RHEUMATOLOGY

doi:10.1093/rheumatology/ket218
Advance Access publication 9 July 2013

Original article

Reliability and sensitivity to change of the Bristol Rheumatoid Arthritis Fatigue scales

Emma K. Dures[1], Sarah E. Hewlett[1], Fiona A. Cramp[1], Rosemary Greenwood[2], Joanna K. Nicklin[1], Marie Urban[2] and John R. Kirwan[3]

Abstract

Objective. To examine the reliability (stability) and sensitivity of the Bristol Rheumatoid Arthritis Fatigue scales (BRAFs) and patient-reported outcome measures (PROMs) developed to capture the fatigue experience. The Multi-Dimensional Questionnaire (BRAF-MDQ) has a global score and four subscales (Physical Fatigue, Living with Fatigue, Cognitive Fatigue and Emotional Fatigue), while three numerical rating scales (BRAF-NRS) measure fatigue Severity, Effect and Coping.

Methods. RA patients completed the BRAFs plus comparator PROMs. Reliability (study 1): 50 patients completed questionnaires twice. A same-day test–retest interval (minimum 60 min) ensured both time points related to the same 7 days, minimizing the capture of fatigue fluctuations. Reliability (study 2): 50 patients completed the same procedure with a re-worded BRAF-NRS Coping. Sensitivity to change (study 3): 42 patients being given clinically a single high dose of i.m. glucocorticoids completed questionnaires at weeks 0 and 2.

Results. The BRAF-MDQ, its subscales and the BRAF-NRS showed very strong reliability ($r = 0.82$–0.95). BRAF-NRS Coping had lower moderate reliability in both wording formats ($r = 0.62$, 0.60). The BRAF-MDQ, its subscales and the BRAF-NRS Severity and Effect were sensitive to change, with effect sizes (ESs) of 0.33–0.56. As hypothesized, the BRF-NRS Coping was not responsive to the pharmaceutical intervention (ES 0.05). Preliminary exploration suggests a minimum clinically important difference of 17.5% for improvement and 6.1% for fatigue worsening.

Conclusion. The BRAF scales show good reliability and sensitivity to change. The lack of BRAF-NRS Coping responsiveness to medication supports the theory that coping with fatigue is a concept distinct from severity and effect that is worth measuring separately.

Key words: rheumatoid arthritis, fatigue, patient-reported outcome, reliability, sensitivity, MCID.

CLINICAL SCIENCE

Introduction

Fatigue in RA is probably caused by a dynamic interaction of clinical factors (inflammation, pain, disability) and psychosocial issues (coping, mood, illness beliefs), which will vary between and within individuals and over time [1–4]. Fatigue affects up to 70% of patients with RA and is experienced as overwhelming, unpredictable and challenging for them to manage [5–8]. The acknowledgement of fatigue as a symptom that is important to patients has resulted in international agreement that fatigue should be measured in all RA trials alongside the core set [9]. Consequently, when a review of existing instruments found that none adequately captured RA fatigue from the patient perspective [10], the Bristol Rheumatoid Arthritis Fatigue scales (BRAFs) were developed in collaboration with patients [11, 12]. The Multi-Dimensional Questionnaire (BRAF-MDQ) is a 20-item questionnaire giving an overall global score and four subscale scores reflecting physical fatigue, living with fatigue, cognitive fatigue and emotional fatigue. In addition, three single-item scales measure fatigue severity, effect and coping using numerical rating scales (BRAF-NRS). Face and content validity of the BRAFs was established in a series of

[1]Faculty of Health and Life Sciences, University of the West of England, Bristol, [2]Research Design Service, University Hospitals Bristol and [3]Faculty of Medicine and Dentistry, University of Bristol, Bristol, UK.

Submitted 6 December 2012; revised version accepted 26 April 2013.

Correspondence to: Emma Dures, Academic Rheumatology, Bristol Royal Infirmary, Bristol BS2 8HW, UK.
E-mail: emma2.dures@uwe.ac.uk

studies including qualitative interviews to identify important concepts, focus groups to draft items and cognitive interviewing to examine the language and understanding of the proposed questions [11]. Factor analysis in a cohort of 229 patients identified 20 items out of 45 draft items that provided four internally consistent subscales, along with construct and criterion validity [12]. The BRAFs have been translated into 34 languages [13]. The recommended methodology of iterative backward and forward translations, independent review and harmonization meetings with the developers was used, and the translated BRAFs were subsequently pilot tested with patients using cognitive interviewing to ensure the comprehension of the individual questions [14–16].

It is essential that patient-reported outcomes (PROMs) are both reliable (stable) and sensitive (responsive). Reliability is when an instrument yields similar results on repeated applications when the concept being measured has not changed [17–20]. There is no consensus about the length of time between the two completions, but it needs to be sufficiently long that participants cannot simply recall their previous answers, yet short enough to minimize the possibility that the concept being measured has meaningfully changed. Thus the key factor is the stability of the construct being measured [17–20]. In the case of RA fatigue, patients report that it can occur without warning, to the extent that they suddenly have to stop their activities and sit or lie down [6–8]. Such unpredictability restricts the concept of stability in RA fatigue to a narrow time frame.

Sensitivity to change is when a PROM is responsive or able to detect meaningful change over time in the concept being measured, for example after an effective intervention [17–20]. While there are currently no pharmacological interventions offered purely for RA fatigue, for some patients fatigue is a feature of inflammatory flares, and therefore the symptom might respond to medications designed to reduce inflammation (e.g. i.m. glucocorticoids), providing a useful scenario in which to test the sensitivity of a fatigue scale. However, as this does not attempt to alter coping with fatigue, we hypothesized that the BRAF-NRS Coping would not be sensitive to such an intervention. While a PROM might reflect change after an intervention, this does not necessarily indicate the change is important to the patient. A minimum clinically important difference (MCID) is the smallest amount of change in a particular PROM that reflects a meaningful change for the patient [21, 22]. MCID can be calculated by comparing change in the PROM with change determined by either a transition question that asks patients whether they consider their symptom to be better, the same or worse [22] or with change in a related (surrogate) concept such as pain [21]. Previous research using surrogate anchors in multiple studies, followed by consensus techniques to resolve uncertainty, suggest an MCID of −1 for improvement on a visual analogue scale (VAS) of fatigue of 0–10 [21]. Another study, this time using a 5-point fatigue transition question (much worsened to much better) as an anchor, gave an MCID for fatigue of −0.82 to −1.12 for improved fatigue and a larger +1.13 to +1.26 for worsened fatigue on a 0–10 VAS [22]. The aim of these three studies was therefore to examine the reliability and sensitivity to change of the BRAF scales and provide preliminary data on an MCID.

Patients and methods

Patients

Consecutive patients were recruited from the rheumatology outpatient department at a single teaching hospital in the Southwest of England. Inclusion criteria for all three studies were a confirmed diagnosis of RA [23], with no other co-morbidities in which fatigue is a common feature (e.g. SLE, multiple sclerosis or cancer). In addition, patients recruited for study 3 (sensitivity to change) had to have been prescribed a single high dose of i.m. glucocorticoids for clinical reasons during their appointment. Ethics approval for the studies was granted by the North Somerset and South Bristol Research Ethics Committee (06/Q2006/104) and written consent was obtained.

Measures

Demographic data were collected on age, sex, disease duration and disability (HAQ) [24] at baseline. The questionnaire packs used in all three studies comprised the BRAF and comparator fatigue scales. All BRAF scales ask about fatigue over the last 7 days. The BRAF-MDQ has 20 items: item 1 (severity) response is 0–10, item 2 (number of days) is 0–7 and item 3 (episode length) is <1 h, several hours or all day, scored 0–2, while items 4–20 have response options of not at all, a little, quite a bit and very much (scored 0–3). All 20-item scores are totalled for the global fatigue score (range 0–70), while four subscales are created for physical fatigue (items 1–4, range 0–22), living with fatigue (items 5–11, range 0–21), cognitive fatigue (items 12–16, range 0–15) and emotional fatigue (items 17–20, range 0–12), with higher scores representing worse fatigue [12]. Missing data were handled according to the BRAF scoring instructions [13]. The BRAF-NRS Severity asks the average level of fatigue (anchors: no fatigue, totally exhausted, 0–10), BRAF-NRS Effect asks about the effect fatigue has had on your life (anchors: no effect, a great deal of effect, 0–10) and BRAF-NRS Coping asks how well you have coped with fatigue (anchors: not at all well, very well, 0–10), thus the reversed BRAF-NRS Coping anchors mean high scores reflect better coping. Following analysis of reliability study 1, the coping anchors were further clarified to read not coping at all well and coping very well, and the study repeated with a fresh cohort of participants (reliability study 2).

The four comparator fatigue PROMs for all three studies were the Functional Assessment of Chronic Illness Therapy Fatigue Scale (FACIT-F) [25], the Multi-dimensional Assessment of Fatigue (MAF) [26], the Profile of Mood States (POMS) [27] and the Short Form-36 Health Survey Vitality subscale (SF-36 VT) [28]. Pain and patient global opinion of disease were measured using a 10-cm

VAS [29, 30]. Two versions of the questionnaire packs (A and B) contained the PROMs in different orders, to try and minimize recall and order bias. In all studies, alternate patients were given version A (or B) at baseline and the opposite version at their second time point.

Procedure

Reliability study 1

The BRAFs measure fatigue over the last 7 days, and as RA fatigue has an inherent variability [6–8], a long period between assessments would risk capturing unpredictable fatigue fluctuations or sudden onset of a new episode, potentially altering the patient's judgement of their fatigue over the preceding week. Such data would reflect the nature of the construct rather than the reliability of the PROM. Patients therefore completed the first questionnaire pack [time 1 (T1)] on arrival at the clinic and the second questionnaire pack [time 2 (T2)] the same day, after a minimum of 60 min, ensuring that both questionnaire completions related to the same 7 days.

Reliability study 2

The BRAF-NRS Coping was reworded and the same procedure repeated with a new cohort of patients.

Sensitivity study 3

A single high dose of i.m. glucocorticoids generally produces a rapid improvement in clinical status, such as a reduction in swollen joints, pain, disability and fatigue within a few days [31]. Patients who were prescribed a single high dose of i.m. glucocorticoids as part of clinical care during their appointment completed the first questionnaire pack (T1) in the clinic while waiting to have their injection. The second pack (T2) was posted to the patients' homes 2 weeks later and returned in a prepaid envelope. At T2, the questionnaire pack contained a transition question (Has your fatigue changed? Better/Same/Worse), in order to provide pilot data for calculating the MCID of the BRAFs.

Sample size

For reliability (studies 1 and 2), a comparison study estimate of 50 complete data sets is needed to produce correlation coefficients of $\geqslant 0.4$ as statistically significant with a power of 82%. For sensitivity (study 3), 40 patients are required to detect a fatigue effect size of 0.46 (shown in a previous study of i.m. glucocorticoids) [31] with 80% power (5% significance). Sample size for a Bland and Altman plot and limits of agreement cannot be formally calculated, but we estimated that 50 sets of data would produce a plot that would be reasonably consistent and easy to interpret.

Analysis

Reliability was assessed using Pearson's correlation coefficient to measure the strength of the association between the two sets of scores. Previous guidelines on interpreting Pearson's correlation coefficients for the BRAFs were used: considered weak at $r = 0.3$–0.49, moderate at $r = 0.5$–0.74 and strong at $\geqslant 0.75$ [32]. Correlation coefficients provide information on how the two sets of scores vary together but does not assess measurement error, therefore Bland and Altman plots were also used. This method assumes the true score is likely to be represented by the mean of the two scores, therefore each patient's difference between their two scores is plotted against this mean, with 2 S.D. of the mean considered the limits of agreement [33].

Sensitivity to change was examined using effect sizes (mean change divided by baseline S.D.), with an effect size of 0.2 considered to be small, 0.5 medium and $\geqslant 0.8$ large [32]. This was compared with the effect sizes seen in other fatigue PROMs. MCID was provisionally explored by using the fatigue transition question as an anchor of worse, same or better fatigue, and calculating the mean BRAF change for each category (95% CI).

Results

Reliability study 1

Ninety consecutive patients were invited to take part in the study: 57 accepted and 33 declined. Of those 57, 7 were withdrawn from the study due to non-completion of the questionnaires at one time point (defined as not returning a completed questionnaire pack to the research team). Fifty patients participated, with a range of disease duration, moderate pain and disability (Table 1) and moderate fatigue (Table 2). There was a very strong correlation between the T1 and T2 scores for the BRAFs: BRAF-MDQ global ($r = 0.95$), each of the four BRAF-MDQ subscales ($r = 0.89$–0.94), BRAF-NRS Severity ($r = 0.92$) and BRAF-NRS Effect ($r = 0.85$) (Table 2). The BRAF-NRS Coping results showed a lower but still moderate agreement ($r = 0.62$).

Bland and Altman limits were relatively narrow, indicating good levels of agreement. The BRAF-MDQ global score agreement was within 11 points (scale range 0–70, observed values 4–70) (Fig. 1), the BRAF-MDQ subscales were all within 5 points (Living with Fatigue, Cognitive Fatigue and Emotional Fatigue scale ranges and observed values 0–12, 0–15 and 0–21, respectively; Physical Fatigue scale range 0–22, observed values 3–22) (Fig. 1). The BRAF-NRS Severity and Effect scores were both within 3 points (scale ranges 0–10; observed values 0–10 for Effect and 1–10 for Severity) (Table 3). The BRAF-NRS Coping performed less strongly, with limits of agreement within 5 points of the 0–10 scale (observed values 0–10), therefore the wording of the BRAF-NRS Coping was clarified by adding cope to the anchors, and the study was repeated with a fresh cohort of participants (study 2).

Reliability study 2

Eighty-three consecutive patients were invited to take part in the study: 59 accepted and 24 declined. Of those 59, 9 were withdrawn from the study due to non-completion of the questionnaires at one time point. Fifty patients participated, with similar characteristics to those in study 1 (Tables 1 and 2). Study 1 findings were replicated, with

Table 1 Demographic and clinical data, studies 1, 2 and 3 at baseline

	Reliability study 1 ($n = 50$)		Reliability study 2 ($n = 50$)		Sensitivity study 3 ($n = 42$)	
	Mean (S.D.)	Range	Mean (S.D.)	Range	Mean (S.D.)	Range
Women, *n* (%)	36 (72)		34 (68)		34 (81)	
Age, years	56.3 (12.8)	25–76	55.8 (13.6)	21–78	56.7 (12.6)	30–76
Disease duration, years	13.4 (12.7)	0.5–60	8.9 (11.3)	0.2–50	10.7 (10.7)	0.05–43
Disability (HAQ 0–3)	1.6 (0.8)	0–3	1.4 (0.9)	0–3	1.9 (0.7)	0.875–3
Pain (VAS 0–100)	54.9 (26.1)	8–98	51.5 (30.1)	1–96	70.8 (19.9)	22–96
Patient global (VAS 0–100)	46.7 (26.1)	7–96	46.6 (27.1)	2–95	58.2 (24.9)	3–95

Table 2 Correlation, reliability studies 1 and 2

	Reliability study 1 ($n = 50$)			Reliability study 2 ($n = 50$)		
	T1 mean (S.D.)	T2 mean (S.D.)	Corr *r*	T1 mean (S.D.)	T2 mean (S.D.)	Corr *r*
BRAF-MDQ						
Global (0–70)	37.2 (16.8)	37.5 (17.3)	0.95	36.0 (17.8)	35.4 (17.4)	0.93
Physical (0–22)	14.8 (5.2)	15.1 (5.2)	0.94	14.3 (5.4)	14.5 (5.3)	0.93
Living (0–21)	9.5 (6.0)	9.7 (6.0)	0.89	9.6 (6.5)	9.4 (6.0)	0.89
Cognition (0–15)	6.8 (4.8)	6.9 (4.3)	0.89	6.3 (4.5)	6.0 (4.6)	0.89
Emotion (0–12)	6.1 (3.8)	5.9 (3.8)	0.92	5.3 (3.9)	5.5 (3.7)	0.84
BRAF-NRS						
Severity (0–10)	6.2 (2.3)	6.3 (2.3)	0.92	5.9 (2.3)	6.0 (2.4)	0.92
Effect (0–10)	6.0 (2.6)	6.1 (2.6)	0.85	5.9 (2.8)	5.9 (2.7)	0.88
Coping (0–10)[a]	5.7 (2.9)	5.2 (2.9)	0.62	6.3 (2.5)	5.7 (2.5)	0.60
SF-36 VT (0–100)[a]	33.4 (21.2)	37.4 (19.6)	0.61	35.9 (20.1)	36.8 (21.5)	0.76
FACIT-F (0–52)[a]	24.1 (11.5)	23.6 (11.4)	0.95	23 (11.1)	25.4 (10.2)	0.88
POMS (0–28)	16.0 (7.4)	15.9 (7.4)	0.90	15.6 (7.9)	15.6 (7.3)	0.93
MAF (1–50)	31.6 (10.1)	31.3 (9.6)	0.91	32.0 (8.1)	31.6 (9.2)	0.88

[a]Higher score indicates better outcome.

very strong correlations for the T1 and T2 BRAF-MDQ global and subscales, and BRAF-NRS Severity and Effect ($r = 0.84$–0.93) (Table 2). BRAF-NRS Coping reliability was unchanged ($r = 0.60$). Bland and Altman limits of agreement were also replicated, with BRAF-MDQ global scores within 12 points, BRAF-MDQ subscales within 6 points and BRAF-NRS Severity and Effect within 3 points. The limits of agreement for BRAF-NRS Coping were again within 5 points.

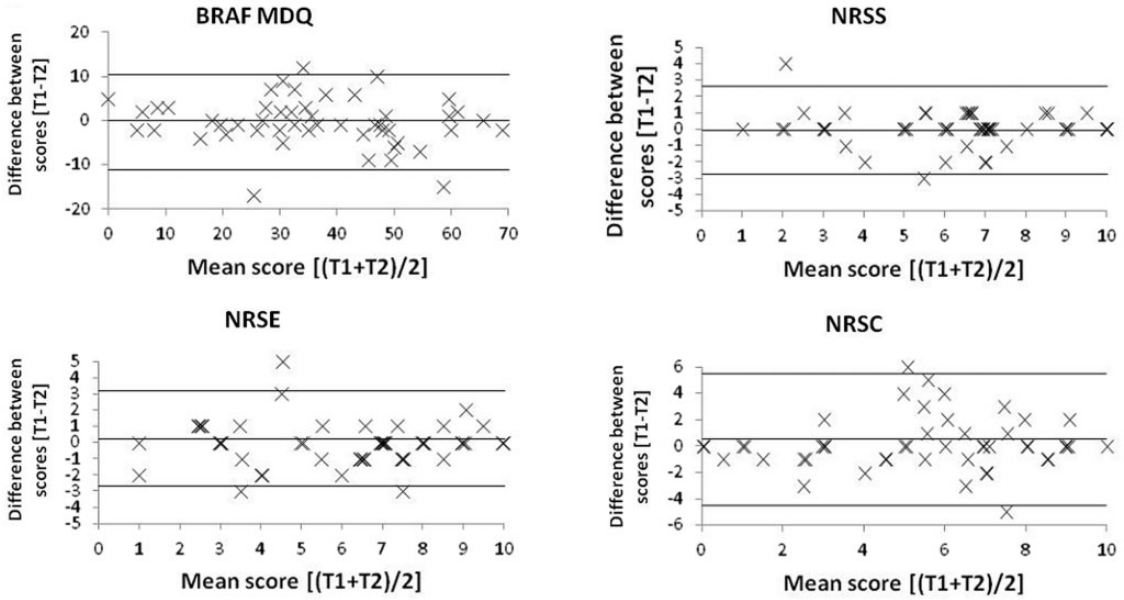
Sensitivity study 3

Seventy patients were invited to take part in the study: 64 accepted and 6 declined. Of those 64, 22 were withdrawn from the study: 16 participants due to non-completion of the questionnaires at one time point and 6 participants due to not meeting the inclusion criteria of a confirmed diagnosis of RA [23]. Forty-two patients participated, with moderate disease duration and disability but high fatigue and pain levels (Tables 1 and 3). The BRAF-MDQ global scale was sensitive to change (ES 0.56), as were the four subscales (ES 0.33–0.54) and the BRAF-NRS Severity and Effect (ES 0.46, 0.47) (Table 4). As hypothesized, the BRAF-NRS Coping was not sensitive to change in this pharmacological intervention (ES 0.05).

Using the transition question, 'Has your fatigue changed?', as an anchor, the mean (S.D.) change in the BRAF-MDQ for the 14 patients who reported no change in their fatigue was −2.42 (8.29), with a 95% CI of −7.43 to 2.58. This means that any patient in whom the BRAF-MDQ score improved by more than 7.43 will have reported an improvement in their fatigue in response to therapy. This represents 17.5% of the pre-treatment value and provides an initial estimate of MCID for improvement. Any patient whose BRAF-MDQ score deteriorated by more than 2.58 will have reported a worsening of their fatigue in response to therapy. This represents 6.1% of the pre-treatment value and provides an initial estimate of MCID for worsening.

The four comparator fatigue scales had similar reliability and sensitivity to the BRAFs (Tables 2 and 3). However, missing data meant that 12% of MAF and 10% of POMS questionnaires could not be scored, compared with 3% of

FIG. 1 Limits of agreement for BRAF-MDQ and -NRS Severity, Effect and Coping.



Mean difference (central line) and limits of agreement (outer lines, at 2 S.D.) for the BRAF-MDQ and -NRS for Severity (S), Effect (E) and Coping (C) measured at times T1 and T2. Coincident points slightly separated for clarity.

TABLE 3 Bland and Altman levels of agreement

	Reliability study 1 ($n = 50$)		Reliability study 2 ($n = 50$)	
	Mean difference	95% levels of agreement	Mean difference	95% levels of agreement
BRAF-MDQ				
Global (0–70)	−0.36	−11.1, 10.4	−0.13	−12.6, 12.3
Physical (0–22)	−0.34	−3.8, 3.1	−0.40	−4.1, 3.3
Living (0–21)	−0.18	−5.6, 5.3	0.2	−5.5, 5.9
Cognition (0–15)	−0.08	−4.3, 4.1	0.26	−3.9, 4.4
Emotion (0–12)	−0.04	−1.8, 1.7	−0.16	−4.5, 4.1
BRAF-NRS				
Severity (0–10)	−0.08	−2.8, 2.6	−0.14	−1.9, 1.6
Effect (0–10)	0.24	−2.7, 3.2	−0.12	−2.8, 2.5
Coping (0–10)[a]	0.5	−4.5, 5.5	0.45	−3.8, 4.7
SF-36 VT (0–100)[a]	2.84	−25, 30.7	−0.41	−29, 28.2
FACIT-F (0–52)[a]	0.17	−7.2, 7.5	−2.21	−12.2, 7.7
POMS (0–28)	−0.03	−6.6, 6.6	0.47	−4.8, 5.7
MAF (1–50)	0.45	−7.8, 8.7	0.82	−7.2, 8.8

[a]Higher scores indicate better outcome.

FACIT, 1% of the SF-36 VT and 0.5% of the BRAFs (based on the percentage of questionnaires that could not be scored due to missing data in the test–retest reliability studies, including time points 1 and 2).

Discussion

The BRAFs were rigorously developed with input from patients. The process began with semi-structured interviews with patients in which they identified their experiences of fatigue and the language they used to describe them [6]. These qualitative data were subsequently discussed with patients in focus groups, ensuring that the concepts of fatigue captured the full experience. This included different fatigue dimensions such as cognitive fatigue and physical fatigue and also perceived fatigue coping and effect. Once the BRAFs had been drafted, patients' understanding and interpretation of

TABLE 4 Mean change and effect sizes, sensitivity study 3 ($n = 42$)

	T1, mean (S.D.)	T2, mean (S.D.)	Mean change	95% CI	*P* value (paired *t*-test)	Effect size
BRAF-MDQ						
Global (0–70)	42.4 (14.7)	34.6 (17.9)	−7.74	−12.11, −3.38	0.001	0.56
Physical (0–22)	16.8 (3.5)	13.9 (6.1)	−2.89	−4.59, −1.2	0.001	0.54
Living (0–21)	11.4 (5.7)	9.1 (6.1)	−2.32	−3.69, −0.95	0.001	0.53
Cognition (0–15)	8.0 (4.3)	6.8 (4.4)	−1.20	−2.34, −0.05	0.041	0.33
Emotion (0–12)	6.2 (3.6)	4.8 (3.3)	−1.34	−2.34, −0.34	0.010	0.42
BRAF-NRS						
Severity (0–10)	7.2 (1.7)	5.9 (2.5)	−1.25	−2.07, −0.43	0.004	0.47
Effect (0–10)	7.0 (2.3)	5.8 (2.6)	−1.25	−2.09, −0.41	0.005	0.46
Coping (0–10)[a]	5.6 (2.2)	5.4 (2.5)	−0.13	−1.03, 0.77	0.771	0.05
SF-36 VT (0–100)[a]	29.3 (17.9)	37.0 (21.0)	7.68	0.48, 14.88	0.037	0.34
FACIT-F (0–52)[a]	19.3 (9.8)	24.9 (12.8)	5.57	2.43, 8.71	0.001	0.55
POMS (0–28)	19.1 (6.1)	15.7 (7.7)	−3.46	−5.59, −1.32	0.002	0.53
MAF (1–50)	35.9 (8.9)	30.5 (10.1)	−5.34	−8.65, −2.04	0.002	0.57

[a]Higher scores indicate better outcome.

the concepts, questions and response options were explored during cognitive interviewing [11]. Data from the current study show that the BRAFs are reliable when the patient's condition does not change, and sensitive when it does, adding further to the validation evidence. Strong reliability was shown, evidenced consistently in two different cohorts of patients (studies 1 and 2). The choice of a short time frame (minimum 60 min between completions) was designed to accommodate the sudden and dramatic onset of fatigue reported in the literature [6–8] and to allow patients to consider the same 7-day period for both completions. The reliability of the BRAF-NRS Coping was less strong, but adequate, at $r = 0.62$, based on accepted levels for PROM validation [17, 18]. It was postulated that this was due to the absence of the word cope in the anchor statements, as a reminder of the question, but reliability was unchanged when this wording was clarified (reliability study 2). Another possibility is that presenting the BRAF-NRS Coping as the third in a page of three NRSs and the only one to be reverse-scored might have been unclear for patients. However, during the design of the BRAFs, patient focus groups expressed a strong consensus that positive coping must be represented by a higher number if it is to make sense to them [11]. Such patient involvement in the clarity of wording during the development of PROMs is recommended best practice [20]. A very similar coping scale has been used in a randomized controlled trial (RCT) of cognitive behavioural therapy (CBT) for fatigue self-management [34], the only difference being in the lower anchor, which was 'very poorly' (compared with the BRAF-NRS 'not at all well'). In the RCT this VAS was also reliable, being unchanged over 18 weeks in the control group [week 0 mean 6.0 (S.D. 2.3); week 18 mean 5.98 (S.D. 2.5), $n = 62$] [34].

The BRAF scales showed sensitivity to change following a single i.m. injection of high-dose glucocorticoids, even though patients were not recruited with fatigue as an inclusion criterion. As hypothesized, the BRAF-NRS Coping did not reflect change after a pharmacological intervention that was not designed to change coping skills. This probably reflects a true state in the underlying construct (i.e. coping with fatigue did not change) rather than a lack of responsiveness in the BRAF-NRS Coping item. This is supported by data from the RCT of CBT for RA fatigue in which the previously described, almost identical coping VAS was used and was very responsive to an intervention designed to improve fatigue coping skills (ES 0.79) [34].

The BRAF-NRS for Coping with fatigue therefore behaves differently from those for fatigue severity and effect. Previous research has shown that a patient can have a disconnect between the severity of fatigue, the effect of fatigue and their perceived ability to cope with it [12]. These findings are coherent with the theory of coping as a distinct concept [35] that can be assessed separately from symptom severity and effect. As coping is conceptualized as a multidimensional construct [36], it is possible that a multidimensional measure would be able to detect changes in aspects of coping that cannot be captured using a unidimensional instrument. In the future, it could be useful to use a multidimensional measure of coping with fatigue, particularly in relation to psychosocial interventions aimed at enhancing coping ability. The measurement of patient outcomes has traditionally focussed on severity levels, but in the past decade, evidence has increased that patients and professionals perceive severity levels differently [37, 38]. It is likely that patients' assessments of symptom severity or change are not formed by severity level alone [39], but rather are a combination of a triad of symptom severity, its importance in their lives and their ability to manage it [40]. All parts of this triad require measuring and, for the symptom of RA fatigue, this is now possible.

The provisional exploration of MCIDs suggested that a greater change in fatigue was required to trigger a perceived improvement on the transition question than was needed to perceive fatigue worsening (BRAF-MDQ 17.5% *vs* 6.1% or −7.43 *vs* +2.58). Similar differences in MCID between improvement and worsening have also been reported in the measurement of disease activity [41]. The sensitivity study was not powered for calculation of MCID, therefore these provisional findings need further exploration in a larger data set. In addition, fatigue was not necessarily the primary clinical issue for the participants receiving the single i.m. injection of high-dose glucocorticoids. It is possible that the findings would have been enhanced if participants had been recruited because of fatigue rather than because of an inflammatory flare. Nonetheless, they suggest that the mean BRAF scores for better, same and worse fatigue on the transition question are in the expected direction and pattern. A strength of these studies is that the larger number of PROMs, administered in different orders at each time point, meant that it was unlikely that patients would recall their previous answers.

The BRAFs performed as well as three of the existing comparator fatigue PROMs and better than the SF-36 vitality subscale. In addition, two of the existing scales had a large proportion of missing data, making them unusable, similar to rates reported in other studies [12, 42]. Furthermore, the additional value of the BRAFs beyond the existing fatigue scales is that for the first time they allow the measurement of different dimensions of RA fatigue and to separate measurement of fatigue severity from fatigue effect and fatigue coping. Evidence of validation for PROMs is never completed but depends on building a substantial body of evidence from a number of sources. The BRAFs have been translated into 34 languages and are currently being used in at least four multinational RCTs, and it is anticipated this will contribute further information to the body of evidence on validity, reliability and sensitivity.

Conclusions

These studies provide evidence contributing to the reliability (stability) and sensitivity of the BRAFs. It is now possible to measure four different dimensions of RA fatigue (such as physical and cognitive fatigue) separately, as well as distinguishing between fatigue severity, coping and effect. In future we should therefore be able to assess which components of an individual patient's fatigue are most troublesome, so that we can move toward developing and testing a range of interventions targeted appropriately to those individual needs.

Rheumatology key messages

- BRAFs are validated PROMs designed in collaboration with patients.
- BRAFs were reliable in stable patients and sensitive to change after intervention.
- In RA, coping with fatigue is distinct from severity and effect, requiring separate measurement.

Acknowledgements

The authors would like to thank the patients who participated in the studies.

Funding: This work was supported by Above and Beyond, University Hospitals Bristol NHS Foundation Trust (grant number 52301).

Disclosure statement: The authors have declared no conflicts of interest.

References

1 Hewlett S, Nicklin J, Treharne G. Fatigue in musculoskeletal conditions *Topical Reviews: Reports on the Rheumatic Diseases Series 6*. ISSN 1469–3097, Autumn 2008.

2 Pollard LC, Choy EH, Gonzalez J *et al*. Fatigue in rheumatoid arthritis reflects pain, not disease activity. Rheumatology 2006;45:885–9.

3 Selwyn C, Richards S, Westlake S. Fatigue in rheumatoid arthritis. Rheumatology 2012;51:1–2.

4 Hewlett S, Chalder T, Choy E *et al*. Fatigue in rheumatoid arthritis: time for a conceptual model. Rheumatology 2011;50:1004–6.

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

6 Hewlett S, Cockshott Z, Byron M *et al*. Patients' perceptions of fatigue in rheumatoid arthritis: overwhelming, uncontrollable, ignored. Arthritis Rheum 2005;53:697–702.

7 Repping-Wuts H, Uitterhoeve R, van Riel P *et al*. Fatigue as experienced by patients with rheumatoid arthritis (RA): a qualitative study. Int J Nurs Stud 2008;45:995–1002.

8 Nikolaus S, Bode C, Taal E *et al*. New insights into the experience of fatigue among patients with rheumatoid arthritis: a qualitative study. Ann Rheum Dis 2010;69:895–7.

9 Kirwan J, Minnock P, Adebajo A *et al*. Patient perspective: fatigue as a recommended patient-centred outcome measure in rheumatoid arthritis. J Rheumatol 2007;34:1174–7.

10 Hewlett S, Hehir M, Kirwan J. Measuring rheumatoid arthritis: a systematic review of scales in use. Arthritis Rheum 2007;57:429–39.

11 Nicklin J, Cramp F, Kirwan J *et al*. Collaboration with patients in the design of patient-reported outcome measures: capturing the experience of fatigue in rheumatoid arthritis. Arthritis Care Res 2010;62:1552–8.

12 Nicklin J, Cramp F, Kirwan J *et al*. Measuring fatigue in rheumatoid arthritis: a cross-sectional study to evaluate the Bristol Rheumatoid Arthritis Fatigue Multi-Dimensional questionnaire, visual analogue scales, and numerical rating scales. Arthritis Care Res 2010;62:1559–68.

13 BRAF scales. http://www1.uwe.ac.uk/hls/research/healthandclinicalresearch/researchareas/longtermconditions/fatiguescales/brafscales.aspx (4 March 2013, date last accessed).

14 Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. J Clin Epidemiol 1996;46:1417–32.

15 Beaton D, Bombardier C, Guillemin F *et al*. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine 2000;25:3186–91.

16 Wild D, Grove A, Martin M *et al*. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. Value Health 2005;8:94–104.

17 Fitzpatrick R, Davey C, Buxton MJ *et al*. Evaluating patient-based outcome measures for use in clinical trials. Health Technol Assess 1998;2:1–86.

18 Frost MH, Reeve BB, Liepa AM *et al*. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? Value Health 2007;10:S94–105.

19 Patrick DL, Burke LB, Powers JH *et al*. Patient-reported outcomes to support medical product labeling claims: FDA perspective. Value Health 2007;10(Suppl 2):S125–37.

20 US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Center for Devices and Radiological Health. Guidance for industry: patient-reported outcome measures. Use in medical product development to support labeling claims 2009. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf (21 May 2013, date last accessed).

21 Wells G, Li T, Maxwell L *et al*. Determining the minimal clinically important differences in activity, fatigue, and sleep quality in patients with rheumatoid arthritis. J Rheumatol 2007;34:280–9.

22 Khanna D, Pope J, Khanna PP *et al*. The minimally important difference for the fatigue visual analog scale in patients with rheumatoid arthritis followed in an academic clinical practice. J Rheumatol 2008;35:2339–43.

23 Arnett F, Edworthy S, Bloch D *et al*. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988; 31:315–24.

24 Fries J, Spitz P, Kraines G *et al*. Measurement of patient outcome in arthritis. Arthritis Rheum 1980;23: 137–45.

25 Cella D, Young S, Sorenson M *et al*. Validation of the functional assessment of chronic illness therapy fatigue scale relative to other instruments in patients with rheumatoid arthritis. J Rheumatol 2005;32:811–9.

26 Belza BL. Comparison of self-reported fatigue in rheumatoid arthritis and controls. J Rheumatol 1995;22: 639–43.

27 McNair D, Lorr M, Droppleman L. Profile of mood states manual. New York: Multihealth Systems, 1992.

28 Ware J, Sherbourne C. The MOS 36-item short form health survey (SF-36). Conceptual framework and item selection. Med Care 1992;30:473–83.

29 Wolfe F, Pincus T, O'Dell J. Evaluation and documentation of rheumatoid arthritis disease status in the clinic: which variables best predict change in therapy? J Rheumatol 2001;28:1712–7.

30 Meenan RF, Gertman PM, Mason JH. Measuring health status in arthritis: the Arthritis Impact Measurement Scales. Arthritis Rheum 1980;23:146–52.

31 Hewlett S, Hehir M, Kirwan J *et al*. The Personal Impact HAQ: sensitivity to change of a measure of the impact of disability. Rheumatol 2005;44S1:413 (abstract).

32 Nicklin J K. The development of scales to measure fatigue in people with rheumatoid arthritis. PhD 2009, University of the West of England, Bristol, UK.

33 Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307–10.

34 Hewlett S, Ambler N, Cliss A *et al*. Self-management of fatigue in rheumatoid arthritis: a randomised controlled trial of group cognitive-behavioural therapy. Ann Rheum Dis 2011;70:1060–7.

35 Connor-Smith JK, Flachsbart C. Relations between personality and coping: a meta-analysis. J Pers Soc Psychol 2007;93:1080–107.

36 Vriezekolk J, Van Lankveld W, Geenen R *et al*. Longitudinal association between coping and psychological distress in RA: a systematic review. Ann Rheum Dis 2011;70:1243–50.

37 Suarez-Almazor ME, Conner-Spady B, Kendall CJ *et al*. Lack of congruence in the ratings of patients' health status by patients and their physicians. Med Decis Making 2001; 21:113–21.

38 Kwoh CK, Ibrahim SA. Rheumatology patient and physician concordance with respect to important health and symptom status outcomes. Arthritis Care Res 2001; 45:372–7.

39 Beaton DE, Tarasuk V, Katz JN *et al*. 'Are you better?' A qualitative study of the meaning of recovery. Arthritis Rheum 2001;45:270–9.

40 Sanderson T, Hewlett S, Flurey C *et al*. The impact triad (incorporating severity, importance and self-management) as a method of enhancing the measurement of the personal life impact of rheumatic diseases. J Rheumatol 2011;38:191–4.

41 Leeb BF, Sautner J, Leeb BA *et al*. Lack of agreement between patients' and physicians' perspectives of rheumatoid arthritis disease activity changes. Scand J Rheumatol 2006;35:441–6.

42 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.