RA study, where the SF-36v2 VT items loaded across 2 separate factors: “Full of life” and “Lot of energy” loaded on a factor with items feeling happy, peaceful, and healthy, while “Feel tired” and “Worn out” loaded on a factor with items feeling down and feeling sad ($n = 401$) (115).

Construct validity. In RA ($n = 86$), SF-36 VT correlated strongly with disability ($r = 0.56$), and weakly to moderately with physician global assessment, patient global assessment, pain, tender joints, and inflammatory markers (-0.27 to -0.37) (112); correlation with anxiety, depression, and helplessness is reported as 0.28–0.50 ($n = 229$) (14). SF-36 VT discriminated between patients with RA with low versus moderate Disease Activity Score in 28 joints (DAS28) but not moderate versus high DAS28 ($n = 200$) (114).

Criterion validity. Data on criterion validity in rheumatology populations are varied for SF-36 VT. For example, correlation with the MAF ranges from very strong (0.79) in RA (70), to strong in OA (-0.54) (55) but only moderate in AS (-0.37) (72). Correlation with a fatigue visual analog scale (VAS) ranges from very strong (0.8) in RA (70), to strong in AS (0.64) (83). Correlation with the facet and domain scores of the Profile of Fatigue ranges from very strong (-0.84) to strong (-0.63) (PSS, $n = 50$) (98) and with the Multi-Dimensional Fatigue Inventory domains from strong (0.73) to only moderate (0.42) (AS, $n = 812$) (83). In the evaluation of the Bristol RA Fatigue (BRAf) Multi-Dimensional Questionnaire and its 4 subscales, correlations with SF-36 VT were moderate to strong (-0.40 to -0.68) but in every instance these were lower than the strong correlations between BRAf and MAF or FACIT-F (-0.52 to -0.83) (14).

Ability to detect change. See article on Adult Measures of General Health for detailed review of entire SF-36 ability to detect change. In patients with RA ($n = 631$) receiving 24 weeks of anti-tumor necrosis factor (TNF) therapy, SF-36 VT showed a mean improvement of 5.2 in patients who did not achieve American College of Rheumatology 20% criteria for improvement in disease activity (ACR20; effect size 0.25), compared to 31.4 in those who achieved

ACR70 (effect size 1.52), which were similar to changes demonstrated by the FACIT-F (57). In PsA ($n = 313$), 24-week treatment with anti-TNF therapy produced a mean \pm SD improvement of 12.8 ± 21 compared to 1.7 ± 19.1 in placebo (59). In SLE ($n = 93$), while the Chalder Fatigue Questionnaire showed significant improvements in fatigue following exercise compared to relaxation or no intervention, the SF-36 VT did not show improvement, but neither did the Fatigue Severity Scale (FSS) or a fatigue VAS (26). In patients with AS ($n = 40$) randomized to etanercept or placebo, SF-36 VT showed an effect size of 0.54 for treatment at 1 month and 0.69 at 4 months, while the FSS was not responsive at 1 month (effect size 0.15) but showed a similar effect size at 4 months (0.43) (50). In patients with OA of the hip ($n = 135$) and knee ($n = 59$) receiving total joint replacement, SF-36 VT showed effect sizes of 1.0 and 0.6, respectively, at 6 months (97). In anti-TNF therapy for patients with RA ($n = 258$), SF-36 VT showed a change of 16%, which was smaller than change in a fatigue VAS (23%), and changes in tender joint count and patient global assessment (24% and 25%, respectively) (117). Based on linear regression analysis on comparative fatigue ratings from patients after paired interviews, the effect size (mean change/SD at baseline) required for an average patient to move to a different fatigue category (i.e., much, somewhat or a little, less or more fatigued) is calculated as 0.67 in RA ($n = 61$) (39) and 0.44 (95% CI 0.25, 0.60) in SLE ($n = 80$), which the authors also present as an SF-36 VT minimum clinically important difference score of -10.7 (95% CI -15.5 , -5.9) (53).

Critical Appraisal of Overall Value to the Rheumatology Community

See article on Adult Measures of General Health for overview of the entire SF-36.

Strengths. The SF-36 VT has been used across many rheumatologic conditions and in many studies. Internal consistency, construct validity, and sensitivity are good. The SF-36 VT may be useful when wishing to compare fatigue with other conditions and healthy populations.

Caveats and cautions. In rheumatology populations, there are conceptual concerns over the assumption of fatigue and energy as opposite ends of a single continuum, as energy is a positive health state, rather than an absence of fatigue, which is supported by data demonstrating the 2 energy and 2 fatigue items load on 2 separate factors. There are some reports that vitality is higher in OA and RA than in healthy controls, reports of SF-36 VT ceiling effects, and item response theory suggests that SF-36 VT may not capture higher levels of fatigue. While criterion validity is good with the MAF and a VAS in RA, there are a range of correlations with other fatigue PROMs in rheumatology populations, some as low as 0.37. The conflicting data on test-retest performance in rheumatology (ranging from 0.03–0.92) are concerning.

Clinical usability. The SF-36 VT would be easy to use in clinical practice, but although it was designed for both clinical practice and population surveys, it is not commonly used in clinical care.

Research usability. The SF-36 VT is frequently used in rheumatology research, and provides a global fatigue

score. However, the above caveats from data in rheumatology populations should be noted. If the entire SF-36 is being administered in order to capture many health domains to compare with other populations, then researchers may wish to consider whether an additional brief fatigue measure would be helpful.

VISUAL ANALOG SCALES (VAS)

Description

Purpose. Fatigue VAS are unidimensional measures aiming to capture an aspect of fatigue, typically severity or intensity.

Content. Fatigue VAS typically comprise a 100-mm horizontal line, anchored by 2 statements representing extreme ends of a single fatigue continuum (e.g., severity or intensity). However, there is no standardized fatigue VAS for use in rheumatology populations. A systematic review (1996–2004) identified 26 rheumatology studies reporting a fatigue VAS, of which only 4 provided a validation reference, and these related to pain VAS validation; only 10 of 26 were described in detail and only 3 of 26 were identical for content (7). The more recent rheumatology literature, explored for this review, shows the situation continues with multiple VAS versions frequently not described in detail or referenced. Therefore, it appears that researchers often create their own fatigue VAS (stem question and anchors) for individual studies. In terms of content, the stem question may describe tiredness, fatigue, fatigue/tiredness, or unusual fatigue (118–121).

Number of items. A single-item scale.

Response options. Respondents are typically instructed to make a mark across or on the VAS line to describe the point between the 2 anchors that best reflects their fatigue status. Response options are not standardized and depend on the nature of the question, with researchers creating their own. Examples include “Not at all tired” to “Very tired,” “No fatigue” to “Total exhaustion,” “None” to “As bad as it could be,” “No problem” to “Major problem,” “Absence of fatigue” to “Worst condition imaginable,” “No fatigue” to “Complete fatigue,” and “No fatigue” to “Intolerable fatigue” (28,118–120,122–124).

Recall period. Usually 1 week (but often not reported in papers).

Endorsements. None found for rheumatologic conditions.

Examples of use. Used extensively in rheumatologic conditions, e.g., rheumatoid arthritis (RA), systemic lupus erythematosus (SLE), ankylosing spondylitis (AS), psoriatic arthritis (PsA), primary Sjögren’s syndrome (PSS), fibromyalgia syndrome (FMS), and osteoarthritis (OA) (14–16,28,70,84,85,88,118–132).

Practical Application

How to obtain. Researchers often create their own VAS.

Method of administration. Self-report by patient using pen and paper.

Scoring. A ruler is used to measure the distance from the left hand anchor to the respondent’s mark on the VAS

line. While most fatigue VAS range from 0–100 mm, some use a 0–10-cm scale. One variation uses a 15-cm VAS and calculates a score ranging from 0–3, although the rationale for this variation is not provided (125,126). Caution should be taken when scoring VAS, as photocopying can distort (lengthen) the line (133).

Score interpretation. Typically, 0–100 or 0–10 with a higher score representing a greater severity or intensity of fatigue. In terms of normative data, VAS fatigue mean \pm SD scores (mm) have been reported in healthy controls ($n = 144$) as 20.5 ± 0.02 (124). In comparison, examples of rheumatology population means \pm SD are 49.7 ± 2.0 in RA, 43.3 ± 2.0 in hand OA, 50.4 ± 30.6 in SLE, 40.8 ± 31.7 in PsA, 74.4 ± 12.9 in PSS, 6.7 ± 2.0 on a scale of 0–10 in AS, and 7.21 ± 1.91 on a scale of 0–10 in FMS ($n = 20$ – 202) (84,124,127–130). In 1 RA study, researchers defined fatigue as clinically relevant at VAS ≥ 20 mm and high fatigue at VAS ≥ 50 mm (9). In an AS study, researchers defined fatigue as a major symptom at ≥ 50 mm (83); elsewhere, researchers have defined substantial fatigue in patients with RA, OA, and FMS as ≥ 2 of 3 on a VAS scaled from 0–3, with a cut off of ≥ 1 for mild fatigue (125). However, none of these studies report the rationale for the cut points.

Respondent burden. A VAS scale usually takes <1 minute to complete.

Administrative burden. VAS scales are easy to administer and to score. The availability to the patient of their prior VAS score may affect subsequent responses; therefore, researchers should be consistent in whether or not these are available during completion (133).

Translations/adaptations. There is no standard fatigue VAS to translate, but researchers will create their own versions in their own languages. Ideally these should be grounded in patients' words and concepts (13).

Psychometric Information

Method of development. VAS scales have their theoretical foundations in psychological theories of response to sensory stimuli and have a long history in psychometric research to measure subjective states (134). Reports of how the stem question and wording are developed are rare. The Bristol RA Fatigue VAS (3 single items on each of severity, coping, and effects) were developed in collaboration with patients, based on qualitative interviews, then focus groups and cognitive interviewing to design the VAS (13,14): patients chose severity anchors of "No fatigue" to "Totally exhausted," effect anchors of "No effect" to "A great deal of effect," and coping anchors of "Not at all well" to "Very well." This wording is also used in the 3 Bristol Rheumatoid Arthritis Fatigue Numerical Rating Scales (BRAFNRS), which the developers recommend in preference to using VAS, as the NRS versions show stronger psychometric properties and better practical and conceptual considerations than VAS (see BRAFNRS section) (14–16).

Acceptability. In general, most patients find VAS scales easy to understand, and 1 FMS study reports a 99.4% completion rate (87). However, some patients do not understand the VAS measurement concept and may mark

above the line or beyond the anchors (133). In RA ($n = 7,760$, 307), 6.4–9% scored best possible score and 1.8–2% scored worst possible (9,70). Fatigue VAS covers most of the full range of fatigue levels (70).

Reliability. *Test-retest.* In RA over 1–2 days, the intraclass correlation coefficient (ICC) of a fatigue VAS was 0.74 (95% confidence interval [95% CI] 0.65, 0.81; $n = 122$) (121). In PSS over a median 14 days the ICC was 0.66 (95% CI 0.39, 0.83; $n = 48$) (84).

Validity. *Content validity.* VAS are unidimensional measures and as they are not standardized, the content largely depends on the construct the researchers wish to explore, and the language they use to capture it.

Construct validity. In a study of 2 RA populations ($n = 238$ and 274 , respectively), fatigue VAS was positively associated with Disease Activity Score at $r = 0.43$ and $r = 0.69$, and with pain at $r = 0.63$ and $r = 0.68$ (9); in another RA study ($n = 22$) fatigue VAS correlated very strongly with pain (0.8) and strongly with sleep (0.6) (119). In AS ($n = 639$), fatigue VAS correlated strongly with axial pain (0.58) but weakly with global pain (0.24) or not with C-reactive protein (-0.07) (120). In FMS ($n = 50$), fatigue VAS correlated strongly with pain (0.6) but moderately with sleep (0.3), which was not statistically significant (119).

Criterion validity. In RA, fatigue VAS very strongly correlated with the Multi-Dimensional Assessment of Fatigue (MAF) at 0.80, and strongly with Short Form 36 vitality subscale (SF-36 VT) at 0.71 ($n = 7,760$) (70). In FMS and in PSS, fatigue VAS correlated strongly with Multi-Dimensional Fatigue Inventory (MFI) total (general fatigue) at 0.62 and 0.70, but moderately with MFI mental fatigue and reduced motivation (0.32–0.39), and ranged between moderate and strong for physical fatigue and reduced activity (0.36–0.67) (84,87). In AS ($n = 812$), fatigue VAS correlated with SF-36 VT at -0.64 (83).

Ability to detect change. In RA ($n = 5,155$), fatigue VAS was more sensitive to changes in pain and patient global opinion over 6 months than MAF or SF-36 VT, although there was no difference in performance between them in relation to disability or quality of life (70). In anti-tumor necrosis factor (anti-TNF) therapy for patients with RA ($n = 391$), fatigue VAS showed change of 23% (treatment VAS difference -16.8 ; 95% CI -22.8 , -10.8), similar to improvement in tender joint count and patient global assessment (24% and 25%, respectively) and greater than change in SF-36 VT (16%) (117). Similar changes were seen with anti-TNF therapy in patients with RA and in patients with PsA ($n = 30$ and 146 , respectively) with improvements in fatigue VAS of -17 and -12.0 (9,131). In FMS ($n = 40$), fatigue VAS (scale 0–10) improved by 2.7 (SD 0.75) after an aquatic exercise program compared to 0.26 (SD 0.35) in controls, with similar changes seen in SF-36 VT (130). In AS ($n = 40$), a fatigue VAS showed an effect size of 0.89 from spa therapy (83) while in another study ($n = 256$) improvement in hemoglobin was associated with improvement in fatigue VAS (129). In RA ($n = 307$), minimum clinically important difference (MCID) for a fatigue VAS of 0–10 was between -0.82 and -1.12 for improvement and 1.13 and 1.26 for worsening, based on a

transition question (122); this is similar to the MCID of 10 in a fatigue VAS of 0–100, found by Wells et al (123). In SLE ($n = 202$), the MCID for a fatigue VAS (0–100) was -13.9 for improvement and 9.1 for worsening, based on a transition question (127). In PsA ($n = 200$), smaller MCIDs for a fatigue VAS were found at -8.15 for improvement and 3.63 for worsening, also based on a transition question (128). Patients with lower scores required a larger change in their fatigue VAS to report worsening, and people with higher scores required a larger change to perceive improvement, which might be related to floor/ceiling effects or different interpretations at different points in the VAS (122).

Critical Appraisal of Overall Value to the Rheumatology Community

Strengths. VAS are one of the most frequently used tools to measure fatigue and have been used for many years, with a number of studies supporting their validity for measuring fatigue. They are quick and simple to administer and score, and minimal in terms of respondent burden. In rheumatology, test–retest is good in RA, but weak in PSS, and construct validity is good. Criterion validity is good with MAF, but weaker with SF-36 VT and MFI, while sensitivity to change is good and may be stronger than SF-36 VT. VAS are suitable for use where a global fatigue assessment only is required.

Caveats and cautions. There are many practical and conceptual concerns with VAS, including: VAS length distorts with photocopying; some patients have difficulty understanding the abstract nature of a VAS; the VAS format cannot be administered online or by phone; patients avoid the extreme ends of a VAS; the precision of a 100-mm line may not be appreciated by respondents who tend to consider responses in multiples of 5–10 mm blocks; and when patients' VAS are plotted against their ordinal fatigue scales of none/mild/moderate/severe, VAS scores show considerable overlap across categories (e.g., VAS ratings of 10 and 100 both appear in "moderate" categories) (18,133,135,136). Lack of a standardized fatigue VAS limits comparisons between studies and makes replication across studies difficult, and validation is largely based on accumulative data on a number of differently phrased VAS. A standardized fatigue VAS format, developed with patients with RA (BRAFF VAS) has been tested, and found to be marginally less robust than the identical NRS versions; therefore, in view of the other VAS concerns listed here, the NRS versions are recommended by the developers (see BRAFF NRS section) (13–16). However, evaluation of the BRAFF short scales found that the NRS scored higher than the VAS indicating that the two different patient-reported outcome measures formats are not interchangeable (14). Researchers often create their own VAS, and should take note of studies with pain VAS, which led to recommendations that VAS should be 100-mm long as VAS of <100 mm are inclined to greater error variance, horizontal VAS should be used as they have a more uniform distribution of scores than vertical VAS, anchor wording should be at each end and not below or above the VAS, and that end markers should be

placed at right angles to the VAS (not arrows or other markers) (136).

Clinical usability. The fatigue VAS is easy to use in clinical practice, to identify patient concerns and response to treatment. As a single item, VAS are limited in the information that they yield.

Research usability. The fatigue VAS is frequently used in rheumatology research and provides a global fatigue score. However, the above caveats should be noted, and the use of NRS considered. A multidimensional assessment may provide a more complete picture and improve understanding of the clinical relationships of fatigue and hence potential treatment.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published.

ROLE OF THE STUDY SPONSOR

GlaxoSmithKline had no control over or contribution to the BRAFF Scales design, data collection, or analyses. Publication was not contingent on their approval.

REFERENCES

- Hewlett S, Nicklin J, Trehan GJ. Fatigue in musculoskeletal conditions. Topical Reviews Series 6: No 1. Arthritis Research UK; 2008. URL: http://www.arthritisresearchuk.org/files/6641_25022010173830.pdf.
- 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 Rheum* 2005;53:697–702.
- Repping-Wu H, Uitterhoeve R, van Riel P, van Achterberg T. Fatigue as experienced by patients with rheumatoid arthritis (RA): a qualitative study. *Int J Nurs Stud* 2008;45:995–1002.
- Power JD, Badley EM, French MR, Wall AJ, Hawker GA. Fatigue in osteoarthritis: a qualitative study. *BMC Musculoskelet Disord* 2008; 9:63.
- Kirwan JR, Minnock P, Adebajo A, Bresnihan B, Choy E, de Wit M, et al. Patient perspective: fatigue as a recommended patient centered outcome measure in rheumatoid arthritis. *J Rheumatol* 2007;34:174–7.
- Choy EH, Arnold LM, Clauw DJ, Crofford LJ, Glass JM, Simon LS, et al. Content and criterion validity of the preliminary core dataset for clinical trials in fibromyalgia syndrome. *J Rheumatol* 2009;36:2330–4.
- Hewlett S, Hehir M, Kirwan JR. Measuring fatigue in rheumatoid arthritis: a systematic review of scales in use. *Arthritis Rheum* 2007; 57:429–39.
- Zonna-Nacach A, Roseman JM, McGwin G Jr, Friedman AW, Baethge BA, Reveille JD, et al. Systemic lupus erythematosus in three ethnic groups. VI. Factors associated with fatigue within 5 years of criteria diagnosis. *Lupus* 2000;9:101–9.
- Pollard LC, Choy EH, Gonzalez J, Khoshaba B, Scott DL. Fatigue in rheumatoid arthritis reflects pain, not disease activity. *Rheumatology (Oxford)* 2006;45:885–9.
- Hewlett S, Chalder T, Choy E, Cramp F, Davis B, Dures E, et al. Fatigue in rheumatoid arthritis: time for a conceptual model [editorial]. *Rheumatology (Oxford)* 2011;50:1004–6.
- Cohen J. Statistical power analysis for the behavioural sciences. 2nd ed. Hillsdale (NJ): Lawrence Erlbaum Associates; 1988. p. 79–81.
- Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcomes measures for use in clinical trials. *Health Technol Assess* 1998;2:1–74.
- 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 Res* 2010;62:1552–8.
- Nicklin J, Cramp F, Kirwan J, Greenwood R, Urban M, Hewlett S. Measuring fatigue in RA: a cross-sectional study to evaluate the Bristol Rheumatoid Arthritis Fatigue Multi-Dimensional Questionnaire, Visual Analog Scales, and Numerical Rating Scales. *Arthritis Care Res* 2010;62:1559–68.

15. Nicklin J. The development of scales to measure fatigue in people with rheumatoid arthritis. Bristol (UK): University of the West of England; 2009.
16. Dures E, Hewlett S, Kirwan J, Cramp F, Nicklin J, Greenwood R. Test-retest reliability of the Bristol Rheumatoid Arthritis Fatigue Scales (BRAFs) [abstract]. *Rheumatology (Oxford)* 2011;50:Suppl 3:109.
17. Hewlett S, Dures E, Kirwan JR, Cramp F, Nicklin J, Almeida C, et al. Sensitivity to change of the Bristol Rheumatoid Arthritis Fatigue Scales [abstract]. *Arthritis Rheum* 2011;63 Suppl.
18. Downie WW, Leatham PA, Rhind VM, Wright V, Branco JA, Anderson JA. Studies with pain rating scales. *Ann Rheum Dis* 1978;37:378–81.
19. Chalder T, Berelowitz G, Pawlikowska T, Watts L, Wessely S, Wright D, et al. Development of a fatigue scale. *J Psychosom Res* 1993;37:147–53.
20. Cella M, Chalder T. Measuring fatigue in clinical and community settings. *J Psychosom Res* 2010;69:17–22.
21. Naschitz JE, Rozenbaum M, Fields MC, Enis S, Manor H, Dreyfuss D, et al. Cardiovascular reactivity in fibromyalgia: evidence for pathogenic heterogeneity. *J Rheumatol* 2005;32:335–9.
22. Tench CM, McCurdie I, White PD, D'Cruz PD. The prevalence and associations of fatigue in systemic lupus erythematosus. *Rheumatology (Oxford)* 2000;39:1249–54.
23. Strombeck B, Ekdahl C, Manthorpe R, Jacobsson LT. Physical capacity in women with primary Sjogren's syndrome: a controlled study. *Arthritis Rheum* 2003;49:681–8.
24. Tench C, Bentley D, Vleck V, McCurdie I, White P, D'Cruz D. Aerobic fitness, fatigue, and physical disability in systemic lupus erythematosus. *J Rheumatol* 2002;29:474–81.
25. White PD, Henderson M, Pearson RM, Coldrick AR, White AG, Kidd BL. Illness behaviour and psychosocial factors in diffuse upper limb pain disorder: a case-control study. *J Rheumatol* 2003;30:139–45.
26. Tench CM, McCarthy JM, McCurdie I, White PD, D'Cruz PD. Fatigue in systemic lupus erythematosus: a randomized controlled trial of exercise. *Rheumatology (Oxford)* 2003;42:1050–4.
27. Henderson M, Kidd BL, Pearson RM, White PD. Chronic upper limb pain: an exploration of the biopsychosocial model. *J Rheumatol* 2005;32:118–22.
28. Lwin CT, Bishay M, Platts RG, Booth DA, Bowman SJ. The assessment of fatigue in primary Sjogren's syndrome. *Scand J Rheumatol* 2003;32:33–7.
29. Vercoulen J, Swanink C, Fennis J, Galama J, van der Meer J, Bleijenberg G. Dimensional assessment of chronic fatigue syndrome. *J Psychosom Res* 1994;38:383–92.
30. Raterman HG, Hoving JL, Nurmohamed MT, Herenius MM, Sluiter JK, Lems WF, et al. Work ability: a new outcome measure in rheumatoid arthritis? *Scand J Rheumatol* 2010;39:127–31.
31. Van Hoogmoed D, Fransen J, Bleijenberg G, van Riel P. Physical and psychosocial correlates of severe fatigue in rheumatoid arthritis. *Rheumatology (Oxford)* 2010;49:1294–302.
32. De Croon EM, Sluiter JK, Nijssen TF, Kammeijer M, Dijkmans BA, Lankhorst GJ, et al. Work ability of Dutch employees with rheumatoid arthritis. *Scand J Rheumatol* 2005;43:277–83.
33. Hoving JL, Bartelds GM, Sluiter JK, Sadiraj K, Groot I, Lems WF, et al. Perceived work ability, quality of life and fatigue in patients with rheumatoid arthritis after a 6 month course of TNF inhibitors: prospective intervention study and partial economic evaluation. *Scand J Rheumatol* 2009;38:246–50.
34. Evers AW, Kraaijmaat FW, van Riel PL, de Jong AJ. Tailored cognitive-behavioral therapy in early rheumatoid arthritis for patients at risk: an RCT. *Pain* 2002;100:141–53.
35. Zijlstra TR, Taal E, van de Laar MA, Rasker JJ. Validation of a Dutch translation of the fibromyalgia impact questionnaire. *Rheumatology (Oxford)* 2007;46:131–4.
36. Van Koulil S, Kraaijmaat FW, van Lankveld W, van Riel PL, Evers AW. A patient's perspective on multidisciplinary treatment gain for fibromyalgia: an indicator for pre-post treatment effects? *Arthritis Rheum* 2009;61:1626–32.
37. Van Hoogmoed D, Fransen J, Bleijenberg G, van Riel PL. How to assess fatigue in rheumatoid arthritis: validity and reliability of the Checklist Individual Strength [abstract]. *Arthritis Rheum* 2008;58 Suppl:S868.
38. Knoop H, Van der Meer JW, Bleijenberg G. Guided self-instructions for people with chronic fatigue syndrome: randomized controlled trial. *Br J Psychiatry* 2008;193:340–1.
39. Beurskens A, Bultmann U, Kant I, Vercoulen J, Bleijenberg G, Swaen G. Fatigue among working people: validity of a questionnaire measure. *Occup Environ Med* 2000;57:353–7.
40. Krupp LB, LaRocca NG, Muir-Nash J, Steinberg AD. The Fatigue Severity Scale: application to patients with multiple sclerosis and systemic lupus erythematosus. *Arch Neurol* 1989;46:1121–3.
41. Ad Hoc Committee on Systemic Lupus Erythematosus Response Criteria for Fatigue. Measurement of fatigue in systemic lupus erythematosus: a systematic review. *Arthritis Rheum* 2007;57:1348–57.
42. Schentag CT, Beaton M, Rahman P, Husted J, Gladman DD. Prevalence and correlates of fatigue in psoriatic arthritis (PsA) [abstract]. *Arthritis Rheum* 2000;43 Suppl:S105.
43. Mancuso CA, Rincon M, Sayles W, Paget SA. Psychosocial variables and fatigue: a longitudinal study comparing individuals with rheumatoid arthritis and healthy controls. *J Rheumatol* 2006;33:1496–502.
44. Harboe E, Greve OJ, Beyer M, Goransson LG, Tjensvoll AB, Maroni S, et al. Fatigue is associated with cerebral white matter hyperintensities in patients with systemic lupus erythematosus. *J Neurol Neurosurg Psychiatry* 2008;79:199–201.
45. Schentag CT, Cichon J, MacKinnon A, Gladman DD, Urowitz MB. Validation and normative data for the 0–10 point scale version of the fatigue severity scale (FSS) [abstract]. *Arthritis Rheum* 2000;43 Suppl: S177.
46. Mattsson M, Moller B, Lundberg IE, Gard G, Bostrom C. Reliability and validity of the Fatigue Severity Scale in Swedish for patients with systemic lupus erythematosus. *Scand J Rheumatol* 2008;37:269–77.
47. Austin JS, Maisiak RS, Macrina DM, Heck LW. Health outcome improvements in patients with systemic lupus erythematosus using two telephone counseling interventions. *Arthritis Care Res* 1996;9:391–9.
48. Omdal R, Mellgren SI, Koldingsnes W, Jacobsen EA, Husby G. Fatigue in patients with systemic lupus erythematosus: lack of association to serum cytokines, anti-phospholipid antibodies or other disease characteristics. *J Rheumatol* 2002;29:482–6.
49. Chandran V, Behella S, Schentag C, Gladman DD. Functional Assessment of Chronic Illness Therapy-Fatigue Scale is valid in patients with psoriatic arthritis. *Ann Rheum Dis* 2007;66:936–9.
50. Wanders AJ, Gorman JD, Davis JC, Landewe RB, van der Heijde DM. 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 Rheum* 2004;51:1–8.
51. Ramsey-Goldman R, Schilling EM, Dunlop D, Langman C, Greenland P, Thomas RJ, et al. A pilot study on the effects of exercise in patients with systemic lupus erythematosus. *Arthritis Care Res* 2000;13:262–9.
52. Pouchot J, Kherani RB, Brant R, Lacaille D, Lehman AJ, Ensworth S, et al. Determination of the minimal clinically important difference for seven fatigue measures in rheumatoid arthritis. *J Clin Epidemiol* 2008;61:705–13.
53. Goligher EC, Pouchot J, Brant R, Kherani R, Avina-Zubieta JA, Lacaille D, et al. Minimal clinically important difference for seven fatigue measures in patients with systemic lupus erythematosus. *J Rheumatol* 2008;35:635–42.
54. Cross M, Lapsley H, Barcenilla A, Brooks P, March L. Association between measures of fatigue and health-related quality of life in rheumatoid arthritis and osteoarthritis. The Patient: Patient-Centered Outcomes Research 2008;1:97–104.
55. Schwartz JE, Jandorf L, Krupp LB. The measurement of fatigue: a new instrument. *J Psychosom Res* 1993;37:753–62.
56. 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.
57. 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;32:811–9.
58. Walker J, Gordon T, Lester S, Downie-Doyle S, McEvoy D, Pile K, et al. Increased severity of lower urinary tract symptoms and daytime somnolence in primary Sjogren's syndrome. *J Rheumatol* 2003;30:2406–12.
59. Gladman DD, Mease PJ, Cifaldi MA, Perdok RJ, Sasso E, Medich J. Adalimumab improves joint-related and skin-related functional impairment in patients with psoriatic arthritis: patient-reported outcomes of the Adalimumab effectiveness in psoriatic arthritis trial. *Ann Rheum Dis* 2007;66:163–8.
60. Mittendorf T, Dietz B, Sterz R, Kupper H, Cifaldi MA, von der Schulenburg JM. Improvement and long-term maintenance of quality of life during treatment with adalimumab in severe rheumatoid arthritis. *J Rheumatol* 2007;34:2343–50.
61. Cohen SB, Emery P, Greenwald MW, Dougados M, Furie RA, Genovese MC, et al. Rituximab for rheumatoid arthritis refractory to anti-tumor necrosis factor therapy: results of a multicenter, randomized, double-blind, placebo-controlled, phase III trial evaluating primary efficacy and safety at twenty-four weeks. *Arthritis Rheum* 2006;54:2793–806.
62. Weinblatt ME, Keystone EC, Furst DE, Moreland LW, Weisman MH, Birbara CA, et al. Adalimumab, a fully human anti-tumor necrosis factor α monoclonal antibody, for the treatment of rheumatoid arthritis in patients taking concomitant methotrexate. *Arthritis Rheum* 2003;48:35–45.
63. Segal B, Bowman SJ, Fox PC, Vivino FB, Murukutla N, Brodscholl J, et al. Primary Sjogren's syndrome: health experiences and predictors

- of health quality among patients in the United States. *Health Qual Life Outcomes* 2009;7:46.
64. Cella D, Lai JS, Chang CH, Peterman A, Slavin M. Fatigue in cancer patients compared with fatigue in the general United States population. *Cancer* 2002;94:528–38.
 65. Nikolaus S, Bode C, Taal E, van de Laar MA. Four different patterns of fatigue in rheumatoid arthritis patients: results of a Q-sort study. *Rheumatology (Oxford)* 2010;49:2191–9.
 66. Tack B. Dimensions and correlates of fatigue in older adults with rheumatoid arthritis. San Francisco: University of California; 1991.
 67. Piper B, Lindsey A, Dodd M, Ferketich S, Paul S, Weller S. The development of an instrument to measure the subjective dimension of fatigue. In: Funk S, Tornquist E, Champagne M, Wiese R, editors. Key aspects of comfort: management of pain, fatigue, and nausea. New York: Springer; 1989. p. 199–207.
 68. Belza BL. Comparison of self-reported fatigue in rheumatoid arthritis and controls. *J Rheumatol* 1995;22:639–43.
 69. Belza B, Henke C, Yelin E, Epstein W, Gilliss C. Correlates of fatigue in older adults with rheumatoid arthritis. *Nurs Res* 1993;42:93–9.
 70. Wolfe F. Fatigue assessments in rheumatoid arthritis: comparative performance of visual analog scales and longer fatigue questionnaires in 7,760 patients. *J Rheumatol* 2004;31:1896–902.
 71. Stebbings S, Herbison P, Doyle TC, Treharne G, Highton J. A comparison of fatigue correlates in rheumatoid arthritis and osteoarthritis: disparity in associations with disability, anxiety and sleep disturbance. *Rheumatology (Oxford)* 2010;49:361–7.
 72. Turan Y, Duruoz MT, Bal S, Guvenc A, Cerrahoglu L, Gurgan A. Assessment of fatigue in patients with ankylosing spondylitis. *Rheumatol Int* 2007;27:847–52.
 73. 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 Nurs Health* 1997; 20:195–204.
 74. Arnold LM, Zlateva G, Sadosky A, Emir B, Whalen E. Correlations between fibromyalgia symptom and function domains and patient global impression of change: a pooled analysis of three randomized, placebo-controlled trials of Pregabalin. *Pain Med* 2011;12:260–7.
 75. Arnold LM, Chatamra K, Hirsch I, Stoker M. Safety and efficacy of Esreboxetine in patients with fibromyalgia: an 8-week, multicenter, randomized, double-blind, placebo-controlled study. *Clin Ther* 2010; 32:1618–32.
 76. Yacoub YI, Amine B, Laatiris A, Abouqal R, Hajjaj-Hassouni N. Assessment of fatigue in Moroccan patients with ankylosing spondylitis. *Clin Rheumatol* 2010;29:1295–9.
 77. Jump RL, Fifield J, Tennen H, Reisine S, Giuliano AJ. History of affective disorder and the experience of fatigue in rheumatoid arthritis. *Arthritis Rheum* 2004;51:239–45.
 78. Bouzubar FF. Self-reported fatigue in individuals with knee osteoarthritis. Pittsburgh: University of Pittsburgh; 2003.
 79. Meek PM, Nail LM, Barsevick A, Schwartz AL, Stephen S, Whitmer K, et al. Psychometric testing of fatigue instruments for use with cancer patients. *Nurs Res* 2000;49:181–90.
 80. Smets E, Garssen B, Bonke B, De Haes J. The Multidimensional Fatigue Inventory (MFI): psychometric qualities of an instrument to assess fatigue. *J Psychosomatic Res* 1995;39:315–25.
 81. Smets E, Garssen B, Cull A, de Haes J. Application of the multidimensional fatigue inventory (MFI-20) in cancer patients receiving radiotherapy. *Br J Cancer* 1996;73:241–5.
 82. Barendregt P, Visser M, Smets E, Tulen J, van den Meiracker A, Boomsma F, et al. Fatigue in primary Sjogren's syndrome. *Ann Rheum Dis* 1998;57:291–5.
 83. Van Tubergen A, Coenen J, Landewe R, Spoorenberg A, Chorus A, Boonen A, et al. Assessment of fatigue in patients with ankylosing spondylitis: a psychometric analysis. *Arthritis Rheum* 2002;47:8–16.
 84. Forsblad d'Elia H, Rehnberg E, Kvist G, Ericsson A, Kontinen YT, Mannerkorpi K. Fatigue and blood pressure in primary Sjogren's syndrome. *Scand J Rheumatol* 2008;37:284–92.
 85. Rupp I, Boshuizen HC, Jacobi CE, Dinant HJ, van den Bos GA. Impact of fatigue on health-related quality of life in rheumatoid arthritis. *Arthritis Rheum* 2004;51:578–85.
 86. Goodchild CE, Treharne GJ, Booth DA, Kitas GD, Bowman SJ. Measuring fatigue among women with Sjogren's syndrome or rheumatoid arthritis: a comparison of the Profile of Fatigue (ProF) and the Multidimensional Fatigue Inventory (MFI). *Musculoskelet Care* 2008;6: 31–48.
 87. Ericsson A, Mannerkorpi K. Assessment of fatigue in patients with fibromyalgia and chronic widespread pain: reliability and validity of the Swedish version of the MFI-20. *Disabil Rehabil* 2007;29:1665–70.
 88. Clauw DJ, Mease P, Palmer RH, Gendreau RM, Wang Y, Milnacipran for the treatment of fibromyalgia in adults: a 15 week, multicenter, randomized, double-blind, placebo-controlled, multiple-dose clinical trial. *Clin Ther* 2008;30:1988–2004.
 89. Varni JW, Burwinkle TM, Katz ER, Meeske K, Dickinson P. The PedsQL in pediatric cancer: reliability and validity of the Pediatric Quality of Life Inventory Generic Core scales, multidimensional fatigue scale, and cancer module. *Cancer* 2002;94:2090–106.
 90. Varni JW, Seid M, Rode CA. The PedsQL: measurement model for the Pediatric Quality of Life Inventory. *Med Care* 1999;37:126–39.
 91. Varni JW, Burwinkle TM, Szer IS. The PedsQL Multidimensional Fatigue Scale in pediatric rheumatology: reliability and validity. *J Rheumatol* 2004;31:2494–500.
 92. Varni JW, Burwinkle TM, Limbers CA, Szer IS. The PedsQL as a patient-reported outcome in children and adolescents with fibromyalgia: an analysis of OMERACT domains. *Health Qual Life Outcomes* 2007;5:9.
 93. Ringold S, Wallace CA, Rivara FP. Health-related quality of life, physical function, fatigue and disease activity in children with established polyarticular juvenile idiopathic arthritis. *J Rheumatol* 2009; 36:1330–6.
 94. Varni JW, Limbers CA. The PedsQL Multidimensional Fatigue Scale in young adults: feasibility, reliability and validity in a University student population. *Qual Life Res* 2008;17:105–14.
 95. Chiang YC, Hinds PS, Yeh CH, Yang CP. Development and psychometric testing of a Chinese version of the Fatigue Scale-Children in Taiwan. *J Clin Nursing* 2008;17:1201–10.
 96. Bowman SJ, Booth DA, Platts RG, and the UK Sjogren's Interest Group. Measurement of fatigue and discomfort in primary Sjogren's syndrome using a new questionnaire tool. *Rheumatology (Oxford)* 2004; 43:758–64.
 97. Bowman SJ, Hamburger J, Richards A, Barry RJ, Sauz S. Patient-reported outcomes in primary Sjogren's syndrome: comparison of the long and short versions of the Profile of Fatigue and Discomfort-Sicca Symptoms Inventory. *Rheumatology (Oxford)* 2009;48:140–3.
 98. Strombeck B, Theander E, Jacobsson LT. Assessment of fatigue in primary Sjogren's syndrome: the Swedish version of the Profile of Fatigue. *Scand J Rheumatol* 2005;34:455–9.
 99. Barry RJ, Sutcliffe N, Isenberg DA, Price E, Goldblatt F, Adler M, et al. The Sjogren's Syndrome Damage Index: a damage index for use in clinical trials and observational studies in primary Sjogren's syndrome. *Rheumatology (Oxford)* 2008;47:1193–8.
 100. Dass S, Bowman SJ, Vital EM, Ikeda K, Pease CT, Hamburger J, et al. Reduction of fatigue in Sjogren's syndrome with rituximab: results of a randomised, double-blind, placebo-controlled pilot study. *Ann Rheum Dis* 2008;67:1541–4.
 101. Goodchild EC, Treharne GJ, Booth DA, Bowman SJ. Daytime patterning of fatigue and its associations with the previous night's discomfort and poor sleep among women with primary Sjogren's syndrome or rheumatoid arthritis. *Musculoskelet Care* 2010;8:107–17.
 102. Segal B, Thomas W, Rogers T, Leon JM, Hughes P, Patel D, et al. Prevalence, severity, and predictors of fatigue in subjects with primary Sjogren's syndrome. *Arthritis Rheum* 2008;59:1780–7.
 103. Ware JE, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36): conceptual framework and item selection. *Med Care* 1992;30: 473–83.
 104. Keystone E, Burmester GR, Furie R, Loveless JE, Emery P, Kremer J, et al. Improvement in patient-reported outcomes in a Rituximab trial in patients with severe rheumatoid arthritis refractory to anti-tumor necrosis factor therapy. *Arthritis Rheum* 2008;59:785–93.
 105. 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;28:266–74.
 106. Wolfe F, Michaud K, Li T, Katz RS. EQ-5D and SF-36 quality of life measures in systemic lupus erythematosus: comparisons with rheumatoid arthritis, non-inflammatory rheumatic disorders and fibromyalgia. *J Rheumatol* 2010;37:296–304.
 107. Soderman P, Malchau H. Validity and reliability of the Swedish WOMAC osteoarthritis index: a self-administered disease-specific questionnaire (WOMAC) versus generic instruments (SF-36 and NHP). *Acta Orthop Scand* 2000;71:39–46.
 108. Tuttleman M, Pillemer SR, Tilley BC, Fowler SE, Budkley LM, Alarcon GM, et al. A cross sectional assessment of health status instruments in patients with rheumatoid arthritis participating in a clinical trial. *J Rheumatol* 1997;24:1910–5.
 109. 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 Rheum* 2005;53:5–11.
 110. Busija L, Osborne RH, Nilsdotter A, Buchbinder R, Roos EM. Magnitude and meaningfulness of change in SF-36 scores in four types of orthopaedic surgery. *Health Qual Life Outcomes* 2008;6:55.
 111. Davey RC, Matthes Edwards S, Cochrane T. Test-retest reliability of lower extremity functional and self-reported measures in elderly with osteoarthritis. *Adv Physiother* 2003;5:155–60.
 112. Birrell FN, Hassell AB, Jones PW, Dawes PT. How does the Short Form 36 Health Questionnaire (SF-36) in rheumatoid arthritis (RA)

- relate to RA outcome measures and SF-36 population values? A cross-sectional study. *Clin Rheumatol* 2000;19:195–9.
113. Loge JH, Kaasa 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;51:1069–76.
 114. Linde L, Sorensen J, Ostergaard M, Horslev-Peterson, Hetland ML. Health-related quality of life: validity, reliability and responsiveness of SF-36, EQ-15D, EQ-5D, RAQoL and HAQ in patients with rheumatoid arthritis. *J Rheumatol* 2008;35:1528–37.
 115. Koh ET, Leong KP, Tsou IY, Lim VH, Pong LY, Chong SY, et al. The reliability, validity and sensitivity to change of the Chinese version of SF-36 in oriental patients with rheumatoid arthritis. *Rheumatology (Oxford)* 2006;45:1023–8.
 116. Saad AA, Ashcroft DM, Watson KD, Symmons DP, Noyce PR, Hyrich KL. Improvements in quality of life and functional status in patients with psoriatic arthritis receiving anti-tumor necrosis factor therapies. *Arthritis Care Res* 2010;62:345–53.
 117. Wells G, Li T, Maxwell L, MacLean R, Tugwell P. Responsiveness of patient reported outcomes including fatigue, sleep quality, activity limitation, and quality of life following treatment with abatacept for rheumatoid arthritis. *Ann Rheum Dis* 2008;67:260–5.
 118. Riemsma RP, Rasker JJ, Taal E, Griep EN, Wouters JM, Wiegman O. Fatigue in rheumatoid arthritis: the role of self-efficacy and problematic social support. *Br J Rheumatol* 1998;37:1042–6.
 119. Mengshoel AM, Forre O. Pain and fatigue in patients with rheumatic disorders. *Clin Rheumatol* 1993;12:515–22.
 120. Dernis-Labous E, Messow M, Dougados M. Assessment of fatigue in the management of patients with ankylosing spondylitis. *Rheumatology (Oxford)* 2003;42:1523–8.
 121. Rohekar G, Pope J. Test-retest reliability of patient global assessment and physician global assessment in rheumatoid arthritis. *J Rheumatol* 2009;36:2178–82.
 122. Khanna D, Pope J, Khanna PP, Maloney M, Samed N, Norrie D, et al. The minimally important difference for the fatigue visual analog scale in patients with rheumatoid arthritis followed in an academic clinical practice. *J Rheumatol* 2008;35:2339–43.
 123. Wells G, Li T, Maxwell L, MacLean R, Tugwell P. Determining the minimal clinically important differences in activity, fatigue, and sleep quality in patients with rheumatoid arthritis. *J Rheumatol* 2007;34:280–9.
 124. Slatkowsky-Christensen B, Mowinckel P, Loge JH, Kvien TK. Health-related quality of life in women with symptomatic hand osteoarthritis: a comparison with rheumatoid arthritis patients, healthy controls, and normative data. *Arthritis Rheum* 2007;57:1404–9.
 125. Wolfe F, Hawley D, Wilson K. The prevalence and meaning of fatigue in rheumatic disease. *J Rheumatol* 1996;23:1407–17.
 126. Wolfe F. Determinants of WOMAC function, pain and stiffness scores: evidence for the role of low back pain, symptoms counts, fatigue and depression in osteoarthritis, rheumatoid arthritis and fibromyalgia. *Rheumatology (Oxford)* 1999;38:355–61.
 127. Colangelo KJ, Pope JE, Peschken C. The minimally important difference for patient reported outcomes in systemic lupus erythematosus including the HAQ-DI, pain, fatigue, and SF-36. *J Rheumatol* 2009;36:2231–7.
 128. Kwok T, Pope JE. Minimally important difference for patient-reported outcomes in psoriatic arthritis: health assessment questionnaire and pain, fatigue, and global visual analog scales. *J Rheumatol* 2010;37:1024–8.
 129. Braun J, van der Heijde D, Doyle MK, Han C, Deodhar A, Inman R, et al. Improvement in hemoglobin levels in patients with ankylosing spondylitis treated with Infliximab. *Arthritis Rheum* 2009;61:1032–6.
 130. Ide MR, Laurindo IM, Rodrigues-Junior AL, Tanaka C. Effect of aquatic respiratory exercise-based program in patients with fibromyalgia. *Int J Rheum Dis* 2008;11:131–40.
 131. Heiberg MS, Kaufmann C, Rodevand E, Mikkelsen K, Koldingsnes W, Mowinckel P, et al. The comparative effectiveness of anti-TNF therapy and methotrexate in patients with psoriatic arthritis: 6 month results from a longitudinal, observational, multicentre study. *Ann Rheum Dis* 2007;66:1038–42.
 132. Heiberg T, Finset A, Uhlig T, Kvien TK. Seven year changes in health status and priorities for improvement of health in patients with rheumatoid arthritis. *Ann Rheum Dis* 2005;64:191–5.
 133. Bellamy N. *Musculoskeletal clinical metrology*. London: Kluwer Academic Publishers; 1993.
 134. Brazier J, Deverill M, Green C, Harper R, Booth A. A review of the use of health status measures in economic evaluation. *Health Technol Assess* 1999;3:9.
 135. Streiner DL, Norman GR. *Health measurement scales: A practical guide to their development and use*. 4th ed. Oxford: Oxford University Press; 2008.
 136. Huskisson EC. *Visual analogue scales*. In: Melzack R, editor. *Pain measurement and assessment*. New York: Raven Press; 1983.

Summary Table for Fatigue Measures

Scale	Purpose/content	Method of administration	Respondent burden	Administrative burden	Score interpretation	Reliability evidence	Validity evidence	Ability to detect change	Strengths	Cautions
Bristol Rheumatoid Arthritis Fatigue Multi-Dimensional Questionnaire	Measure severity, impact, and dimensions of fatigue in rheumatoid arthritis	Patient self-report	4–5 minutes	2–3 minutes	Higher = worse	Internal consistency: strong; test-retest: strong	Content validity: strong; construct validity: strong; criterion validity: strong	Good, based on data still under peer review	Developed with rheumatoid arthritis patients; contains multiple subscales	Currently, data only available from developmental article; reliability and sensitivity article awaited
Bristol Rheumatoid Arthritis Fatigue Numerical Rating Scales (severity, effect, and coping)	Measure severity, impact, and coping with fatigue in rheumatoid arthritis	Patient self-report	1 minute	1 minute	Severity and effect: higher = worse; coping higher = better	Severity and effect test-retest: strong; coping test-retest: moderate	Content validity: strong; construct validity: strong; criterion validity: strong; moderate for coping	Good, based on data still under peer review	Developed with rheumatoid arthritis patients; measures severity, effect, and coping	Currently, data only available from developmental article; reliability and sensitivity article awaited
Chandler Fatigue Questionnaire	Measure severity in hospital and community populations	Patient self-report	2–3 minutes	2–3 minutes	Higher = worse	Internal consistency: strong; test-retest: strong in other populations	Content validity: good; construct validity: moderate; criterion validity: moderate	Good	Measures different subscales	No reliability data for rheumatology; does not always differentiate between rheumatology patients and controls
Checklist Individual Strength	Measure aspects of fatigue in chronic fatigue syndrome	Patient self-report	4–5 minutes	4–5 minutes	Higher = worse	Internal consistency: strong; test-retest: strong	Content validity: moderate; construct validity: strong; criterion validity: strong	Good	Evaluated in many long-term conditions; contains multiple subscales	3 items might be confounded by rheumatology disease or disability
Fatigue Severity Scale	Measure disabling fatigue in multiple sclerosis and systemic lupus erythematosus	Patient self-report	2–3 minutes	2–3 minutes	Higher = worse	Internal consistency: strong; test-retest: strong	Content validity: strong; construct validity: strong; criterion validity: strong	Good	Recommended fatigue scale for systemic lupus erythematosus	A 10-point version is being used in psoriatic arthritis
Functional Assessment Chronic Illness Therapy (Fatigue)	Measure fatigue in anemic and chronic illness tested in chronic illness	Patient self-report	3–4 minutes	3–4 minutes	Higher = better	Internal consistency: strong; test-retest: strong	Content validity: moderate; construct validity: strong; criterion validity: strong	Good	Evaluated in several rheumatologic conditions	Phrasing of 4 of 13 items potentially confounded by rheumatologic conditions
Multi-Dimensional Assessment of Fatigue	Measure multiple dimensions of fatigue in chronic fatigue syndrome and rheumatoid arthritis	Patient self-report	5–8 minutes	4–5 minutes	Higher = worse	Internal consistency: strong; test-retest: strong	Content validity: moderate; construct validity: strong; criterion validity: strong	Good	Rheumatoid arthritis specific but also assessed a range of long-term conditions	High levels of missing data reported; items might be assessed in relation to disability
Multi-Dimensional Fatigue Inventory	Measure cancer fatigue using a multidimensional, short questionnaire without any somatic items	Patient self-report	4–5 minutes	4–5 minutes	Higher = worse	Internal consistency: strong; test-retest: strong	Content validity: moderate; construct validity: strong; criterion validity: moderate and variable	Good	Contains multiple subscales, evaluated in long-term and short-term conditions	Mental fatigue scale does not correlate strongly with other subscales; potential for confounding by fatigue; criterion validity variable across subscales; phrasing may be confounded by rheumatologic conditions
Pediatric Quality of Life Multi-Dimensional Fatigue Scale	Measure child and parent perceptions of fatigue in pediatric patients; developed in cancer, intended as generic	Patient self-report	4–5 minutes	4–5 minutes	Higher = better	Internal consistency: strong; test-retest: strong	Content validity: moderate; construct validity: strong to moderate; criterion validity: moderate in healthy children	No data in any disease could be located	Measures multiple dimensions of a well-established measurement system	No rheumatology criterion validity; sensitivity data, correlates very strongly with psychological health in rheumatology
Profile of Fatigue	Characterize patterns of fatigue in primary Sjögren's syndrome	Patient self-report	4–5 minutes	4–5 minutes	Higher = worse	Internal consistency: strong; test-retest: moderate	Content validity: strong; construct validity: moderate; criterion validity: strong	Good	Primary Sjögren's developed with multiple subscales	Test-retest reliability data moderate; studies have small numbers to test 6 facets
Short Form 36 vitality subscale	Measure vitality (energy and fatigue) in general and clinical populations	Patient self-report	1 minute	1–2 minutes	Higher = better	Internal consistency: strong; test-retest: very weak to strong	Content validity: moderate; construct validity: strong; criterion validity: variable, moderate to strong	Good	Widely used, can compare across studies; developed with general population	Concerns over concepts of fatigue vs. energy; criterion validity not tested in rheumatology populations
Visual analog scales	Measure various fatigue constructs are required	Patient self-report	1 minute	1 minute	Higher = worse	Test-retest: strong	Content validity: no standard for construct validity; strong criterion validity; variable, moderate to strong	Good	Widely used, a quick screening patient-reported outcome	Standardized version developed but was not used in studies; rating scales, visual analog scale can be confusing for some patients; photocopying distorts the line