

Original article

The revised Bristol Rheumatoid Arthritis Fatigue measures and the Rheumatoid Arthritis Impact of Disease scale: validation in six countries

Sarah Hewlett¹, John Kirwan², Christina Bode³, Fiona Cramp¹, Loreto Carmona⁴, Emma Dures¹, Matthias Englbrecht⁵, Jaap Fransen⁶, Rosemary Greenwood⁷, Sofia Hagel⁸, Maart van de Laar³, Anna Molto⁹, Joanna Nicklin⁷, Ingemar F. Petersson⁸, Marta Redondo¹⁰, Georg Schett⁵ and Laure Gossec¹¹

Abstract

Objective. To evaluate the Bristol Rheumatoid Arthritis Fatigue Multidimensional Questionnaire (BRAFM-DQ), the revised Bristol Rheumatoid Arthritis Numerical Rating Scales (BRAFNRS V2) and the Rheumatoid Arthritis Impact of Disease (RAID) scale in six countries.

Methods. We surveyed RA patients in France, Germany, The Netherlands, Spain, Sweden and the UK, including the HAQ, 36-item Short Form Health Survey (SF-36) and potential revisions of the BRAFNRS coping and Spanish RAID coping items. Factor structure and internal consistency were examined by factor analysis and Cronbach's α and construct validity by Spearman's correlation.

Results. A total of 1276 patients participated (76% female, 25% with a disease duration <5 years, median HAQ 1.0). The original BRAFM-DQ four-factor structure and RAID single-factor structure were confirmed in every country with $\geq 66\%$ of variation in items explained by each factor and all item factor loadings of 0.71–0.98. Internal consistency for the BRAFM-DQ total and subscales was a Cronbach's α of 0.75–0.96 and for RAID, 0.93–0.96. Fatigue construct validity was shown for the BRAFM-DQ and BRAFNRS severity and effect scales, correlated internally with SF-36 vitality and with RAID fatigue ($r=0.63$ – 0.93). Broader construct validity for the BRAFs and RAID was shown by correlation with each other, HAQ and SF-36 domains ($r=0.46$ – 0.82), with similar patterns in individual countries. The revised BRAFNRS V2 Coping item had stronger validity than the original in all analyses. The revised Spanish RAID coping item performed as well as the original.

Conclusion. Across six European countries, the BRAFM-DQ identifies the same four aspects of fatigue, and along with the RAID, shows strong factor structure and internal consistency and moderate–good construct validity. The revised BRAFNRS V2 shows improved construct validity and replaces the original.

Key words: PROMs, validation, cultural, fatigue, impact, rheumatoid arthritis

¹Nursing & Midwifery, University of the West of England,
²Musculoskeletal Research, Bristol University, Bristol, UK, ³Department of Psychology, Health and Technology, University of Twente, Twente, The Netherlands, ⁴Instituto de Salud Musculoesquelética, Madrid, Spain, ⁵Department of Internal Medicine 3, Friedrich-Alexander-University Erlangen-Nürnberg and Universitätsklinikum Erlangen, Erlangen, Germany, ⁶Rheumatology Department, Radboud University, Nijmegen, The Netherlands, ⁷Department of Research and Innovation, Bristol Royal Infirmary, Bristol, UK, ⁸Orthopaedics, Clinical Sciences Lund, Lund University, Lund, Sweden, ⁹Rheumatology Department,

Cochin Hospital, AP-HP, Paris Descartes University, Paris, France, ¹⁰Department of Psychology, Universidad Camilo José Cela, Madrid, Spain and ¹¹Rheumatology Department, Sorbonne Universités, UPMC Univ Paris 06, Hôpital Pitié Salpêtrière, AP-HP, Paris, France

Submitted 6 February 2017; revised version accepted 31 August 2017

Correspondence to: Sarah Hewlett, Academic Rheumatology Unit, Zone B-504, Bristol Royal Infirmary, Bristol BS2 8HW, UK.
E-mail: Sarah.Hewlett@uwe.ac.uk

Rheumatology key messages

- Rheumatoid arthritis fatigue and impact measures demonstrate similar structure and validity across six European countries.
- Fatigue comprises similar multidimensional components across western cultures, challenging the value of global fatigue measures.
- The revised Bristol RA Fatigue Numerical Rating Scales demonstrates stronger validation for coping with fatigue.

Introduction

Pharmacological and non-pharmacological interventions aim to reduce the impact of RA, such as pain, disability, fatigue and distress [1–3]. Such outcomes are evaluated using patient-reported outcome measures (PROMs) such as the Bristol Rheumatoid Arthritis Fatigue Multidimensional Questionnaire (BRAF-MDQ) and Bristol Rheumatoid Arthritis Numerical Rating Scales (BRAF-NRS, a trio of simple NRSs for fatigue severity, effect and coping) and the RA Impact of Disease (RAID) scale, developed to measure broader impacts of RA not captured by existing single item PROMs for pain, disability and function [4–9]. As studies are increasingly conducted internationally, the RAID was developed simultaneously in 12 European countries. After development in the UK, the BRAFs were translated into 37 languages using formal protocols. BRAF translations were conducted by a medical PROM translation company, supported by the BRAF authors and based on internationally published guidelines [10]. Initial RAID translations were undertaken by rheumatology clinicians using a similar but slightly simpler protocol, although later translations were professionally performed. In a systematic review identifying >30 published guidelines on the translation and cultural adaptation of PROMs [11], and in cognitive interviewing studies [12], these two methods appeared to be equally robust. Whether the questionnaires retain equivalence in their full internal structure and appropriate construct validity is the subject of this report.

While terminologies differ, classic texts and authoritative review articles highlight the requirement for ensuring that adaptations have conceptual equivalence (e.g. meaning of distress), item equivalence (e.g. relevance of gardening in a high-rise city), operational equivalence (e.g. scoring systems such as an NRS) and semantic equivalence (e.g. terms for coping) across cultures before a PROM is ready for testing of psychometric (measurement) equivalence [13, 14]. Cognitive interviewing is a systematic methodology for this preparatory work [15–17], and a recent study demonstrated item and semantic equivalence for the BRAF-MDQ, BRAF-NRS and RAID across six European countries [12]. The cognitive interviews potentially identified an operational and a conceptual weakness in two PROM items relating to coping [12]. For the BRAF-NRS coping item, many Dutch participants struggled with the operational issue of low scores reflecting poor coping (0, not at all well–10, very well) while for the other traditionally worded BRAF-NRS (severity and effect), high scores reflect a worse situation. This direction of layout had been requested by patients during BRAF development, as they felt a high score should reflect strong coping skills [4]. The phraseology of the RAID item on coping confused many

Spanish participants [12], and this appeared to relate to a conceptual issue about capturing the fundamental meaning of coping. Therefore the present study included different versions of these two items so their performance could be compared.

The cognitive interviewing study prepared the way for the current study, a psychometric evaluation of validity across the six countries. Psychometric equivalence includes many aspects of measurement performance, including criterion validity (accuracy), construct validity (association with variables that make theoretical sense), a stable structure (e.g. factors or subscales), internal consistency (items capturing cohesive concepts), sensitivity to change and test-retest reliability (stability) [18]. Construct validity, a consistent structure of four distinct subscales that reflect different elements of fatigue, and internal consistency have been demonstrated for the BRAF-MDQ in the UK [5] but have not yet been tested in other European countries. If the translations capture concepts that mirror those demonstrated in the UK, then the items within each factor should also represent a single dimension in each country that can be tested by an appropriate factor analysis (see Methods section). If these factors are inconsistent across countries, this might indicate important cultural differences in fatigue perception. Further, construct validity of the RAID, a stable structure and internal consistency have been demonstrated in many countries [8] but not yet evaluated in Sweden, where the Swedish RAID has been newly translated.

While desirable, it would be impractical to examine the construct of RA fatigue in all 37 BRAF translations simultaneously. However, examining the performance in several countries could provide important proof-of-concept information on whether the fatigue components appear common. Since the impact of RA is a useful construct for comparison with fatigue, and since the new Swedish RAID has not yet been validated, collaboration with the RAID developers in a combined study is pragmatic. The aims of this study were therefore to evaluate the factor structure, internal consistency and construct validity of the BRAF-MDQ, BRAF-NRS and RAID across six European countries as exemplars and to test the construct validity of a possible revision of the BRAF-NRS coping item and of the Spanish RAID coping item, which arose from a previous cognitive interviewing study in these countries [12].

Methods

Study design

A cross-sectional study was performed across France, Germany, The Netherlands, Spain, Sweden and the UK during 2013–14.

Patients

Consecutive patients attending hospital rheumatology outpatient departments were invited to participate or were invited by post using departmental databases (Spain included a private practice). Patients had to be ≥ 18 years of age with a diagnosis of RA [19], and were excluded only if they lacked the capacity to consent.

Data collection

Demographic data comprised gender, age (<40, 40–59, ≥ 60 years), disease duration (<5 years, ≥ 5 years), medication and patient global NRS (Considering all the ways your arthritis affects you, please mark below how well you are doing: 0, very well–10, very poor) [20]. Patients completed the relevant language translations of the BRAF-MDQ, BRAF-NRS, revised BRAF-NRS V2 and RAID (supplementary data, section BRAF-MDQ, Revised BRAF-NRS V2 and the RAID: The Scales, available at *Rheumatology* Online), with Spanish patients also completing a revised Spanish RAID coping item. The BRAF-MDQ comprises 20 items (yielding a total score of 0–70) and four subscales of physical fatigue (0–22), living with fatigue (0–21), cognitive fatigue (0–15) and emotional fatigue (0–12), with high scores representing worse fatigue [4, 5]. The BRAF-NRS comprises three items measuring fatigue severity, effect and coping. The BRAF-NRS for severity and effect have high scores reflecting worse situations (0–10). However, as requested by patients during development, the BRAF-NRS coping scale is scored with 10 reflecting strong coping skills (Please circle the number which shows how well you have coped with fatigue over the past 7 days: 0, not at all well–10, very well) [4]. Based on the cognitive interviewing study [12], a potential revision of the BRAF-NRS coping item was developed with 10 as worse coping (0, very well–10, not at all well) and tested here as an additional item. The RAID contains seven NRSs to capture the impact of RA on pain, functional disability, fatigue, sleep, physical and emotional well-being and coping [7–9]. Each NRS score is weighted for importance using standardized weightings derived from a patient survey, then all seven are summed to yield a single, composite score from 0 to 10, where a high score indicates worse health [7–9]. A potential revised RAID coping item for Spain was included based on the cognitive interviewing results, using the phrase ‘arreglado (ha llevado, afrontado, se ha apañado)’ [sort out (deal with, cope with, manage)] [12]. Comparator questionnaires for validation of the BRAFs and RAID were the HAQ and 36-item Short Form Health Survey (SF-36). The HAQ assesses disability using 20 items, yielding a total score of 0–3, where 3 represents worse disability [21]. The SF-36 comprises eight health domains (physical function, physical role, emotional role, social function, pain, mental health, vitality and general health) yielding scores of 0–100, where 0 represents worse health [22]. To prevent any order effects, two versions of the questionnaire packs were created with the PROMs in different orders. Furthermore, the two potential revised coping items were separated from their original versions by other questionnaires and also presented in two different orders. Paper questionnaire packs were distributed in no particular order

and patients completed them either in the clinic or at home. BRAFs are freely available from <http://www1.uwe.ac.uk/hls/research/healthandclinicalresearch/researchareas/longtermconditions/fatiguescales.aspx> and the RAID from <http://oml.eular.org/>.

Ethics

Research ethics committee approval was obtained in the UK (London, City Road and Hampstead Ethics Committee, 12/LO/1198). The principal investigator in each country then obtained local approvals as required. Patients provided written informed consent. Anonymous questionnaires were returned to Bristol, UK for data entry and analysis.

Analysis

Missing data from individual PROMs were dealt with according to individual PROM scoring guidelines. A statistical analysis plan was agreed to in advance and analysis was performed on the total dataset and by individual country. Factor structure was examined for the BRAF-MDQ by testing whether the items in each of the original factors also represent a single dimension in the new data collected for the present study. Using the items within each of the original factors, a separate factor analysis was conducted for each country to identify the number of factors with eigenvalues >1 . If only a single factor emerged from each analysis, this confirmed agreement with the original UK factor. In addition, the robustness of these results was investigated by repeated factor analyses on random samples of 50% of patients (bootstrapping, 20 samples \times 6 countries \times 4 factors = 480 tests). A similar factor analysis was also performed on the RAID to check for the original unidimensional structure across countries. The internal consistency of the BRAF-MDQ and RAID was tested using Cronbach's α . Construct validity for the BRAF-MDQ and BRAF-NRS severity and effect was evaluated using Spearman's correlation to examine the pattern of correlations between the BRAF-MDQ subscales, with the RAID fatigue item and with the SF-36 vitality subscale. Broader construct validity of the BRAFs and RAID was evaluated through Spearman's correlation with the remaining eight SF-36 health subscales and the HAQ. The performance of potential revisions to the coping items was evaluated by comparing the construct validity of the revisions with the originals. Construct validity for the revised BRAF-NRS V2 (coping) was first evaluated by correlation with the original item, all other fatigue items and the RAID coping item and then with the remaining SF-36 health subscales, RAID and HAQ, as was the revised Spanish RAID coping item (Spearman's correlations). Findings similar to the original validation during development [5, 8] would indicate acceptability, with a high proportion of variance accounted for and meeting recommended criteria of Cronbach's $\alpha \geq 0.7$ and correlation ≥ 0.45 [23]. Factor analysis requires 4–10 respondents per item; to accommodate the 20-item BRAF-MDQ, a sample size of 100–200 patients per country was sought

[24]. Analysis was performed using SPSS version 21.0 (IBM, Armonk, NY, USA).

Results

A total of 1276 patients participated, achieving ≥ 150 patients per country. There was a preponderance of female participants, people >60 years of age and those with longer disease duration, reflecting the wider RA population (Table 1). Biologic agents were being used by 37% of patients (UK 16% to Sweden 65%). The mean fatigue and impact scores were moderate with a good range of responses, as indicated by the wide standard deviations. The UK participants had slightly poorer health status than other countries for most indices. The overall proportion of missing data was small at 2.7% (345 of 12 760 items; supplementary Table S1, available at *Rheumatology* Online).

Factor structure

The factor analysis showed that the original four factors of the BRAF-MDQ (physical, living, cognitive and emotional fatigue) and the single RAID factor were upheld in the total patient sample taken together and in each of the individual countries. All items loaded strongly on their respective factors (0.71–0.98; supplementary Table S2, available at *Rheumatology* Online), explaining a high proportion of the variance for the items in each factor (all $\geq 66\%$; Table 2).

Repeated analyses using random 50% samples of the data demonstrated the robustness of the factor structure, with 476/480 tests (99%) showing the same (four factor) structure for the BRAF-MDQ and indicating that the structure was not dependant on including particular respondents. In one of each of the 20 tests for France, Germany, Spain and Sweden, the living factor was split into two factors (daily life, social life).

Internal consistency

Cronbach's α demonstrated internal consistency by country and for the total sample for the BRAF-MDQ total score, the BRAF-MDQ subscales and the RAID (Table 3). The internal consistency for the BRAF-MDQ physical subscale was slightly lower than other subscales, but still satisfactory at ≥ 0.75 in all countries.

Construct validity with fatigue items

Construct validity for the BRAF-MDQ and BRAF-NRS for severity and effect was demonstrated by moderate to strong correlation between the fatigue subscales and NRS, and with the SF-36 vitality subscale and the RAID fatigue item (Table 4; $r = 0.63$ – 0.93). This pattern was similar in all individual countries (supplementary Table S3, available at *Rheumatology* Online).

Construct validity with related broader concepts

The BRAF-MDQ, BRAF-NRS severity, BRAF-NRS effect and RAID showed good construct validity through moderate to strong correlations with each other, disability (HAQ) and the SF-36 health domains (Table 5; $r = 0.46$ – 0.82). This pattern of correlations was replicated

in individual countries (supplementary Table S4, available at *Rheumatology* Online).

Validation of potential revised coping items

The revised BRAF-NRS V2 coping item showed stronger construct validity than the original NRS coping item overall (Table 6; revised $r = 0.15$ – 0.74 , original $r = 0.09$ – 0.48) and in each country (supplementary Table S5, available at *Rheumatology* Online). An unexpected finding during validity testing of the BRAF-NRS coping item was that for Spain ($n = 157$), but none of the other countries ($n = 1119$), the construct correlations for the original BRAF-NRS coping version were always in the inappropriate direction (except for correlation with BRAF-MDQ total score). These inappropriate correlational directions were rectified with the revised BRAF-NRS V2 coping item (Table 6). Based on these results, the BRAF-NRS V2 containing the original severity and effect items but the revised coping item now replaces the original BRAF-NRS. Since the original BRAF-NRS coping item correlates only weakly with the revised BRAF-NRS V2 coping item ($r = -0.348$ for a five country sample, -0.082 for Spain), patient ratings already obtained using the original version cannot simply be reverse scored to make them comparable to V2.

The Spanish alternative wording for the RAID coping item correlated well with the original item ($r = 0.89$, $n = 157$) and demonstrated a similar pattern of construct validity with the eight SF-36 health subscales (original $r = 0.51$ – 0.66 , revised $r = 0.39$ – 0.74 ; supplementary Table S6, available at *Rheumatology* Online). Therefore the original Spanish wording for this item was retained.

Discussion

This study tested the performance of the BRAF-MDQ, the BRAF-NRS and the RAID in six European countries for the first time and found that they retained their internal structure and satisfactory construct validity in all countries. The study also demonstrated the improved performance of the revised BRAF-NRS V2 coping item compared with the original, which it now replaces, and confirmed that there was no need to replace the original RAID coping item for Spain, which performed satisfactorily. These novel data suggest these PROMs can be used with confidence in a range of countries, facilitating international comparison.

Factor analysis demonstrated that the four original fatigue subscales developed in the BRAF-MDQ (physical fatigue, living with fatigue, cognitive fatigue and emotional fatigue) are confirmed in a fresh UK population and are also confirmed in five other European countries. This suggests fatigue has a similar construct from the patient's perspective across these western cultures, although this has yet to be tested in other cultures [25]. Such generalizability was hypothesized from cognitive interviewing across these European countries [12] and strengthens the argument that fatigue is not a unidimensional concept but has aspects that may vary between patients [4, 5], potentially questioning the value of the BRAF-MDQ total score and global fatigue questions or PROMs.

TABLE 1 Demographic data overall (*n* = 1276) and by country

Characteristic	France (<i>n</i> = 206)	Germany <i>n</i> = 216)	The Netherlands (<i>n</i> = 317)	Spain (<i>n</i> = 157)	Sweden (<i>n</i> = 170)	UK (<i>n</i> = 210)	Total (<i>n</i> = 1276)
Female, <i>n</i> (%)	176 (85)	60 (70)	194 (61)	138 (88)	130 (77)	165 (79)	954 (76)
Age, <40 years, <i>n</i> (%)	28 (14)	14 (7)	14 (5)	26 (17)	13 (8)	29 (14)	124 (10)
Age, >60 years, <i>n</i> (%)	88 (43)	109 (51)	192 (62)	42 (27)	108 (65)	91 (44)	630 (50)
Disease duration, ≤5 years, <i>n</i> (%)	43 (22)	59 (28)	65 (21)	30 (19)	36 (22)	80 (42)	313 (26)
Biologics, <i>n</i> (%)	95 (46)	65 (30)	105 (33)	66 (42)	108 (64)	33 (16)	472 (37)
BRAF-MDQ total (0–70)	27.5 (16.6)	22.0 (14.4)	22.0 (14.0)	26.1 (16.7)	27.1 (16.0)	34.2 (17.3)	26.0 (16.2)
Physical (0–22)	11.7 (5.7)	10.4 (5.5)	11.4 (5.6)	11.3 (6.2)	12.5 (5.7)	14.0 (5.6)	11.8 (5.8)
Living (0–21)	6.6 (5.7)	5.8 (5.1)	5.3 (4.7)	6.6 (5.4)	6.6 (5.2)	8.7 (6.0)	6.5 (5.4)
Cognitive (0–15)	4.4 (3.9)	3.5 (3.2)	3.1 (3.2)	4.7 (4.1)	4.7 (3.9)	6.2 (4.4)	4.3 (3.9)
Emotional (0–12)	4.6 (3.7)	2.4 (2.6)	2.2 (2.5)	3.6 (3.0)	3.5 (3.2)	5.0 (3.4)	3.4 (3.2)
BRAF-NRS							
Severity (0–10)	4.7 (2.4)	4.1 (2.5)	4.6 (2.3)	4.9 (2.8)	5.2 (2.5)	5.9 (2.7)	4.8 (2.6)
Effect (0–10)	4.4 (2.5)	3.8 (2.5)	4.3 (2.4)	4.8 (2.9)	4.9 (2.6)	5.8 (2.9)	4.6 (2.7)
Coping (original) (0–10) ^a	6.0 (2.4)	5.8 (2.7)	6.1 (2.4)	5.7 (2.8)	5.8 (2.5)	6.0 (2.6)	6.0 (2.6)
Coping V2 (revised) (0–10)	3.8 (2.5)	3.4 (2.4)	3.6 (2.5)	4.2 (2.9)	4.1 (2.5)	5.2 (2.9)	4.0 (2.7)
RAID (0–10)	3.8 (2.2)	3.7 (2.3)	3.7 (2.0)	4.4 (2.4)	4.5 (2.2)	5.2 (2.5)	4.2 (2.3)
HAQ (0–3)	0.8 (0.6)	0.9 (0.7)	1.0 (0.7)	1.1 (0.7)	1.1 (0.6)	1.3 (0.8)	1.0 (0.7)
Patient global (0–10)	4.3 (2.4)	3.9 (2.5)	4.1 (2.3)	4.6 (2.5)	5.0 (2.4)	4.8 (2.6)	4.4 (2.4)
SF-36							
Physical function (0–100) ^a	59.0 (26.5)	61.6 (28.9)	57.3 (25.5)	52.1 (26.7)	50.4 (24.2)	37.8 (29.5)	53.5 (28.0)
Physical role (0–100) ^a	37.4 (42.4)	46.0 (42.1)	41.0 (42.3)	38.9 (42.5)	36.7 (40.2)	23.0 (35.5)	37.6 (41.6)
Emotion role (0–100) ^a	53.6 (42.3)	59.9 (45.3)	66.8 (42.2)	54.7 (45.7)	56.7 (44.0)	44.7 (44.1)	57.1 (44.3)
Social role (0–100) ^a	59.9 (21.6)	63.1 (23.4)	65.0 (19.4)	56.4 (24.2)	58.4 (23.7)	48.2 (24.9)	59.2 (23.3)
Pain (0–100) ^a	53.8 (23.4)	52.2 (24.1)	55.1 (20.6)	47.2 (22.9)	44.7 (21.0)	40.4 (25.2)	49.6 (23.4)
Mental health (0–100) ^a	63.0 (18.7)	63.2 (20.8)	72.9 (17.2)	58.9 (20.3)	68.0 (22.3)	60.7 (20.5)	65.3 (20.3)
Vitality (0–100) ^a	51.5 (20.2)	46.9 (22.2)	54.2 (19.4)	44.4 (21.1)	44.7 (24.9)	35.5 (19.9)	47.0 (22.0)
General health (0–100) ^a	46.4 (20.09)	50.3 (20.5)	49.6 (19.0)	40.4 (18.7)	45.4 (22.8)	39.4 (21.7)	45.9 (20.9)

Values presented as mean (s.d.) unless otherwise stated. ^aLow score = worse health. BRAF-MDQ: British Rheumatoid Arthritis Fatigue – Multidimensional Questionnaire; BRAF-NRS: British Rheumatoid Arthritis Fatigue – Numerical Rating Scale; RAID: RA Impact of Disease; SF: Short-form Health Survey (*n* = 36).

For example, while the Functional Assessment of Chronic Illness Therapy–Fatigue has been used in RA clinical trials, it was not developed for or with RA patients and provides only a unidimensional global fatigue score, as does the Patient-Reported Outcomes Measurement Information System [26, 27]. The three unidimensional BRAF-NRS items nonetheless provide a rapid assessment of fatigue severity, effect and coping where questionnaire burden would otherwise be high and fatigue is not the primary aim of treatment or research.

The internal consistency of the BRAF-MDQ, BRAF-NRS and RAID was further supported by construct validity in six European countries, providing novel data for the BRAF scales and the new Swedish RAID. Construct validity for the BRAFs was demonstrated against the SF-36 vitality domain and the RAID fatigue question (both fatigue severity items). These had strong correlations with the BRAF-MDQ total, physical and living with fatigue items and with the BRAF-NRS severity and effect items and more moderate correlations with the BRAF-MDQ emotional and cognitive fatigue subscales. The pattern of slightly lower correlations for emotional and cognitive fatigue are

appropriate since these concepts likely include an individual's response to fatigue; they also reflect the original validation patterns seen during UK development [5].

The similar performance of the original and potential revised wording for the Spanish RAID coping item shows that more than one expression may capture the concept of coping and produce similar results (original version was therefore retained). There is no consensus on the common core concepts to be measured in a coping scale, and a review of 100 coping PROMs identified >400 coping strategies that could be inquired about [28]. These specific behavioural, emotional and cognitive coping strategies may relate to the concept of self-management highlighted as an important element to be captured in RA, but for which there is yet no specific PROM [29]. In contrast, both the BRAF-MDQ and RAID coping items are simple ratings of the patient's perspective of their current overall coping success, rather than asking about strategies, and might be tapping into the broad construct of flexible coping seen in RA [30].

In contrast, the issue for the original BRAF-NRS coping item was operational, relating to the direction of scoring.

TABLE 2 Percentage of variation in the items accounted for by each factor

Factor	Number of items	France (n = 206)	Germany (n = 216)	The Netherlands (n = 317)	Spain (n = 157)	Sweden (n = 170)	UK (n = 210)	Total (n = 1276)
BRAF-MDQ								
Physical	4	66.1	69.0	70.4	73.1	73.5	72.5	70.5
Living	7	67.0	73.8	68.9	74.8	69.0	73.5	71.1
Cognitive	5	74.1	75.2	76.2	84.4	78.6	84.8	80.0
Emotional	4	74.7	70.1	72.4	73.3	75.9	75.6	75.0
RAID	7	71.7	79.7	74.2	78.0	74.7	76.0	75.6

BRAF-MDQ: British Rheumatoid Arthritis Fatigue – Multidimensional Questionnaire; RAID: RA Impact of Disease.

TABLE 3 Internal consistency of BRAF-MDQ (Cronbach's α , overall and by country)

Factor	France (n = 206)	Germany (n = 216)	The Netherlands (n = 317)	Spain (n = 157)	Sweden (n = 170)	UK (n = 210)	Total (n = 1276)
BRAF-MDQ total	0.938	0.934	0.932	0.943	0.947	0.954	0.944
Physical	0.750	0.752	0.763	0.782	0.791	0.785	0.770
Living	0.918	0.941	0.924	0.943	0.923	0.940	0.932
Cognitive	0.912	0.916	0.920	0.953	0.931	0.955	0.937
Emotional	0.885	0.852	0.873	0.878	0.893	0.892	0.888
RAID	0.932	0.955	0.938	0.951	0.941	0.945	0.947

BRAF-MDQ: British Rheumatoid Arthritis Fatigue – Multidimensional Questionnaire; RAID: RA Impact of Disease.

TABLE 4 Construct validity of fatigue impact, severity and effect with other fatigue items (correlation coefficient)

Factor	BRAF-MDQ physical	BRAF-MDQ living	BRAF-MDQ cognitive	BRAF-MDQ emotion	BRAF-NRS severity	BRAF-NRS effect	RAID fatigue	SF-36 ^a vitality
BRAF-MDQ								
Physical							0.858	−0.767
Living	0.779						0.733	−0.723
Cognitive	0.652	0.707					0.623	−0.632
Emotional	0.676	0.758	0.788				0.648	−0.646
Total	0.900	0.925	0.845	0.869	0.850	0.854	0.822	−0.790
BRAF-NRS								
Severity	0.903	0.745	0.633	0.659			0.892	−0.732
Effect	0.866	0.771	0.646	0.696	0.915		0.857	−0.735

^aLow score = worse health. BRAF-MDQ: British Rheumatoid Arthritis Fatigue – Multidimensional Questionnaire; BRAF-NRS: British Rheumatoid Arthritis Fatigue – Numerical Rating Scale; SF: Short-form Health Survey (n = 36).

This direction was formulated in research with patients who felt that using 0 for not coping and 10 for coping well made better sense than the traditional direction [4], although the majority of fatigue PROMs used in rheumatology are scored in the opposite direction [26]. In the cognitive interviewing study that preceded the present report, it was Dutch patients who queried this [12], but in the present study it was Spanish patients who scored the original BRAF-NRS coping item in the opposite direction, generating inappropriately negative correlations (Table 6). In contrast, for the revised BRAF-NRS V2

coping item (0 = coping well), these correlations were in the appropriate direction for all countries, including Spain. It is possible that the common use of the double negative in Spain [31] may be related to their interpretation of the original 0 for not coping. This illustrates the care required to ensure equivalence of questionnaires across countries.

The construct validity for the original BRAF-NRS coping item is acceptable, but that for the revised BRAF-NRS V2 coping item is improved. The correlation between the original and revised coping items is weak ($r = 0.342$),

TABLE 5 Construct validity of disease impact, fatigue impact, severity and effect with related concepts (correlation coefficient)

Factor	RAID	HAQ	SF-36 ^a physical function	SF-36 ^a physical role	SF-36 ^a emotional role	SF-36 ^a social role	SF-36 ^a pain	SF-36 ^a mental health	SF-36 ^a general health
RAID		0.639	-0.616	-0.641	-0.497	-0.680	-0.783	-0.574	-0.697
BRAF-MDQ									
Total	0.786	0.592	-0.603	-0.677	-0.580	-0.760	-0.657	-0.630	-0.604
Physical	0.781		-0.558	-0.613	-0.465	-0.664	-0.634	-0.515	-0.567
Living	0.730		-0.622	-0.678	-0.547	-0.736	-0.639	-0.569	-0.571
Cognitive	0.592		-0.427	-0.521	-0.520	-0.620	-0.496	-0.577	-0.476
Emotional	0.640		-0.456	-0.554	-0.569	-0.671	-0.527	-0.646	-0.508
BRAF-NRS									
Severity	0.815		-0.522	-0.576	-0.457	-0.653	-0.629	-0.503	-0.549
Effect	0.821		-0.569	-0.610	-0.501	-0.710	-0.650	-0.564	-0.562

^aLow score = worse health. BRAF-MDQ: British Rheumatoid Arthritis Fatigue - Multidimensional Questionnaire; BRAF-NRS: British Rheumatoid Arthritis Fatigue - Numerical Rating Scale; RAID: RA Impact of Disease; SF: Short-form Health Survey (n = 36).

TABLE 6 Construct validity for the revised BRAF-NRS V2 coping item (correlation coefficient)

Factor	Original BRAF-NRS ^a coping (10 = good) (n = 1119)	Revised BRAF-NRS V2 coping (10 = bad) (n = 1119)	Original BRAF-NRS ^a coping (10 = good) (Spain, n = 157)	Revised BRAF-NRS V2 coping (10 = bad) (Spain, n = 157)
With other fatigue items				
BRAF-NRS coping original ^a		-0.348		-0.082
BRAF-NRS V2				
Severity	-0.137	0.545	0.445	0.390
Effect	-0.154	0.562	0.438	0.366
BRAF-MDQ				
Total score	-0.389	0.714	-0.290	0.744
Physical	-0.153	0.525	0.478	0.350
Living	-0.166	0.464	0.348	0.214
Cognitive	-0.143	0.380	0.242	0.296
Emotional	-0.149	0.404	0.289	0.287
RAID question 7, coping	-0.175	0.536	0.369	0.291
SF-36 vitality subscale ^a	0.172	-0.412	-0.311	-0.258
With related concepts				
RAID	-0.135	0.558	0.396	0.344
HAQ	-0.135	0.391	0.304	0.163
SF-36				
Physical function ^a	0.140	-0.380	-0.314	-0.136
Physical role ^a	0.184	-0.401	-0.252	-0.209
Emotional role ^a	0.142	-0.334	-0.085	-0.186
Social role ^a	0.178	-0.419	-0.202	-0.325
Pain ^a	0.126	-0.391	-0.315	-0.265
Mental health ^a	0.171	-0.355	-0.104	-0.270
General health ^a	0.144	-0.382	-0.135	-0.153

^aHigh score = better health. BRAF-MDQ: British Rheumatoid Arthritis Fatigue - Multidimensional Questionnaire; BRAF-NRS: British Rheumatoid Arthritis Fatigue - Numerical Rating Scale; RAID: RA Impact of Disease; SF: Short-form Health Survey (N = 36).

suggesting that different patients are interpreting the original version in different directions. As it is not known which patients these are, investigators using the original BRAF-NRS coping item in ongoing studies should not be

tempted to reverse those scores, as they are not mirror images of each other. Investigators should continue to use the original version for studies currently under way, as its performance is acceptable.

A possible limitation of this study might be that the health status of the UK patient sample was slightly poorer than the other countries (HAQ, SF-36). This might have occurred through the approach to patients in the UK, which was largely at their clinic appointment and thus, when patients are stable, they only have routine check-ups every 2 years. In the meantime they have direct access to rapid appointments via a nurse helpline if they have a flare [32], meaning patients attending clinics are generally seeking help for their RA disease control. The UK sample also comprised fewer patients on biologic drugs than the other European countries (16% vs 30–64%), and lower UK biologic prescribing rates have been reported previously [33]. Nonetheless, analysis shows that the BRAFs and RAID performed similarly across the countries, thus demonstrating effective measurement at a range of different levels of health. The strength of this study is that it was preceded by systematic cognitive interviewing [12] and conducted in a large group of patients ($n = 1276$) with a sufficient number from each of the six countries to allow robust conclusions per country regarding questionnaire performance. There are examples in RA studies where PROMs have not been subjected to such detailed post-translation analysis, making it difficult to interpret the research findings, which may have significant clinical implications [34]. Validation of the BRAFs and RAID in other countries and cultures has yet to be undertaken.

This study demonstrates the potential usefulness of the BRAFs and RAID across European countries to measure and compare RA fatigue and impact. Further, it indicates that fatigue has a common four-factor structure across countries, paving the way for interventions targeting different issues (e.g. lifestyle, cognition, emotion) that could be tested internationally. Three further psychometric steps could build upon the work presented here. First, clarification of individual item performance by country could be explored, perhaps using Item Response Theory [35]. This would identify any potential for improving scale sensitivity by differential scoring within countries. Second, international longitudinal studies could measure reliability and sensitivity to change. This is particularly important to explore changes in the four fatigue dimensions, which might respond differentially depending on the nature of the intervention (e.g. physical and living with fatigue responding to pharmacological interventions and cognitive and emotional fatigue responding to cognitive interventions). Third, it would be useful to examine the BRAFs and RAID for evaluating fatigue, impact and coping beyond these western cultures.

Acknowledgements

The authors would like to thank the patients for participating and the study team patient research partners for their guidance [Beverley Davis (UK) and Gabrielle Von Krause (France)]. Contributions: S. Hewlett, J.K., C.B., L.C., E.M., J.N. and L.G. led the study design. C.B., L.C., M.E., J.F., S. Hagel, A.M. J.N., M.R. and L.G. collected the data. Data analysis was performed by R.G., with all authors

contributing to interpretation and discussion of the findings. All authors reviewed drafts of the manuscript and approved the final submission.

Funding: This study was funded by a EULAR grant (EP1009).

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

Supplementary data

Supplementary data are available at *Rheumatology* Online.

References

- Young A, Dixey J, Cox N *et al*. How does functional disability in early rheumatoid arthritis affect patients and their lives? Results of 5 years of follow-up in 732 patients from the Early RA Study (ERAS). *Rheumatology* 2000;39:603–11.
- Stoffer MA, Schoels MM, Smolen JS *et al*. Evidence for treating rheumatoid arthritis to target: results of a systematic literature search update. *Ann Rheum Dis* 2016;75:16–22.
- Pollard L, Choy EH, Scott DL. The consequences of rheumatoid arthritis: quality of life measures in the individual patient. *Clin Exp Rheumatol* 2005;235:43–52.
- 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.
- Nicklin J, Cramp F, Kirwan J *et al*. Measuring fatigue in RA: a cross-sectional study to evaluate the BRAF multi-dimensional questionnaire, visual analogue and numerical rating scales. *Arthritis Care Res* 2010;62:1559–6.
- Dures E, Cramp F, Greenwood R, Kirwan J, Hewlett S. Reliability and sensitivity to change of the Bristol Rheumatoid Arthritis Fatigue scales. *Rheumatology* 2013;52:1832–9.
- Gossec L, Dougados M, Rincheval N *et al*. Elaboration of the preliminary Rheumatoid Arthritis Impact of Disease (RAID) score: a EULAR initiative. *Ann Rheum Dis* 2009;68:1680–5.
- Gossec L, Paternotte S, Aanerud G *et al*. Finalisation and validation of the Rheumatoid Arthritis Impact of Disease (RAID) score: a patient-derived composite measure of impact of RA. A EULAR initiative. *Ann Rheum Dis* 2011;70:935–42.
- Heiberg T, Austad C, Kvien TK *et al*. Performance of the Rheumatoid Arthritis Impact of Disease (RAID) score in relation to other patient-reported outcomes in a register of patients with rheumatoid arthritis. *Ann Rheum Dis* 2011;70:1080–2.
- 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.

- 11 Epstein J, Miyuki Santo R, Guillemin F. A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. *J Clin Epidemiol* 2015;68:435–41.
- 12 Hewlett S, Nicklin J, Bode C *et al.* Translating patient reported outcome measures: methodological issues explored using cognitive interviewing with three rheumatoid arthritis measures in 6 European languages. *Rheumatology* 2016;55:1009–16.
- 13 Streiner DL, Norman GR. Devising the items. In: Streiner DL, Norman GR, eds. *Health Measurement Scales: A Practical Guide to Their Development and Use*. Oxford: Oxford University Press, 2008:29–33.
- 14 Herdman M, Fox-Rushby J, Badia X. A model of equivalence in the cultural adaptation of HRQoL instruments: the universalist approach. *Qual Life Res* 1998;7:323–35.
- 15 Drennan J. Cognitive interviewing: verbal data in the design and pretesting of questionnaires. *J Adv Nurs* 2003;42:57–63.
- 16 Collins D. Pretesting survey instruments: an overview of cognitive methods. *Qual Life Res* 2003;12:229–38.
- 17 Tourangeau R. Cognitive aspects of survey measurement and mismeasurement. *Int J Public Opin Res* 2003;15:3–7.
- 18 Boers M, Kirwan J, Tugwell P *et al.* The OMERACT Handbook. http://www.omeract.org/pdf/OMERACT_Handbook.pdf (14 July 2016, date last accessed).
- 19 Arnett FC, Edworthy SC, Bloch DA *et al.* The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988;31:315–24.
- 20 Meenan RF, Gertman PM, Mason JH. Measuring health status in arthritis: the arthritis impact measurement scales. *Arthritis Rheum* 1980;23:146–52.
- 21 Fries J, Spitz P, Kraines G *et al.* Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;23:137–45.
- 22 Ware J, Sherbourne C. The MOS 36-item short form health survey (SF-36). *Med Care* 1992;30:473–83.
- 23 DeVon H, Block M, Moyle-Wright P *et al.* A psychometric toolbox for testing validity and reliability. *J Nurs Scholarsh* 2007;39:155–64.
- 24 Lidwine B, Mokkink LB, Terwee CB, *et al.* Cosmin Checklist Manual. <http://www.cosmin.nl/images/upload/files/COSMIN%20checklist%20manual%20v9.pdf> (24 January 2017, date last accessed).
- 25 Malaviya A. Outcome measures in rheumatoid arthritis: Indian perspective. *Indian J Rheumatol* 2013;8:S36–7.
- 26 Hewlett S, Dures E, Almeida C. Measures of fatigue: Bristol Rheumatoid Arthritis Fatigue Multi-Dimensional Questionnaire (BRAFMQ), Bristol Rheumatoid Arthritis Fatigue Numerical Rating Scales (BRAFNRS) for severity, effect, and coping, Chalder Fatigue Questionnaire (CFQ), Checklist Individual Strength (CIS20R and CIS8R), Fatigue Severity Scale (FSS), Functional Assessment Chronic Illness Therapy (Fatigue) (FACIT-F), Multi-Dimensional Assessment of Fatigue (MAF), Multi-Dimensional Fatigue Inventory (MFI), Pediatric Quality of Life (PedsQL) Multi-Dimensional Fatigue Scale, Profile of Fatigue (Prof), Short Form 36 Vitality Subscale (SF-36 VT), and Visual Analog Scales (VAS). *Arthritis Care Res* 2011;63(Suppl 11):263–86.
- 27 Bartlett S, Orbai AM, Duncan T *et al.* Reliability and validity of selected PROMIS measures in people with rheumatoid arthritis. *PLoS One* 2015;10:e0138543.
- 28 Skinner EA, Edge K, Altman J *et al.* Searching for the structure of coping: a review and critique of category systems for classifying ways of coping. *Psychol Bull* 2003;129:216–69.
- 29 Bykerk V, Lie E, Alten R *et al.* Establishing a core domain set to measure rheumatoid arthritis flares: report of the OMERACT 11 RA flare workshop. *J Rheumatol* 2014;41:799–809.
- 30 Vriezekolk J, Van Lankveld WGJM, Eijsbouts AMM *et al.* The coping flexibility questionnaire: development and initial validation in patients with chronic rheumatic diseases. *Rheumatol Int* 2012;32:283–91.
- 31 Batchelor RE, San Jose MA. *A Reference Grammar of Spanish*. Cambridge: Cambridge University Press, 2010:416–21.
- 32 Hewlett S, Kirwan J, Pollock J *et al.* Patient-initiated outpatient follow-up in rheumatoid arthritis: six year randomised controlled trial. *BMJ* 2005;330:171–6.
- 33 Richards M. Extent and causes of international variations in drug usage. A report for the Secretary of State for Health. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/216249/dh_117977.pdf (14 July 2016, date last accessed).
- 34 Katz PP, Barton J, Trupin L. Poverty, depression, or lost in translation? Ethnic and language variation in patient-reported outcomes in rheumatoid arthritis. *Arthritis Care Res* 2016;68:621–8.
- 35 Streiner DL, Norman GR. Item response theory. In: Streiner DL, Norman GR, eds. *Health Measurement Scales: A Practical Guide to their Development and Use*. Oxford: Oxford University Press, 2008:299–330.