

Identifying flares in rheumatoid arthritis: reliability and construct validation of the OMERACT RA Flare Core Domain Set

Vivian P Bykerk,^{1,2} Clifton O Bingham,³ Ernest H Choy,⁴ Daming Lin,² Rieke Alten,⁵ Robin Christensen,⁶ Daniel E Furst,^{7,8} Sarah Hewlett,⁹ Amye Leong,¹⁰ Lyn March,¹¹ Thasia Woodworth,¹² Gilles Boire,¹³ Boulos Haraoui,¹⁴ Carol Hitchon,¹⁵ Shahin Jamal,¹⁶ Edward C Keystone,² Janet Pope,¹⁷ Diane Tin,¹⁸ J Carter Thorne,¹⁸ Susan J Bartlett,^{3,19} on behalf of the OMERACT RA Flare Group and CATCH Investigators

To cite: Bykerk VP, Bingham CO, Choy EH, *et al.* Identifying flares in rheumatoid arthritis: reliability and construct validation of the OMERACT RA Flare Core Domain Set. *RMD Open* 2016;**2**:e000225. doi:10.1136/rmdopen-2015-000225

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/rmdopen-2015-000225>).

VPB and COB are co-primary authors.

Received 8 December 2015
Revised 21 April 2016
Accepted 22 April 2016



CrossMark

For numbered affiliations see end of article.

Correspondence to
Dr Vivian P Bykerk;
bykerkv@hss.edu

ABSTRACT

Objective: To evaluate the reliability of concurrent flare identification using 3 methods (patient, rheumatologist and Disease Activity Score (DAS)28 criteria), and construct validity of candidate items representing the Outcome Measures in Rheumatology Clinical Trials (OMERACT) RA Flare Core Domain Set.

Methods: Candidate flare questions and legacy measures were administered at consecutive visits to Canadian Early Arthritis Cohort (CATCH) patients between November 2011 and November 2014. The American College of Rheumatology (ACR) core set indicators were recorded. Concordance to identify flares was assessed using the agreement coefficient. Construct validity of flare questions was examined: convergent (Spearman's r); discriminant (mean differences between flaring/non-flaring patients); and consequential (proportions with prior treatment reductions and intended therapeutic change postflare).

Results: The 849 patients were 75% female, 81% white, 42% were in remission/low disease activity (R/LDA), and 16–32% were flaring at the second visit. Agreement of flare status was low–strong (κ 's 0.17–0.88) and inversely related to RA disease activity level. Flare domains correlated highly (r 's ≥ 0.70) with each other, patient global (r 's ≥ 0.66) and corresponding measures (r 's 0.49–0.92); and moderately highly with MD and patient-reported joint counts (r 's 0.29–0.62). When MD/patients agreed the patient was flaring, mean flare domain between-group differences were 2.1–3.0; 36% had treatment reductions prior to flare, with escalation planned in 61%.

Conclusions: Flares are common in rheumatoid arthritis (RA) and are often preceded by treatment reductions. Patient/MD/DAS agreement of flare status is highest in patients worsening from R/LDA. OMERACT RA flare questions can discriminate between patients with/without flare and have strong evidence of construct and consequential validity. Ongoing work will

Key messages

- Flares are common in rheumatoid arthritis (RA) and are often preceded by treatment reductions.
- Patients and MDs generally agree when patients flare, especially when previously in remission/low disease activity.
- OMERACT RA flare questions show evidence of reliability and construct, discriminant and consequential validity.

identify optimal scoring and cut points to identify RA flares.

INTRODUCTION

People living with rheumatoid arthritis (RA) frequently experience transient increases in joint pain, swelling, and other symptoms such as stiffness and fatigue that indicate increased inflammation and worsening of their RA.^{1–2} These episodes vary widely in frequency, duration, and intensity. They can be severe and disabling.^{1–3} Patients and rheumatologists (MDs) often use the word 'flare' to describe such episodes. Flares are generally expected to be reversible, though elevated RA disease activity persists in some cases.

Flares become clinically relevant when they are of sufficient intensity and duration to suggest that current therapy may be inadequate and a change in treatment may be required to optimise disease management.^{4–5} There is growing evidence of the importance

of identifying and addressing inflammatory flare episodes as they may contribute substantially to worsening cardiovascular comorbidity, joint damage and other long-term outcomes.⁶ Although flares can occur unexpectedly, the risk of flare increases when RA treatments are tapered or withdrawn.⁷ In clinical trials, early identification of clinically important RA worsening would signal the need for (re)initiation of therapy. In clinical practice, early identification and resolution of flares would reduce the risks associated with persistently active disease to improve long-term outcomes.

Thus, there is a need for criteria and tools that can be used to reliably identify and quantify RA flares that represent clinically important worsening. Several methods have been proposed including a priori specified increases in disease activity scores (DASs),^{8,9} but there is little consensus. We are unaware of any reports that evaluate the reliability of flare identification which represents clinically important worsening by comparing different perspectives (eg, patients, treating rheumatologists, use of DAS worsening criteria). Similarly, a validated method to quantify flares remains elusive.

We have previously described our pathway to create a consensus-based definition of RA flares and identify the domains essential to include in any measure of flare. In brief, the Outcome Measures in Rheumatology Clinical Trials (OMERACT) RA Flare Group defined RA flares as episodes of increased RA disease activity accompanied by worsening symptoms, functional impacts, and clinical indicators of sufficient magnitude and duration to place individuals at greater risk of joint damage and poorer outcomes when left untreated. Our foundational qualitative and quantitative work with patients, clinicians and other scientists identified essential domains that could be used to measure flare severity.^{4,10} At OMERACT 2012, our RA Flare Core Domain Set was ratified by 200 + OMERACT attendees.¹¹ The RA Flare Core Set included the American College of Rheumatology (ACR) RA core set¹² (patient and physician assessment of disease, tender and swollen joints, acute-phase reactants, physical function, pain) and added fatigue, stiffness and participation. Exploration of self-management as a contextual factor¹³ and other research domains (systemic features, coping, sleep and emotional distress) was also endorsed.

The next steps needed to create a reliable tool to measure RA flare are to (1) identify candidate items to measure each RA flare domain; (2) evaluate the reliability of flare identification from different perspectives; and (3) assess the construct validity of our candidate flare items assessing each flare domain.¹¹ To identify items, we reviewed conceptual models and existing patient-reported outcomes (PROs)¹⁴ and collaborated with members from the International Classification of Function and Health framework.^{15,16} Here, we present initial evidence of the reliability of flare identification and the construct validity of the candidate OMERACT RA flare items.

METHODS

Study participants

Data are from a subset of patients with RA seen over the first 2 years of follow-up in the Canadian Early Arthritis Cohort (CATCH) study, a prospective observational study of patients with early RA recruited at 19 centres across Canada initiated in January 2007.¹⁷ CATCH patients are followed every 3 months in year 1, and every 6 months in year 2 using a standardised protocol. Treatment generally follows Canadian guidelines for RA management.^{18,19} At each visit, patients complete validated measures of RA symptoms and function, the treating rheumatologist performs a physical examination to assess disease activity, and blood is drawn for analysis at local laboratories. Ethics boards at each centre approved the study, and written informed consent was obtained.

In November 2011, the candidate OMERACT flare questions were added to each visit. Here, we included all patients who met the 1987 ACR or 2010 ACR/European League Against Rheumatism (EULAR) criteria and had completed flare questions at ≥ 2 visits through November 2014. Data from the two most recent visits 3 months apart were used; when not available, data from the two most recent visits 6 months apart were selected (designated as V1 and V2).

Measures

Flare questions

Patients were asked whether their disease had worsened, remained the same or improved in the past week (7-point scale; much worse to much better) and if they were experiencing a flare of their RA (yes/no). Patients who classified themselves as flaring then rated the severity (11-point Numerical Rating Scale (NRS)) and duration (1–3, 4–7, 7–14 and >14 days) of their flare. All respondents then completed the candidate OMERACT RA Flare Core Domain Set items rating pain, physical function, fatigue, stiffness and participation over the past week due to RA using 11-point NRS (see online supplementary figure 1) and indicated tender and swollen joints on a 40-joint homunculus.²⁰

Legacy PROs

Participants also completed the RAND-12,²¹ Rheumatoid Arthritis Disease Activity Index (RADAI)²² and Work Productivity and Activity Impairment-Rheumatoid Arthritis (WPAI-RA) Questionnaire.²³

RA Indicators

ACR core set measures were recorded, including pain (10 cm Visual Analogue Scale (VAS)), physical function (Health Assessment Questionnaire-Disability Index (HAQ-DI)),²⁴ patient and MD global assessments, MD tender and swollen joint counts (0–28), erythrocyte sedimentation rate (ESR) and C reactive protein (CRP); a DAS28 was calculated.

Classification of flare status reflecting clinically important worsening

Patient definition

Individuals who answered 'yes' to the question 'Are you having a flare of your RA at this time?' were classified as flaring.

MD definition

Physicians were asked: 'Do you think your patient is in a flare today?' (0 not in flare; 10 severe flare). Receiver operating curves were examined to establish the cut point which best reflected endorsement of flare (see online supplementary figures 2A–C).

DAS28 definition

As compared with the first of selected paired visits (V1), DAS28 worsening at the second visit (V2; ≥ 1.2 or ≥ 0.6 units if DAS28 at V1 was ≥ 3.2) was used to identify flares consistent with a definition used in recent studies.^{8 25 26}

Statistical analysis

Distributions were examined, and descriptive statistics were calculated. As data were missing for MD flare and DAS28 measures on some individuals, we used *t* tests to compare sociodemographic and RA characteristics between groups with and without complete data.

Reliability

Concordance between patient, MD and DAS28 identification of flares at V2 was assessed using the agreement coefficient.²⁷ We hypothesised there would be moderate–high agreement (using Cohen's criteria)²⁸ of flare status between patients, MDs and DAS28.

Construct validation

To examine convergent validity, we used Spearman's correlation coefficient to estimate the degree to which scores from flare domain questions correlated with each other, 'legacy' PROs measuring similar domains, and MD and patient-reported joint counts at V1. We hypothesised there would be moderate ($r's \geq 0.30$) to high correlations ($r's \geq 0.50$) between legacy PROs for flare questions assessing pain, function, participation, fatigue and stiffness; and weak ($r's \geq 0.10$) to moderate correlations between pain and MD joint counts. We selected patients with good control of their RA at the V1 (ie, DAS28 < 3.2), and calculated the difference (95% CIs) in mean change scores at V2 in flare item scores, disease activity indicators and patient ratings of RA activity. We hypothesised that as compared with those not in flare at V2, flaring patients would have significantly higher scores on flare items and rate their RA as worse/much worse. Consequential validity was examined by evaluating the proportion of patients with treatment reductions at V1 and rheumatologist intention to escalate therapy at V2. Analyses were performed using SAS V9.3 (SAS Institute); *p* values < 0.05 were considered statistically significant and all tests were two sided.

RESULTS

These analyses include data from 849 patients (see figure 1). The sociodemographic and RA characteristics of patients included in these analyses were similar to those who had completed < 2 flare assessments (data not shown).

As shown in table 1, patients were mostly female, white and 57% were in remission or low disease activity (LDA). At V2, MD flare ratings were available for 85% (718/849), and DAS28 scores were available for 61%

Figure 1 Flow diagram for patient selection. CATCH, Canadian Early Arthritis Cohort; RA, rheumatoid arthritis.

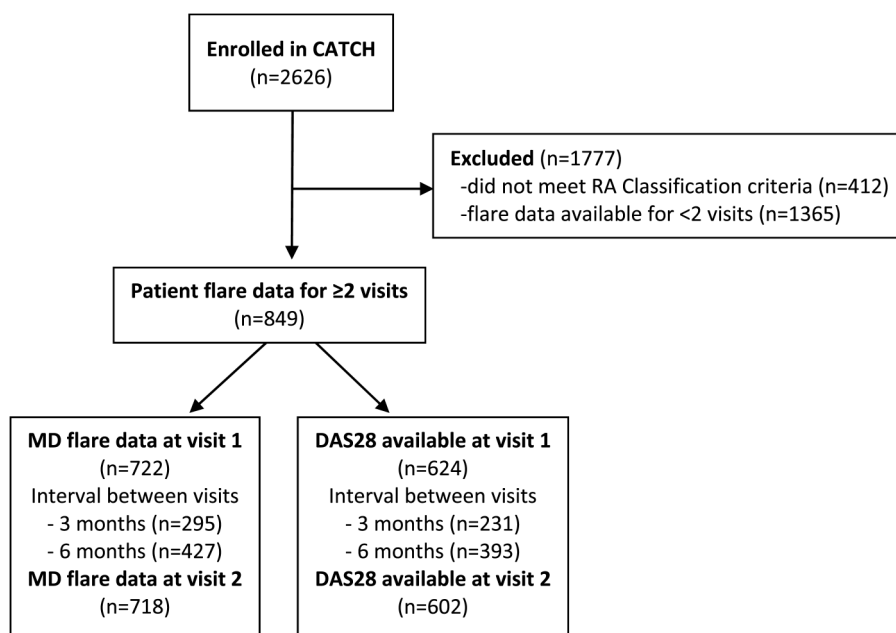


Table 1 Characteristics of participants by flare status

Variable mean (SD) or n (%) unless otherwise stated	All (n=849)	Flare classification					
		Patient (n=849)		MD (n=718)		DAS28 (n=515)	
		Yes (n=201)	No (n=648)	Yes (n=233)	No (n=485)	Yes (n=84)	No (n=431)
Sociodemographic characteristics							
Age (years)	54.3 (15.4)	53.7 (14.3)	54.4 (15.7)	54.9 (15.0)	53.6 (15.4)	55.0 (15.5)	54.7 (14.5)
Female sex	635 (75%)	149 (74%)	486 (75%)	168 (72%)	365 (75%)	61 (73%)	308 (71%)
White	690 (81%)	154 (77%)	536 (83%)	181 (78%)	385 (79%)	69 (82%)	357 (83%)
Education ≤ high school	342 (40%)	95 (47%)	247 (38%)	104 (45%)	172 (35%)	32 (38%)	174 (40%)
Currently smoking ≤	140 (17%)	39 (20%)	101 (16%)	37 (16%)	81 (17%)	17 (20%)	60 (14%)
2 years							
Duration in study (months; median IQR)	18 (15)	18 (15)	18 (15)	12 (18)	18 (12)	21 (12)	18 (15)
RA characteristics							
DAS28	3.2 (1.6)	3.7 (1.7)	3.0 (1.5)	3.9 (1.7)	2.7 (1.4)	2.8 (1.2)	3.2 (1.7)
Remission	277 (44%)	44 (30%)	233 (49%)	55 (30%)	182 (51%)	40 (48%)	199 (46%)
Low	83 (13%)	14 (10%)	69 (14%)	16 (9%)	59 (17%)	20 (24%)	49 (11%)
Moderate	182 (29%)	57 (39%)	125 (26%)	68 (37%)	91 (26%)	21 (25%)	120 (28%)
High	82 (13%)	30 (21%)	52 (11%)	44 (24%)	23 (6%)	3 (4%)	63 (15%)
MD tender joints (28)	3.2 (5.0)	4.9 (5.8)	2.7 (4.6)	5.4 (6.2)	2.2 (3.9)	2.1 (3.7)	3.7 (5.6)
MD swollen joints (28)	2.2 (4.1)	3.2 (4.7)	1.9 (3.9)	4.1 (5.5)	1.2 (2.6)	1.5 (3.6)	2.7 (4.8)
ESR (mm/h)	17.2 (17.4)	18.4 (18.1)	16.8 (17.2)	21.0 (20.0)	15.4 (16.1)	19.3 (20.6)	16.7 (16.7)
CRP (mg/L)	6.5 (10.5)	8.3 (12.8)	6.0 (9.6)	8.6 (11.8)	5.3 (8.8)	7.7 (11.2)	6.1 (9.0)
HAQ (0–3)	0.51 (0.60)	0.75 (0.67)	0.44 (0.55)	0.66 (0.63)	0.42 (0.56)	0.47 (0.59)	0.49 (0.60)
MD global (10 cm VAS)	1.9 (2.5)	2.8 (2.9)	1.6 (2.3)	3.2 (2.9)	1.2 (1.9)	1.3 (1.8)	2.0 (2.6)
Patient global (10 cm VAS)	3.4 (2.9)	4.6 (3.1)	3.0 (2.8)	4.2 (3.0)	2.9 (2.8)	2.7 (2.7)	3.4 (3.0)
Pain (10 cm VAS)	3.2 (2.8)	4.3 (3.0)	2.8 (2.6)	4.0 (2.9)	2.7 (2.7)	2.8 (2.7)	3.2 (2.9)

Sociodemographic characteristics are at enrolment; RA characteristics are at first of paired visits. At second visit, 718 patients had a second MD rating of flare and 515 had a DAS28-ESR at both visits; missing DAS28 values are due to missing ESR. CRP, C reactive protein; DAS, Disease Activity Score; ESR, erythrocyte sedimentation rate; HAQ, Health Assessment Questionnaire; RA, rheumatoid arthritis; VAS, Visual Analogue Scale.

(515/849 – the remaining 39% had missing ESR). Groups that were missing MD flare ratings or DAS28 scores did not differ from those with these data available on sociodemographic or RA characteristics, suggesting data could be treated as missing at random (see online supplementary table S1 for additional characteristics).

Reliability: agreement of flare status

At the second visit, 201/849 (24%) of patients classified themselves as flaring; MDs rated 233/718 (32%) and DAS classified 84/515 (16%) as flaring. Agreement of flare status varied based on disease activity level at the second visit. In patients previously in remission, agreement was high (κ 's \geq 0.73) between patients and physicians and patients and DAS28 (table 2). For patients in LDA, agreement was moderate–strong between patients and physicians and patients and DAS (κ 's=0.44–0.63), but agreement was low for those in moderate–high disease activity (κ 's=0.17–0.35).

Construct validity of flare domains

At V2, OMERACT RA flare domain questions correlated highly with each other (r 's 0.70–0.90; see online supplementary table S2) and with patient global (r 's

0.66–0.88), and moderately with patient joint counts (r 's 0.39–0.62; table 3).

Convergent validity

Across the three flare definitions used, correlations of OMERACT RA flare domain questions with other PROs assessing similar constructs were high, with most >0.60 (see online supplementary table S2). Low–moderate correlations were observed between pain and MD/patient tender joint counts in flaring patients (r 's 0.29–0.48).

Known groups

At V2, mean scores for flare questions, other PROs and RA clinical indicators were significantly higher in patients who self-identified as flaring (table 4). Differences between groups were highest when patients and MD rated the patient as flaring. Patients who self-identified as flaring or also were rated by their doctors as flaring were much more likely ($p<0.0001$) to rate their RA as worse or much worse since the previous visit.

Consequential validity

Similarly, patient flare severity ratings and change in DAS28 scores at V2 were highest when patients and MDs agreed the patient was flaring (p 's=0.028 and <0.001 ,

Table 2 Agreement of flare status between patient, MD and DAS28 worsening criteria by disease activity level

Flare classification (Visit 2)	Flare status	Patient		Agreement		AC1 κ
		Yes	No	Observed	Expected (%)	
MD flare classification (n=718)						
All	Yes	98 (58%)	135 (25%)	512 (71%)	59	0.52
	No	71 (42%)	414 (75%)			
Remission	Yes	5 (23%)	39 (16%)	212 (79%)	78	0.73
	No	17 (77%)	207 (84%)			
Low disease activity	Yes	8 (50%)	17 (29%)	49 (66%)	59	0.44
	No	8 (50%)	41 (71%)			
Moderate disease activity	Yes	37 (69%)	40 (45%)	86 (60%)	49	0.21
	No	17 (31%)	49 (55%)			
High disease activity	Yes	23 (79%)	11 (65%)	29 (63%)	56	0.35
	No	6 (21%)	6 (35%)			
DAS28 criterion flare classification (n=515)						
All	≥1.2/0.6	36 (31%)	48 (12%)	388 (75%)	69	0.64
	<1.2/0.6	79 (69%)	352 (88%)			
Remission	≥1.2/0.6	0 (0%)	1 (0%)	232 (89%)	89	0.88
	<1.2/0.6	27 (100%)	232 (100%)			
Low disease activity	≥1.2/0.6	2 (13%)	7 (12%)	54 (73%)	72	0.63
	<1.2/0.6	13 (87%)	52 (88%)			
Moderate disease activity	≥1.2/0.6	21 (40%)	33 (36%)	80 (56%)	53	0.17
	<1.2/0.6	31 (60%)	59 (64%)			
High disease activity	≥1.2/0.6	13 (62%)	7 (44%)	22 (59%)	51	0.20
	<1.2/0.6	8 (38%)	9 (56%)			

AC, agreement coefficient; DAS, Disease Activity Score.

Table 3 Relationship between items assessing OMERACT RA Flare domains and other measures by flare status*

Domains	Source	All N	Patient		Physician		DAS28†	
			Yes	No	Yes	No	Yes	No
Pain								
How much pain due to RA in past week	HAQ	849	201	648	233	485	84	431
Today's level of pain today	RADAI							
Pain past week	–							
Joint area pain severity (0–48)	RADAI							
Patient tender joint count (28)	Patient							
MD tender joint count (28)	MD							
Physical function								
Disability score (0–3)	HAQ							
Physical function	RAND-12							
Daily activities in past 7 days	WPAI							
Fatigue								
Vitality (RAND-12)	RAND-12							
Unusual fatigue/tiredness past week	–							
Participation								
Role—physical	RAND-12							
Social function	RAND-12							
Productivity while working	WPAI							
RA affecting daily activities	WPAI							
Stiffness								
Morning Joint Stiffness Score	RADAI							

*Spearman correlation coefficients for second visit.

†DAS28 scores <3.2 at second visit required an increase of 1.2 units whereas DAS≥3.2 at second visit required increase of 0.6 units to classify flare.

DAS, Disease Activity Score; HAQ, Health Assessment Questionnaire; OMERACT, Outcome Measures in Rheumatology Clinical Trials; RA, rheumatoid arthritis; RADAI, Rheumatoid Arthritis Disease Activity Index; RAND-12, RAND 12 Health Survey; WPAI, Work Productivity and Activity Impairment.

Table 4 Change in flare question scores and other RA indicators in patients previously in remission/low disease activity

Characteristic (mean SD)	Patient flare*		Difference (95% CI)/ p value	Patient and MD flare N=28	Patient and MD no flare N=219	Difference (95% CI)/p value
	Yes N=58	No N=302				
OMERACT flare domain questions (0–10)						
Pain	1.7 (2.4)	−0.4 (1.8)	2.0 (1.4 to 2.7)	2.3 (2.6)	−0.5 (1.7)	2.7 (1.7 to 3.8)
Stiffness	1.3 (3.0)	−0.3 (1.7)	1.6 (0.8 to 2.4)	2.1 (3.1)	−0.4 (1.7)	2.5 (1.3 to 3.7)
Function	1.6 (2.7)	−0.3 (1.9)	1.9 (1.2 to 2.6)	1.8 (2.7)	−0.4 (1.8)	2.2 (1.1 to 3.2)
Fatigue	0.6 (3.1)	−0.3 (2.1)	0.9 (0.1 to 1.8)	1.6 (3.0)	−0.5 (1.9)	2.1 (0.9 to 3.3)
Participation	1.5 (2.7)	−0.3 (1.8)	1.8 (1.1 to 2.5)	1.8 (2.7)	−0.4 (1.6)	2.2 (1.1 to 3.3)
Patient global	1.9 (3.0)	−0.3 (2.1)	2.2 (1.3 to 3.0)	2.6 (2.8)	−0.3 (2.2)	3.0 (2.1 to 3.9)
Physician measures						
MD global (0–10)	1.2 (2.3)	−0.1 (1.3)	1.3 (0.7 to 2.0)	2.7 (2.1)	−0.3 (1.2)	3.0 (2.1 to 3.8)
MD TJC28*	2.8 (4.5)	0.3 (2.4)	2.4 (1.2 to 3.7)	4.8 (5.5)	0.2 (1.9)	4.6 (2.4 to 6.7)
MD SJC28†	1.6 (4.0)	0.0 (1.4)	1.6 (0.5 to 2.7)	3.4 (5.0)	−0.1 (1.2)	3.6 (1.6 to 5.5)
Acute-phase reactants						
CRP (mg/L)	3.0 (11.0)	0.1 (5.9)	2.9 (−0.2 to 6.1)	5.8 (15.0)	−0.1 (5.6)	5.9 (−0.4 to 12.1)
ESR (mm/h)	4.3 (11.6)	0.6 (8.7)	3.7 (0.4 to 7.0)	6.3 (14.1)	0.4 (8.5)	5.9 (0.2 to 11.5)
RA transition (since previous visit) (n (%))						
Much worse/worse	9 (16%)	5 (2%)	<0.0001	6 (21%)	3 (1%)	<0.0001
Slightly worse/same/ slightly better	41 (71%)	179 (59%)		22 (79%)	128 (58%)	
Better/much better	8 (14%)	118 (39%)		0 (0%)	88 (40%)	

*Tender joint count.

†Swollen joint count.

CRP, C reactive protein; ESR, erythrocyte sedimentation rate; OMERACT, Outcome Measures in Rheumatology Clinical Trials; RA, rheumatoid arthritis.

respectively; see table 5). RA treatment had been reduced or stopped in 36% (10/28), and 61% of rheumatologists reported an intention to increase treatment in response to flare.

DISCUSSION

This is the first study to evaluate the reliability of concurrent flare identification in RA using three methods: patient reports, physician ratings and DAS28 criteria used in recent reduction/withdrawal trials. Agreement about flares using the three methods (patient, MD and DAS) was highest for patients initially in remission and LDA. Our data are also the first to show the construct validity of the five candidate OMERACT RA flare domains (pain, fatigue, stiffness, function and participation). These data support additional evaluation of our patient-centered tool in treatment studies to establish numerical criteria for flare for use in trial and practice.

Agreement about flares varied and was influenced by disease activity level at the initial visit. Others have also reported that the patient and physician do not always agree about RA status, including flares.^{3 29–31} We found that agreement was highest between physician-based measures, either by MD report or DAS worsening, and patients, when disease activity had been in good control at the preceding visit. Conversely, agreement regarding flares was lowest between rheumatologists and DAS28 criteria when patients had previously been in moderate–

high disease activity levels. This may, in part, reflect greater uncertainty in identifying worsening of disease when disease activity is already high. Also, in our study, the mean worsening in DAS28 of flaring patients was nearly two points higher, which is considerably greater than DAS criteria. It is not surprising that the least reliable method of identifying flare was the DAS criteria definition (identifying flare as ‘worsening’ of DAS28 that exceeded measurements error (0.6 units) or twice its value).³² This definition has been used in only a few studies that evaluated RA flares in the context of tapering or withdrawing trials.⁸

Our results support the construct, discriminant and consequential validity of the five candidate items capturing the OMERACT RA Flare Domain Core Set. Flare domain questions correlated highly with PROs measuring similar constructs that are widely used in RA trials and were able to discriminate between patients with/without flare using three definitions. In people who were flaring at the second visit (but not the first), individual flare question scores were significantly higher, and similar increases were observed in other PROs as well as clinical indicators of RA disease activity. In contrast, there was little change in scores of non-flaring patients over these visits. Finally, in flaring patients, 61% of rheumatologists indicated they intended to intensify treatment indicating evidence of the consequential validity of flare identification and consistent with sufficient clinical worsening to justify escalation in therapy.

Table 5 Flare characteristics as classified by patients, MD and patient–MD concordant reports in patients with RA who were previously in remission/low disease activity (n=360)

Characteristics (mean (SD) or n (%))	Patient N=58 (16%)	MD N=71 (20%)	Patient and MD N=28 (8%)
Patient flare severity (0–10)	4.4 (2.1)	4.5 (2.6)	5.0 (2.3)
Duration (days)			
1–3	12 (21%)	5 (18%)	5 (18%)
4–7	6 (10%)	4 (14%)	4 (14%)
8–14	12 (21%)	6 (21%)	6 (21%)
>14	28 (48%)	13 (46%)	13 (46%)
Change in DAS28			
DAS28 at time of flare	3.2 (1.4)	3.0 (1.4)	3.9 (1.4)
DAS28 at previous visit	2.1 (0.7)	2.1 (0.6)	2.1 (0.7)
Worsening of DAS28	1.1 (1.4)	0.9 (1.3)	1.8 (1.2)
Change in DMARD# and/or oral steroids use from previous visit			
Reduced (dose or frequency)	22 (38%)	29 (41%)	9 (32%)
Stopped (without escalating or adding another therapy)	19 (33%)	28 (39%)	7 (25%)
Reduced and/or stopped	24 (41%)	34 (48%)	10 (36%)
MD intent to increase treatment*	25 (45%)	37 (53%)	17 (61%)
Proposed treatment change*			
Non-methotrexate DMARDs added	9 (16%)	10 (14%)	7 (25%)
Methotrexate added or increased†	2 (6%)	3 (7%)	1 (7%)
Biologics added/switched (not due to side effect)	2 (3%)	3 (4%)	2 (7%)
Steroids added (PO/IM or IA; not used in prior visit)	7 (12%)	7 (10%)	4 (14%)
NSAIDs added (not used in the prior visit)	3 (5%)	3 (4%)	2 (7%)

DMARDs could include biologic and synthetic DMARDs.

*Reported increase in treatment at second or subsequent visit.

†Dose increased or changed from oral to subcutaneous.

DAS, Disease Activity Score; DMARD, disease-modifying antirheumatic drug; IA, intra-articular; IM, intramuscular; NSAIDs, non-steroidal anti-inflammatory drugs; PO, per os.

The five RA flare items represent patient-reported core domains identified by patients and providers as essential to measure for RA flare beyond the patient global assessment of disease activity.¹⁰ The high rates of agreement identifying flares between patients and physicians are not unexpected. The flare questions were developed to reflect the usual interchange of information between patients and physicians when discussing disease worsening. Concordance was lowest between patients and DAS-rated and MD-rated flare status when patients were in moderate or high disease activity at the first visit. One reason for this may be that the DAS criteria do not directly incorporate any of the specific domains identified by patients and providers as essential to measuring RA flare, with the exception of patient global. While we found that patient global scores were highly correlated with the five RA Flare domain scores, it is unclear whether patient global alone can be used to identify flares related to inflammation. Notably, only two of the five domains, pain and physical function, are captured in the ACR RA core set, and core set measures make up standard composite measures of disease activity derived to measure improvement. RA composite measures alone may not be sensitive or specific to inflammatory flare, underscoring the need for a tool that can reliably identify and quantify RA flares.

Our results showed that agreement about flare status was high between patients and providers; also, the

largest increases in flare domain scores and DAS28 occurred in flaring patients when there was patient/MD concordance if the patient was flaring. Better understanding of the factors related to discordance in flare reports between the patient and doctor assessments warrant additional study. In the interim, however, an important finding of our work for clinicians is recognition that patients can reliably identify significant increases in inflammatory activity, and that most of the time clinicians agree with patients when patients state they are in a flare. This is especially true in patients that had been in remission or LDA at a previous visit.

While doctors and patients may not always agree on flare status, agreement between both increases confidence that RA flares, which truly reflect worsening inflammation, can be reliably detected. A means to reliably identify and precisely quantify inflammatory flares in RA is needed for clinical trials where drug therapy is reduced or withdrawn as well as comparative effectiveness trials. In this large observational 'real-world' trial, rheumatologists classified 32% of patients as flaring, and treatment had been reduced/stopped at the prior visit in 36% of patients confirming that therapeutic change is an important antecedent to flare in many patients. As evidence grows that tight control is essential to improve long-term outcomes,³³ early identification and treatment of flares seem essential to improve long-term outcomes. The ability to quickly and easily identify flares

incorporating the patient perspective of flare could also facilitate more patient-centred care through the rapid identification of patients potentially requiring escalation of therapy between scheduled visits. Early flare identification enabled by patient report may also help patients know when to initiate additional self-management strategies.

Strengths of this study include the prospective collection of flare data in the context of a national observational study of early RA where many clinical and PROs were systematically collected in a standardised manner. Our sample included patients with all levels of disease activity. Foundational work closely followed OMERACT methods for developing new measurement tools and COSMIN criteria to ensure high methodological quality.^{16 34–36} However, there are limitations. We did not a priori provide guidance or a definition of flare for patients and physicians, and the threshold identified by ROC curves to optimise discrimination of flare was low. Providing a standardised definition and querying physicians directly (yes/no) about flare status may increase agreement. Nevertheless, for patients initially under good control, agreement regarding flares was strong.

PRO measures are evolving, especially for fatigue, stiffness and participation.³⁷ The ability to differentiate patients who are flaring or not was at the group level; additional work is needed to identify whether the questions can be used to reliably identify individuals who are flaring. Worsening assessed by patients and providers was at a single point in time; we did not specifically ask if judgements were made in relation to a previous time period (eg, 3 months), an approach being used by others.³ DAS28 criteria reflect changes from the previous visit, either 3 or 6 months prior. We have no information about symptoms, function, self-management (eg, transient use of glucocorticoids or non-steroidal anti-inflammatory drugs), and limited information on potential treatment changes between visits.

These initial results support the usefulness of the OMERACT RA flare questions for identifying flare. Additional evaluation is needed before they can be recommended for widespread use. Work is ongoing by our group to develop a scoring system for the OMERACT flare questions and to evaluate unidimensionality, responsiveness, and to identify clear thresholds that reflect worsening of RA inflammation signalling a potential need to intensify treatment. Flare data are being collected in international observational studies and randomised controlled trials (RCTs) of early and established RA to help establish criteria for symptom intensity and duration necessary to define inflammatory flare, and to develop thresholds for existing disease activity measures. Identifying and understanding the role of self-management strategies and other contextual factors also need to be considered.¹³ Further exploration of discordance between doctors and patients regarding RA flares and evaluation of patient and MD joint counts is also ongoing.

In conclusion, in routine rheumatology care settings, flares in RA representing clinically important worsening of inflammation are common and are often preceded by treatment reductions. Agreement about flare status among patients, treating rheumatologists, and using DAS28 criteria is high, especially for patients previously in remission or LDA. The five questions that represent the OMERACT RA Flare Core Domain Set, where patients are asked to rate their pain, fatigue, stiffness, function and participation have strong evidence of content validity, known groups and consequential validity. Additional work is also ongoing to develop a scoring system and identify the thresholds of change and flare severity that can be used by clinicians and researchers in order to reliably identify flares using the OMERACT RA flare questions.

Author affiliations

¹Department of Rheumatology, Hospital for Special Surgery, Weill Cornell Medical College, New York, New York, USA

²Rebecca McDonald Center for Arthritis & Autoimmune Disease, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada

³Division of Rheumatology, Johns Hopkins School of Medicine, Baltimore, Maryland, USA

⁴Section of Rheumatology, Arthritis Research UK & Health and Care Research Wales CREATE Centre, Cardiff University, Cardiff, UK

⁵Schlosspark Klinik, Charité University Medicine, Berlin, Germany

⁶Musculoskeletal Statistics Unit, Department of Rheumatology, The Parker Institute, Copenhagen University Hospital, Bispebjerg and Frederiksberg, Denmark

⁷Division of Rheumatology, University of California, Los Angeles, Los Angeles, California, USA (Emeritus)

⁸University of Washington, Seattle Wash; University of Florence, Florence, Italy

⁹University of the West of England, Bristol, UK

¹⁰Bone and Joint Decade, Healthy Motivation, Santa Barbara, California, USA

¹¹Department of Rheumatology, University of Sydney, Institute of Bone and Joint Research, Royal North Shore Hospital, St Leonards, New South Wales, Australia

¹²Division of Rheumatology, University of California, Los Angeles, Los Angeles, California, USA

¹³Division of Rheumatology, Université de Sherbrooke, Sherbrooke, Québec, Canada

¹⁴Rheumatic Disease Unit, Institut de Rhumatologie, Montreal, Québec, Canada

¹⁵Arthritis Center, University of Manitoba, Winnipeg, Manitoba, Canada

¹⁶Vancouver Coastal Health Institute, Vancouver, British Columbia, Canada

¹⁷Division of Rheumatology, St. Joseph's Health Care London, University of Western Ontario, London, Ontario, Canada

¹⁸Southlake Regional Health Centre, Newmarket, Ontario, Canada

¹⁹Divisions of Clinical Epidemiology, Rheumatology, and Respiratory Clinical Trials Unit, McGill University, Montreal, Québec, Canada

Acknowledgements The authors thank Laure Gossec, Maarten Boers, Vibeke Strand and the OMERACT Executive Committee for input which has shaped the direction of this work. The Canadian Early Arthritis Cohort (CATCH) investigators also include Majed Khraishi, Memorial University, St. John's Newfoundland; Murray Baron and Ines Colmegna, McGill University, Montreal Quebec; Michel Zimmer, HÔPITAL MAISONNEUVE ROSEMOUNT, Montreal, Quebec; Pooneh Akhavan, Lawrence Rubin, Bindee Kuriya, University of Toronto, Toronto, Ontario; Vandana Ahluwalia, Headwater's Health Center, Orangeville, Ontario; William Bensen and Maggie Larche, McMaster University, Hamilton, Ontario; Lillian Barra, University of Western Ontario, London, Ontario; Bindu Nair, University of Saskatchewan, Manitoba; Christopher Penney, Dianne Mosher, Cheryl Barnabe, Glen Hazlewood, Calgary Health Sciences Center, University of Calgary, Alberta; Hector Arbillaga, Lethbridge, Alberta; Christopher Lyddell, Grande Prairie, Alberta;

Alice Klinkhoff, University of British Columbia, Vancouver, Canada. They also thank Franci Sniderman for management of the CATCH project and Jim Wang, of McDougall Scientific for statistical support.

Collaborators OMERACT Flare Group: Annelies Boonen, Alfons den Broeder, Bruno Fautrel, Francis Guillemin, Anne Lyddiat, James E. May, Pam Montie, Ana-Maria Orbai, Christoph Pohl and Marieke Scholte Voshaar. 20CATCH investigators: Vandana Ahluwalia, Pooneh Akhavan, Murray Baron, William Bensen, Louis Besette, Gilles Boire, Vivian Bykerk, Ines Colmegna, Boulos Haraoui, Carol Hitchon, Shahin Jamal, Edward Keystone, Alice Klinkhoff, Majed Kraishi, Maggie Larche, Chris Lyddell, Bindu Nair, Chris Penney, Janet Pope, Laurence Rubin, Carter Thorne and Michel Zummer.

Contributors VPB, SJB and COB were involved in the conception, design, acquisition, analysis, interpretation, drafting and revisions of the manuscript. DL was responsible for the analysis, interpretation, drafting and revisions of the manuscript. EHC and LM were involved in the study conception, design, interpretation, drafting and revisions of the manuscript. RA, RC, DEF, SH, AL, LM, TW and KV were responsible for the conception, design, interpretation and revisions of the manuscript. VB, JP, GB, CH, SJ, DT, JCT and ECK participated in the conception, acquisition, interpretation and revisions of the manuscript. All authors reviewed and approved the final manuscript.

Competing interests OMERACT is an international rheumatology outcomes methodology group that has received hands-off funding from more than 23 pharmaceutical and clinical research companies over the last 2 years. COB and LM are members of the OMERACT Executive Committee but receive no financial remuneration for their service in this role. We thank UCB, Inc. for providing funding to support translation of the PFQ into 13 languages (Spanish, German, Dutch, French, Portuguese, Danish, Hungarian, Italian, Polish, Romanian, Swedish, Catalan and Russian), including linguistic adaptations for individual countries (eg, French included versions France and Canada) using validated methods for use in an international RA clinical trial. Additional unrestricted funding for the OMERACT RA Flare Group was provided by Pfizer (Germany), Novartis and Actelion. The CATCH study was designed and implemented by the investigators and financially supported initially by Amgen Canada Inc. and Pfizer Canada Inc. via an unrestricted research grant since the inception of CATCH. As of 2011, further support was provided by Hoffmann-LaRoche Ltd., UCB Canada Inc., Bristol-Myers Squibb Canada Co., AbbVie Corporation (formerly Abbott Laboratories Ltd.), Medexus Inc. and Janssen Biotech Inc. (a wholly owned subsidiary of Johnson & Johnson Inc.). COB and SJB were supported in part by a Patient-Centered Outcomes Research Institute (PCORI) Pilot Project Award (11P2-PI000737-01) and a PCORI Improving Methods for Conducting PCOR Award (SC14-1402-10818). All statements in this report, including its findings and conclusions, are solely those of the authors and do not necessarily represent the views of PCORI, its Board of Governors or Methodology Committee. VPB is supported by the Cedar Hill Foundation, New York, NY and by NIH grant (1UH2AR067691). RC/The Parker Institute is supported by grants from the Oak Foundation.

Ethics approval Central Ethics Committee at each institution where study was performed.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

- Hewlett S, Sanderson T, May J, *et al.* 'I'm hurting, I want to kill myself': rheumatoid arthritis flare is more than a high joint count--an international patient perspective on flare where medical help is sought. *Rheumatology (Oxford)* 2012;51:69–76.
- Flurey CA, Morris M, Richards P, *et al.* It's like a juggling act: rheumatoid arthritis patient perspectives on daily life and flare while on current treatment regimes. *Rheumatology (Oxford)* 2014;53:696–703.
- Berthelot JM, De Bandt M, Morel J, *et al.* A tool to identify recent or present rheumatoid arthritis flare from both patient and physician perspectives: the 'FLARE' instrument. *Ann Rheum Dis* 2012;71:1110–16.
- Bingham CO III, Alten R, Bartlett SJ, *et al.* Identifying preliminary domains to detect and measure rheumatoid arthritis flares: report of the OMERACT 10 RA Flare Workshop. *J Rheumatol* 2011;38:1751–8.
- Bingham CO III, Pohl C, Woodworth TG, *et al.* Developing a standardized definition for disease "flare" in rheumatoid arthritis (OMERACT 9 Special Interest Group). *J Rheumatol* 2009;36:2335–41.
- Myasoedova E, Chandran A, Ilhan B, *et al.* The role of rheumatoid arthritis (RA) flare and cumulative burden of RA severity in the risk of cardiovascular disease. *Ann Rheum Dis* 2016;75:560–5.
- Alten R, Pohl C, Choy EH, *et al.* Developing a construct to evaluate flares in rheumatoid arthritis: a conceptual report of the OMERACT RA Flare Definition Working Group. *J Rheumatol* 2011;38:1745–50.
- van der Maas A, Lie E, Christensen R, *et al.* Construct and criterion validity of several proposed DAS28-based rheumatoid arthritis flare criteria: an OMERACT cohort validation study. *Ann Rheum Dis* 2013;72:1800–5.
- Lie E, Woodworth TG, Christensen R, *et al.* Validation of OMERACT preliminary rheumatoid arthritis flare domains in the NOR-DMARD study. *Annals of the rheumatic diseases* 2013;73:1781–7.
- Bartlett SJ, Hewlett S, Bingham CO III, *et al.* Identifying core domains to assess flare in rheumatoid arthritis: an OMERACT international patient and provider combined Delphi consensus. *Ann Rheum Dis* 2012;71:1855–60.
- Bykerk VP, Lie E, Bartlett SJ, *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.
- Felson DT. Choosing a core set of disease activity measures for rheumatoid arthritis clinical trials. *J Rheumatol* 1993;20:531–4.
- Boers M, Kirwan JR, Wells G, *et al.* Developing core outcome measurement sets for clinical trials: OMERACT Filter 2.0. *J Clin Epidemiol* 2014;67:745–53.
- Gossec L, Paternotte S, Aanerud GJ, *et al.* Finalisation and validation of the rheumatoid arthritis impact of disease score, a patient-derived composite measure of impact of rheumatoid arthritis: a EULAR initiative. *Ann Rheum Dis* 2011;70:935–42.
- Escorpizo R, Boers M, Stucki G, *et al.* Examining the similarities and differences of OMERACT core sets using the ICF: first step towards an improved domain specification and development of an item pool to measure functioning and health. *J Rheumatol* 2011;38:1739–44.
- Boers M, Idzerda L, Kirwan JR, *et al.* Toward a generalized framework of core measurement areas in clinical trials: a position paper for OMERACT 11. *J Rheumatol* 2014;41:978–85.
- Bykerk VP, Jamal S, Boire G, *et al.* The Canadian Early Arthritis Cohort (CATCH): patients with new-onset synovitis meeting the 2010 ACR/EULAR classification criteria but not the 1987 ACR classification criteria present with less severe disease activity. *J Rheumatol* 2012;39:2071–80.
- Bykerk VP, Akhavan P, Hazlewood GS, *et al.* Canadian Rheumatology Association recommendations for pharmacological management of rheumatoid arthritis with traditional and biologic disease-modifying antirheumatic drugs. *J Rheumatol* 2012;39:1559–82.
- Bombardier C, Hazlewood GS, Akhavan P, *et al.* Canadian Rheumatology Association recommendations for the pharmacological management of rheumatoid arthritis with traditional and biologic disease-modifying antirheumatic drugs: part II safety. *J Rheumatol* 2012;39:1583–602.
- Choy EH, Khoshaba B, Cooper D, *et al.* Development and validation of a patient-based disease activity score in rheumatoid arthritis that can be used in clinical trials and routine practice. *Arthritis Rheum* 2008;59:192–9.
- Selim AJ, Rogers W, Fleishman JA, *et al.* Updated U.S. population standard for the Veterans RAND 12-item Health Survey (VR-12). *Qual Life Res* 2009;18:43–52.
- Fransen J, Langenegger T, Michel BA, *et al.* Feasibility and validity of the RADAI, a self-administered Rheumatoid Arthritis Disease Activity Index. *Rheumatology (Oxford)* 2000;39:321–7.
- Zhang W, Bansback N, Boonen A, *et al.* Validity of the work productivity and activity impairment questionnaire--general health version in patients with rheumatoid arthritis. *Arthritis Res Ther* 2010;12:R177.
- Fries JF, Spitz P, Kraines RG, *et al.* Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;23:137–45.

25. den Broeder AA, van Herwaarden N, van der Maas A, *et al.* Dose REduction strategy of subcutaneous TNF inhibitors in rheumatoid arthritis: design of a pragmatic randomised non inferiority trial, the DRESS study. *BMC Musculoskelet Disord* 2013;14:299.
26. van der Maas A, Kievit W, van den Bernt BJ, *et al.* Down-titration and discontinuation of infliximab in rheumatoid arthritis patients with stable low disease activity and stable treatment: an observational cohort study. *Ann Rheum Dis* 2012;71:1849–54.
27. Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. *Br J Math Stat Psychol* 2008;61(Pt 1):29–48.
28. Cohen J, Cohen P. *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1983.
29. Nicolau G, Yogui MM, Vallochi TL, *et al.* Sources of discrepancy in patient and physician global assessments of rheumatoid arthritis disease activity. *J Rheumatol* 2004;31:1293–6.
30. Barton JL, Imboden J, Graf J, *et al.* Patient-physician discordance in assessments of global disease severity in rheumatoid arthritis. *Arthritis Care Res (Hoboken)* 2010;62:857–64.
31. Hirsh JM, Boyle DJ, Collier DH, *et al.* Health literacy predicts the discrepancy between patient and provider global assessments of rheumatoid arthritis activity at a public urban rheumatology clinic. *J Rheumatol* 2010;37:961–6.
32. Den Broeder AA, Creemers MC, van Gestel AM, *et al.* Dose titration using the Disease Activity Score (DAS28) in rheumatoid arthritis patients treated with anti-TNF-alpha. *Rheumatology (Oxford)* 2002;41:638–42.
33. Smolen JS, Breedveld FC, Burmester GR, *et al.* Treating rheumatoid arthritis to target: 2014 update of the recommendations of an international task force. *Ann Rheum Dis* 2016;75:3–15.
34. Kirwan JR, Bartlett SJ, Beaton DE, *et al.* Updating the OMERACT Filter: implications for patient-reported outcomes. *J Rheumatol* 2014;41:1011–15.
35. Mokkink LB, Terwee CB, Patrick DL, *et al.* The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737–45.
36. Mokkink LB, Terwee CB, Knol DL, *et al.* Protocol of the COSMIN study: COnsensus-based Standards for the selection of health Measurement INstruments. *BMC Med Res Methodol* 2006;6:2.
37. 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 (Hoboken)* 2011;63(Suppl 11):S263–86.