BRIEF REPORT

Patient-self assessment of flare in rheumatoid arthritis: translation and reliability of the Flare instrument

Thomas Maribo · Annette de Thurah · Kristian Stengaard-Pedersen

Abstract The Flare instrument (FI) is a French self-administrated questionnaire used to identify flares in disease activity in patients with rheumatoid arthritis. In addition to a total score, the FI has two subscales: one relating to joint symptoms and one relating to general symptoms. The objective of this study was to translate and adapt the French FI into Danish and to determine the reliability of the FI in a consecutive cohort of patients with RA. The FI was translated according to international guidelines, tested among 10 patients and 5 health professionals, and adapted. Test-retest reliability was determined by the standard error of the measurement (SEM) and the intra class correlation coefficients (ICC). The FI was administered to 50 patients with rheumatoid arthritis from an outpatient clinic of a university hospital and re-administered after 10 days. The patients had a mean age of 65.3 years (SD 12.0) and mean disease duration of 18.1 years (range 2–47 years). We found an excellent reliability with ICC higher than 0.95 and SEM between 0.44 and 0.63. Best reliability was found in the total FI score. Thus, the results of the present study show that the FI is a feasible and reliable tool for evaluation of flares in patients with rheumatoid arthritis.

T. Maribo (✉)
Rehabilitation Center Marselisborg, Department of Public Health, Section of Social Medicine and Rehabilitation, Aarhus University, MarselisborgCentret, Bygning 1B, P.P.Ørumsgade 11, DK-8000 Aarhus C, Denmark
e-mail: Thomas.maribo@stab.rm.dk

T. Maribo
CFK, Aarhus, Central Denmark Region, Denmark

A. de Thurah · K. Stengaard-Pedersen
Department of Rheumatology, Aarhus University Hospital, Aarhus, Denmark

Introduction

Rheumatoid arthritis (RA) is characterized by fluctuating between inactivity and activity, a so-called flare [1]. Different definitions of such flares exist in RA [2–4]. However, according to the international network OMERACT (Outcome Measures in Rheumatology), a flare measure should be able to indicate a worsening in disease activity state and should also include the patient's perspective [5]. Traditionally, disease activity in RA has been measured using a composite score, i.e., the Disease Activity Score 28 (DAS28) [6]; however, recent years have seen several attempts at developing simpler, more user-friendly, and less time-consuming flare instruments [7, 8].

The French Flare instrument (FI) is a newly developed patient self-assessment questionnaire aimed at detecting changes in disease activity among RA patients in daily clinical practice [8]. It seeks to determine the perspective and perception of a flare among both patients and clinicians, and is based on semi-structured interviews among 99 RA patients and consensus Delphi rounds among 13 rheumatologists. The Flare instrument can be defined as a patient-related outcome measure (PROM) [9], and thus, as for any PROM, the ability of the FI to capture a flare depends on the psychometric strength of the instrument.

We have translated the French version of the FI into Danish, and the present study is the first of two presenting the translation process and the reliability of the Danish FI. In a second publication, we present the construct and criterion validity of the instrument.

Methods

Patients

For this study, patients visiting a large out-patient clinic at the Department of Rheumatology, Aarhus University Hospital, Denmark were invited. Inclusion criteria were diagnosis of RA according to the criteria set by the American College of Rheumatology [10], age older than 18 years, and fluency in speaking Danish. Clinical characteristics are presented in Table 1.

The Flare instrument

The Flare Instrument (FI) is a newly developed self-administered tool for patients with RA to identify disease activity between consultations. It is designed to detect both past and present disease activity among patients with RA in daily clinical practice. FI is a result of 99 semi-structured interviews together with statements from 13 rheumatologists, generated through a Delphi process. It consists of 12 items and each item represents a statement, which is related to disease activity in RA. FI addresses both the patients and the clinician's experience of disease activity [8]. When scoring FI, the patients are asked to enumerate their degree of agreement on a 10-point Likert scale (0=completely agree, 10=completely disagree), higher scores indicating a flare. The FI gives three scores: the total FI score with potentially two subscales: one related to joint symptoms (FI joint) and one related to general symptoms (FI general). The total FI score, FI joint symptoms, and FI general symptoms were calculated as the arithmetic mean of the 12 FI items, the joint symptoms items (item 1 to 6), and the general symptoms items (item 7 to 12). Only questionnaires with no missing data were included in the calculations.

Recently, FI was shown to have high reproducibility in a randomized controlled trail that included 200 French RA

patients with stable disease (unpublished data). In the present study, we aim to describe the translation and reliability of the FI among a consecutive sample of patients with RA visiting a large outpatient clinic at the department of rheumatology at a university hospital in Denmark.

Translation

The adaptation of the FI to a Danish version (FI-D) was performed according to the guidelines by Guillemin et al. [11] and included the following steps: (i) Translation of the FI from French to Danish by two independent qualified translators; (ii) synthesizing the translations, in order to achieve coherence; (iii) face validity test of this version among patients from the outpatient clinic and health professionals in order to test feasibility of the test; and (iv) back-translation of the consensus version by two independent back-translators was approved by the developer of the original questionnaire.

Reliability

Reliability evaluates the degree to which the measurement is free from measurement error [12]. The reliability was evaluated with 10 days between assessments to reduce the risk of (a) recollection of answers and (b) new flares. The FI was sent to the patients and a new questionnaire (retest) was sent 8 days after receiving the first questionnaire, giving patients at least 10 days between test and retest.

Inclusion and dropout is shown in Fig. 1.

Statistical analysis

The sample size followed the general recommendation of having at least 50 subjects in a method comparison study [12, 13].

Descriptive statistics were calculated for all variables. Possible floor and ceiling effects were examined. Such effects were considered to be present if more than 15% of the respondents achieved the highest or the lowest score, respectively

Table 1 Clinical characteristics of 50 RA patients in the FI study

Age in years	65.3 (12.0) [33–82]
Sex, m/f (%)	17/33 (34/66)
Rheuma-factor positive (%)	45 (90.0)
Disease duration in years (median)[a]	13.5 [8–31]
Receiving corticosteroids (%)[b]	6 (12.5%)
Receiving treatment with biological drugs (%)	11 (22%)
DMARD (%)[c]	45 (90 %)
Days between test and retest	12.6 (2.6) [10; 17]

Values are mean (SD) and [range], *n* and (percentage) or median and [interquartile range]

[a] *n*=42, due to missing data

[b] *n*=48, due to missing data

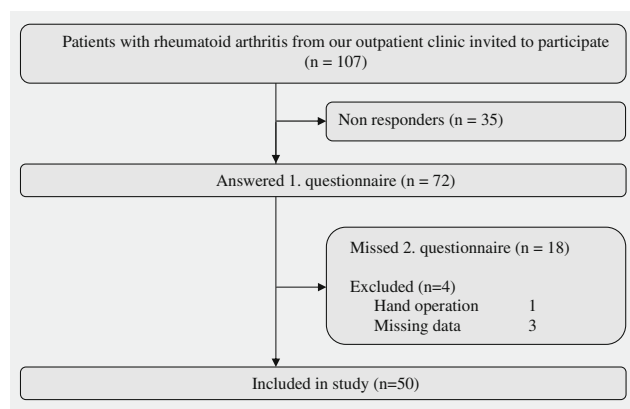[c] Given as mono or combination therapy; Methotrexate 60 %, Salazypyrin: 8 %, other 22 %

Fig. 1 Flow chart for selection of RA patients in the FI study

Table 2 Reliability and agreement parameters for FI total score, FI joint symptoms, and FI general symptoms in 50 RA patients

	Mean, (95 % CI) test Mean, (95 % CI) retest	Difference (95 % CI)	LOA	SEM (95 % CI)	ICC (95 % CI)	MDC
FI total score	2.87 (2.15; 3.58) 2.82 (2.09; 3.55)	0.05 (−0.13; 0.23)	−1.18; 1.27	0.44 (0.37: 0.55)	0.97 (0.95; 0.99)	1.23
FI joint symptoms	3.04 (2.27; 3.82) 3.00 (2.24; 3.76)	−0.04 (−0.21; 0.29)	−1.70; 1.78	0.63 (0.52; 0.78)	0.95 (0.92; 0.98)	1.74
FI general symptoms	2.69 (1.94; 3.43) 2.63 (1.87; 3.40)	0.06 (−0.16; 0.28)	−1.46; 1.57	0.55 (0.46; 0.68)	0.96 (0.94; 0.98)	1.51

FI Flare instrument, *LOA* limits of agreement, *SEM* standard error of measurement, *MDC* minimal detectable change, *ICC* intra class correlation coefficient model 2.1

[14]. Scatters of the differences between test and retest were plotted against the means to indicate if the differences were related to FI score (heteroscedasticity) [15]. Differences between test and retest were calculated, and systematic differences were assessed by paired *t* test. These differences were plotted against the means of the two measurements by Bland–Altman plots with 95 % confidence intervals (CI) and 95 % limits of agreement (LOA). Absolute measurement errors were estimated by calculating the standard error of the measurement (SEM) and converted the SEMs into the minimally detectable change (MDC) ($MDC = 1.96 \times \sqrt{2} \times SEM$). The MDC defines the smallest within-person change that can be interpreted as a "real" change above the measurement error [13]. The intra class correlation coefficients (ICC) model 2.1

with corresponding 95 % CI was used to assess reliability. The ICC can range from 0.0 to 1.0, and according to recommendations, an ICC exceeding 0.90 was considered a sufficient reliability for evaluation of the individual patients [14].

Analysis of internal consistency using Cronbach's α was determined for the total FI score, FI joint symptoms, and FI general symptoms; values above 0.7 were considered acceptable [16].

In order to check whether patients had changed between test and retest sensitivity analyses were performed on item on medication and the global item using weighted kappa with squared weights (K_{w2}) on the medication item.

Statistical analysis was performed with STATA13 software (StataCorp, College station).

Fig. 2 Differences between test and retest plotted against the mean of FI total score, FI joint symptoms, and FI general symptoms
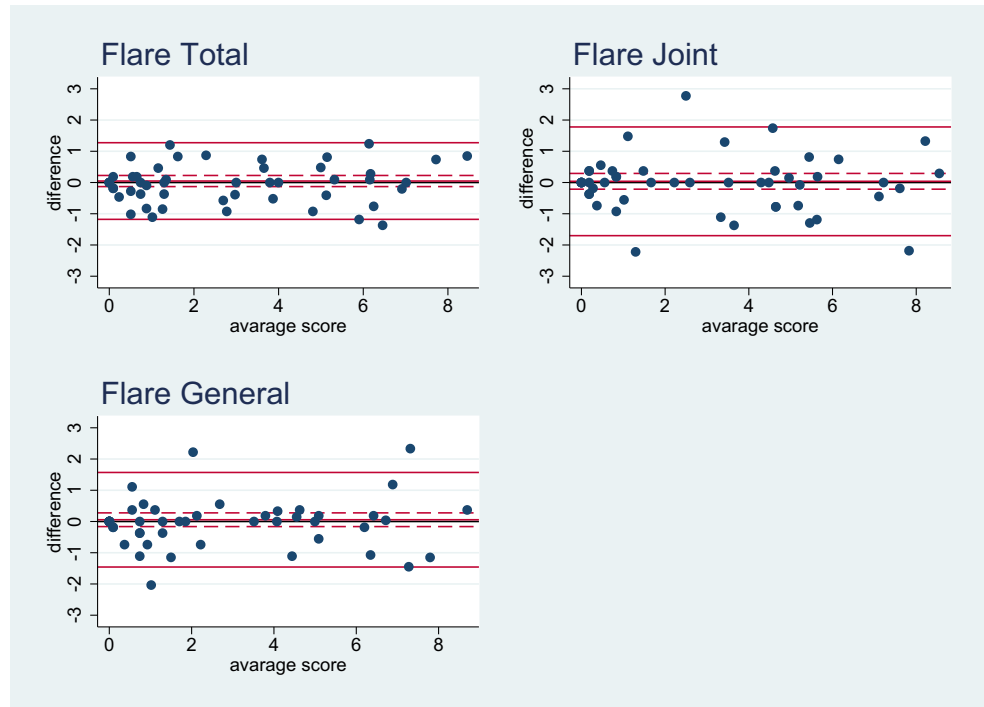
Ethics

All patients gave their written informed consent. The study protocol was approved by the Danish Data Protection Agency (reference number: 1-16-02-577-13), and all procedures were in accordance with the declaration of Helsinki II.

Results

Translation

After the forward translation from French into Danish especially item 1 to 4 did not appear to be mutually exhaustive and exclusive. This became evident through the face validity test, and hence, there was a need for rewording these items. Ten patients from the outpatient clinic and five health professionals (four physicians and one nurse) participated in the test of face validity. The adjusted version was then back translated and this version was approved by the French authors, and this procedure did not result in further corrections. A figure of the translations process is available as extra web material, and the Danish version of FI is available as Appendix 1.

Reliability

Data to the reliability test were collected from March 2013 to August 2013. A total of 107 patients with RA from our outpatient clinic were invited to participate by a postal questionnaire; of these, 72 patients accepted the invitation and returned the first FI. Eighteen patients only completed the first test. The FI is a two-page questionnaire, and three patients completed only the first page and were excluded due to missing data. One patient was excluded due to a hand operation between test and retest. This left 50 patients for the study (Fig. 1).

There were no statistical difference between included patients and the non-responders or the 22 patients excluded (see Fig. 1) in matters of sex, positive rheuma-factor, corticosteroids or disease duration ($p>0.05$). Included patients were significantly older than non-responders (65.3 and 54.3, respectively) ($p<0.01$), but there were no significant difference in age between included patients and the 22 excluded patients ($p=0.29$).

The characteristics of the included patients are presented in Table 1. The mean duration between the two tests was 12.6 days (SD 2.6).

No floor or ceiling effects in total scores and subscales were found except for FI general symptoms, where 20 % of the patients scored 0.

Cronbach's α was 0.96, 0.92, and 0.93 for the total FI score, FI joint symptoms, and FI general symptoms, respectively.

As seen in Table 2, there was no systematic bias between test and retest.

Means for test and retest for, difference, LOA, SEM, and ICC are shown in Table 2.

Differences between test and retest plotted against the mean of the tests are shown in Fig. 2.

Weighted kappa showed excellent agreement on both analgesic and the global item ($_{w2}$ 0.96 and 0.82, respectively).

Discussion

The FI was successfully translated into Danish, and excellent reliability was found for the total FI score and for the two subscales: joint symptoms and general symptoms. Thus, the results of the present study show that the FI is a reliable tool for evaluation of flares in patients with RA.

This was the first reliability study on the FI, and thus, we have no comparable studies.

The translation from French to Danish and backwards was carried out according to international guidelines for cross-cultural adaptation of health-related measures using the procedure introduced by Guillemin et al. [11]. In this process, small adaptations to the original French version were made, including the way arthritis medication was addressed (a field in which the Danish language in general is more imprecise than French) and the form of address in the questionnaire (to a more informal tone).

It is generally recommended that reliability coefficients as a minimum exceed 0.70 to discriminate between groups in clinical trials and 0.90 to evaluate individual patients [14, 17]. In our study, FI and its subscales had an ICC ≥0.95, which indicates that FI and its subscales are well-suited to evaluate disease activity in individual RA patients and also for comparison between groups of patients. The level of the estimated MDC implies that at least 1.23 points are needed to detect a "true" within-person change in the FI total score and a bit higher in FI joint symptoms and FI general symptoms 1.74 and 1.51, respectively.

Cronbach's α was high indicating good internal consistency; as α exceeded 0.90, it could be argued that some items

could be removed [16], but as this is a translation and reliability study, the purpose was not to reduce items.

The present study has some limitations. Participation in this study relied on active enrolment of patients. However, only 46 % of the patients returned both the first and second questionnaire. This dropout was larger than expected [14] and could potentially have introduced selection bias. As we expected, more than 50 % of the patients in our cohort were treated with DMARDs [18], but age was higher and disease duration longer than expected [19]. Because patients with longer disease duration usually have the most stable disease courses, consequently, we may have included patients with a lower prevalence of flares than usually seen in daily clinical practice. This could potentially have caused an overestimation of the ICCs in our study.

A relatively short interval between test and retest was used. A longer time between test and retest has been proposed by some authors [20]. The present time interval was chosen to ensure minimal change in RA, which may have occurred over time. However, such a short interval may be a potential source of bias because the patients may recall previous answers given, and this may affect the objectivity of the tests.

In reliability tests, one of the main questions is whether patients actually had changed between test and retest. As the FI were mailed to the patients, there is no clinical examination to confirm this, but kappa calculations on both the general item and use of analgesic showed excellent reliability indicating no or minor change in clinical symptoms.

In conclusion, we have successfully translated and adapted the FI into the Danish language with the final Danish version demonstrating good psychometric properties with excellent reliability.

Conflict of interest   None

Appendix 1

Flare spørgeskema

De følgende spørgsmål drejer sig om, hvordan leddegigten har præget din hverdag inden for de sidste 3 måneder eller siden sidste konsultation.

Anfør venligst i hvilken grad du er enig eller uenig i nedenstående udsagn ved at sætte kryds i det felt, der svarer bedst til din mening.

Der er ingen rigtige eller forkerte svar. Vi er interesserede i din personlige mening.

Du bedes venligst besvare alle spørgsmål

Inden for de seneste 3 måneder (eller siden den sidste konsultation)

1. har jeg flere dage i træk enten bemærket morgenstivhed i leddene og/eller oplevet en forværring af dette symptom i forhold til tidligere

| Helt UENIG | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Helt ENIG |

2. har jeg flere nætter i træk enten oplevet at vågne på grund af leddegigtsmerter og/eller bemærket en forværring af dette symptom i forhold til tidligere

| Helt UENIG | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Helt ENIG |

3. har jeg flere dage i træk enten haft hævelser i et eller flere led og/eller bemærket en forværring af dette symptom i forhold til tidligere

| Helt UENIG | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Helt ENIG |

4. har jeg flere nætter eller morgener i træk enten haft smerter i flere led og/eller bemærket en forværring af dette symptom i forhold til tidligere

| Helt UENIG | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Helt ENIG |

5. har jeg flere dage i træk oplevet en betydelig forværring af min leddegigt

| Helt UENIG | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Helt ENIG |

6. har jeg flere dage i træk taget mere smertestillende medicin end sædvanligt (gængs smertestillende medicin eller smertestillende gigtmedicin)
Hvis du ikke tager noget smertestillende medicin, så sæt kryds ved « helt uenig »

| Helt UENIG | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Helt ENIG |

7. har jeg flere dage i træk følt mig meget træt

| Helt UENIG | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Helt ENIG |

8. har jeg flere dage i træk følt mig så fysisk begrænset, at jeg "ikke kunne noget som helst"

| Helt UENIG | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Helt ENIG |

9. har jeg flere dage i træk følt mig mere irritabel, end sædvanligt

| Helt UENIG | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Helt ENIG |

10. har jeg flere dage i træk følt mig nedtrykt og har haft svært ved at holde modet oppe

| Helt UENIG | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Helt ENIG |

11. har jeg flere dage i træk haft lyst til at trække mig ind i mig selv eller at isolere mig

| Helt UENIG | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Helt ENIG |

12. har jeg flere dage i træk oplevet at have et større behov for hjælp, end sædvanligt

| Helt UENIG | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Helt ENIG |

Inden for de seneste 3 måneder (eller siden den sidste konsultation):

- vil du da mene at have oplevet forværring af din leddegigt?

☐ Nej

☐ Ja, en enkelt gang

☐ Ja, flere gange

Tak for hjælpen

Fig W1: Translation of the Flare instrument,
a self-administered questionnaire to measure disease activity
in RA patients
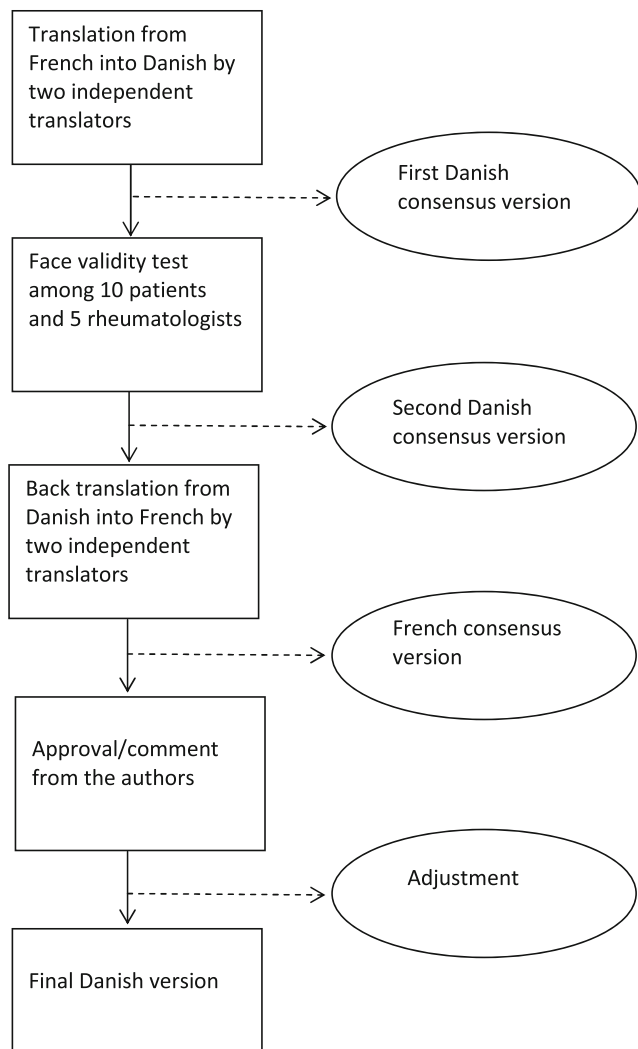
References

1. National Collaborating Centre for Chronic Conditions (UK) (2009) Rheumatoid arthritis: national clinical guideline for management and treatment in adults

2. Rahman MU, Strusberg I, Geusens P et al (2007) Double-blinded infliximab dose escalation in patients with rheumatoid arthritis. Ann Rheum Dis 66:1233–1238

3. Smolen JS, Keystone EC, Emery P et al (2007) Consensus statement on the use of rituximab in patients with rheumatoid arthritis. Ann Rheum Dis 66:143–150

4. Yazici Y, Erkan D, Kulman I, Belostocki K, Harrison MJ (2002) Decreased flares of rheumatoid arthritis during the first year of etanercept treatment: further evidence of clinical effectiveness in the "real world". Ann Rheum Dis 61:638–640

5. Bingham CO III, Pohl C, Woodworth TG (2009) Developing a standardized definition for disease "flare" in rheumatoid arthritis (OMERACT 9 special interest group). J Rheumatol 36:2335–2341

6. Prevoo ML, van 't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL (1995) Modified disease activity scores that include twenty-eight-joint counts. development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. Arthritis Rheum 38:44–48

7. Bykerk VP, Lie E, Bartlett SJ et al (2014) Establishing a core domain set to measure rheumatoid arthritis flares: report of the OMERACT 11 RA flare Workshop. J Rheumatol 41:799–809

8. Berthelot JM, De Bandt M, Morel J et al (2012) A tool to identify recent or present rheumatoid arthritis flare from both patient and physician perspectives: the 'FLARE' instrument. Ann Rheum Dis 71:1110–1116

9. Great Britain. Department of Health (2010) Equity and excellence: liberating the NHS. Stationery Office, Norwich

10. Aletaha D, Neogi T, Silman AJ et al (2010) 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/ European League against rheumatism collaborative initiative. Ann Rheum Dis 69:1580–1588

11. Guillemin F, Bombardier C, Beaton D (1993) Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. J Clin Epidemiol 46:1417–1432

12. Mokkink LB, Terwee CB, Patrick DL et al (2010) The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 19:539–549

13. Terwee CB, Bot SD, de Boer MR et al (2007) Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 60:34–42

14. Fayers PM, Machin D (2007) Quality of life: the assessment, analysis and interpretation of patient-reported outcomes. Wiley, Chichester

15. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1:307–310

16. Streiner DL (2003) Starting at the beginning: an introduction to coefficient alpha and internal consistency. J Pers Assess 80:99–103

17. Hopkins WG (2000) Measures of reliability in sports medicine and science. Sports Med 30:1–15

18. Sokka T, Makinen H (2009) Drug management of early rheumatoid arthritis—2008. Best Pract Res Clin Rheumatol 23:93–102

19. Sokka T, Toloza S, Cutolo M et al (2009) Women, men, and rheumatoid arthritis: analyses of disease activity, disease characteristics, and treatments in the QUEST-RA study. Arthritis Res Ther 11:R7

20. Scholtes VA, Terwee CB, Poolman RW (2011) What makes a measurement instrument valid and reliable? Injury 42:236–240