

Translating Patient Reported Outcome Measures: Methodological issues explored using cognitive interviewing with three rheumatoid arthritis measures in 6 European languages

Hewlett Sarah,¹ Nicklin Joanna,¹ Bode Chistina,² Carmona Loreto,³ Dures Emma,¹ Engelbrecht Matthias,⁴ Hagel Sofia,⁵ Kirwan John,⁶ Molto Anna,^{7,8} Redondo Marta,⁹ Gossec Laure.^{7,8}

¹Dept of Nursing and Midwifery, University of the West of England, Bristol, UK

²Dept of Psychology, Health & Technology, University of Twente, Enschede, The Netherlands

³Instituto de Salud Musculoesquelética, Madrid, Spain

⁴Department of Internal Medicine 3 and Institute of Clinical Immunology, University of Erlangen-Nuremberg, Germany

⁵Dept of Clinical Sciences, Section of Rheumatology, Lund University and Skane University Hospital, Lund Sweden

⁶Academic Rheumatology, University of Bristol, UK

⁷Sorbonne Universités, UPMC Univ Paris 06, Institut Pierre Louis d'Epidémiologie et de Santé Publique, GRC-UPMC 08 (EEMOIS)

⁸AP-HP, Pitié Salpêtrière Hospital, Department of rheumatology, Paris, France

⁹School of Psychology, Universidad Camilo José Cela, Madrid, Spain

Corresponding Author:

Professor Sarah Hewlett
Academic Rheumatology Unit
Bristol Royal Infirmary
Bristol BS2 8HW
Tel: 44 (0)117 342 2903
Fax: 44 (0)117 342 3841
Email: Sarah.Hewlett@uwe.ac.uk

Word count: 2893

Key words:

Patient Reported Outcome Measures; Translation; Cognitive interviewing;
Bristol Rheumatoid Arthritis Fatigue (BRAFF) scales; Rheumatoid Arthritis Impact of Disease scale (RAID)

Running head:

Translating PROMs: Cognitive interviewing

ABSTRACT:

Objectives: Cross-cultural translation of patient-reported outcome measures (PROMs) is a lengthy process, often performed professionally. Cognitive interviewing assesses patient comprehension of PROMs. The objectives were to evaluate the usefulness of cognitive interviewing to assess translations, and compare professional (full) to non-professional (simplified) translations processes.

Methods: A 'full' protocol used for the Bristol RA Fatigue Multi-dimensional Questionnaire and Numerical Rating Scale (BRAf-MDQ, BRAf-NRS) was compared with a 'simplified' protocol used for the RA Impact of Disease scale (RAID). RA patients in UK, France, the Netherlands, Germany, Spain and Sweden completed the PROMS during cognitive interviewing (BRAfs in UK omitted as performed during development). Transcripts were deductively analysed for understanding, information retrieval, judgement, and response options. Usefulness of cognitive interviewing was assessed by the nature of problems identified, and translation processes by percentage of 'consistently problematic items' ($\geq 40\%$ patients/country with similar concerns).

Results: 60 patients participated (72% women). For the BRAfs (full protocol) one problematic item was identified (of 23 items x 5 languages, $1/115=0.9\%$). For the RAID (simplified protocol) two problematic items were identified (of 7 items x 6 languages, $2/42=4.8\%$), of which one was revised (Dutch). Coping questions were problematic in both PROMs.

Conclusions: Conceptual and cultural challenges though rare, were important, as identified by formal evaluation, demonstrating that cognitive interviewing is crucial in PROM translations. Proportionately fewer problematic items were found for the full than the simplified translation procedure, suggesting that whilst both are acceptable, professional PROM translation might be preferable. Coping may be a particularly challenging notion cross-culturally.

Key messages:

- Cognitive interviewing identifies rare but crucial conceptual and cultural challenges in the translation of patient-reported outcome measures.
- Recommended best practice translation methods for patient-related outcome measures can be followed using pre-determined translation protocols.
- Simplified (clinician) and full (professional) translation protocols are effective, but professional translation of patient-reported outcome measures is preferable

INTRODUCTION

In order to provide appropriate treatments it is essential that clinicians are able to measure not only disease process or severity but also the impact on patients, using patient reported outcome measures (PROMs).[1] International studies require the use of PROMs that are adequately translated. However, cross-cultural translation may be complex: evidence from fibromyalgia demonstrates that literal translation of PROMs (even using formal protocols) may not produce versions that are culturally relevant in the target population.[2] Translation processes need to address conceptual equivalence across cultures (eg the meaning of fatigue); item equivalence (eg relevance of climbing stairs in a culture of single-storey dwellings); semantic equivalence (eg “feeling blue”); operational equivalence (eg understanding of Visual Analogue Scales); and measurement (psychometric) equivalence.[3]

Best practice principles for PROM translation are iterative rounds of forward and backward translations, preceded by conceptual elaboration with the source (original) PROM developers, as agreed by the International Society for Pharmaco-economics and Outcomes Research (ISPOR).[4] However, these formal processes can be lengthy and costly and are usually undertaken by professional PROM translation companies. It is not uncommon for researchers or clinicians wishing to utilize a PROM in their own country to translate it themselves, with or without the approval or collaboration of the PROM developer. Reports of what was done are rarely published. A practical method would be for the PROM developers to undertake the translation process themselves using a formal but simplified protocol.

Rheumatoid arthritis (RA) is a systemic, inflammatory condition causing synovitis in multiple joints, leading to pain, fatigue, and disability with accompanying emotional, social, financial and societal burden.[5-8] Three PROMS that capture impact on the patient are the Bristol RA Fatigue Multi-Dimensional Questionnaire and Numerical Rating Scale (BRAf-MDQ, BRAf-NRS)[9-11] and the RA Impact of Disease scale (RAID).[12-14] The BRAFs were translated into 35 languages by a professional PROM company using qualified PROM translators and native bi-lingual speakers (all professional personnel), following the full protocol of ISPOR principles including concept elaboration with the developers.[4] The RAID was conceptualized in English with patients and clinicians from 10 countries, with simultaneous translation into 12 EU languages by the clinician developers (rheumatologists) and native bi-lingual speakers (lay), in a formal but simplified protocol.[15,16] Both protocols (*online supplement A*) included forward and backward translations but to differing degrees.

Patient evaluation is recommended during translation because cultural nuances may differ between countries.[4] Both the full and simplified translation protocols for BRAf and RAID included evaluation by 5 patients/country, although this was not audio-taped or formally analysed (*online*

supplement A). Cognitive interviewing is a formal research methodology, where participants are prompted to “think aloud” as they complete the PROM, with interviews taped and rigorously analysed for understanding, retrieval of information, judgement and response options.[17] By capturing the patient’s cognitive processing prospectively as they complete the questionnaire, the researcher can determine if patients have problems interpreting questions and response options in the way intended, enabling phraseology of PROMs to be clarified.[17-19] Our objective was to use these BRAF and RAID translations to a) evaluate whether the recommended use of cognitive interviewing with patients helps identify any important translation issues, and b) to compare the full versus simplified translation processes.

METHODS

Study design: Cross-sectional study in France, Germany, the Netherlands, Spain, Sweden and UK during 2012-13.

Patients: Patients who attended hospital rheumatology outpatient appointments were invited to participate if they were aged ≥ 18 years with a diagnosis of RA.[20] Patients were purposively selected using a sampling frame to reflect a range of characteristics that might influence PROM completion: age, education, disease duration, gender, disability,[21] patient global opinion of disease activity (Visual Analogue Scale (VAS) 0-10) and fatigue (NRS 0-10).

PROMs: The BRAFs were developed and validated in collaboration with patients to measure RA fatigue.[9-11] The multi-dimensional questionnaire (BRAF-MDQ) contains 20 questions and provides a global fatigue score, plus four distinct subscales (Physical Fatigue, Living with Fatigue, Emotional Fatigue and Cognitive Fatigue), while a trio of single Numerical Rating Scales (BRAF-NRS) measure Fatigue Severity, Coping and Effect. The RAID was developed and validated as a EULAR initiative as a cross-cultural PROM capturing the impact of RA.[12-14] RAID contains 7 NRS (pain, functional disability, fatigue, sleep, coping, emotional and physical well-being), each weighted for importance using standardized weightings derived from a patient survey, yielding a single, composite score. The 35 BRAF translations and 12 RAID translations are freely available: BRAFs can be downloaded from

<http://www1.uwe.ac.uk/hls/research/healthandclinicalresearch/researchareas/longtermconditions/fatiguescales.aspx> and RAID from

http://www.eular.org/index.cfm?framePage=/st_com_clinical_tools.cfm.

Data collection: The interviewers were trained in and practiced cognitive interviewing in a 3 hour session. All 6 interviewers (JN, CB, ME, SHagel, MA, RM) were experienced researchers or clinicians (5 female) and two had prior cognitive interviewing experience. One-to-one cognitive interviews were held in local departments (10 participants/country). As the UK BRAF had undergone extensive cognitive interviewing during development,[9] UK interviews were only performed on the RAID. Each patient completed the PROMs whilst verbalising their understanding

of what each question was asking, and how they reached their answer, prompted by the interviewer where necessary. When a patient experienced problems with a question or response option, the interviewer asked them to clarify their concerns and invited them to rephrase the question or response option.[17-19] Interviews were recorded, transcribed in their native language and anonymised.

Analysis: The local interviewers deductively analysed each question under the recommended categories of understanding, information retrieval (Recalling the necessary information), judgement (What information was considered?), and response options (Were these appropriate?).[17-19] Concerns or confusion identified were extracted onto a standard report form, and reports/coding checked by a second researcher centrally (JN). The steering team considered it unacceptable if more than one-third of patients in any one country had consistent concerns, thus 'consistently problematic' was defined as $\geq 4/10$ patients with consistent concerns. If emergent data suggested important problems, questions would be considered for rewording and re-testing. To compare the two translation methods the percentage of 'consistently problematic' items were calculated: BRAFs 115 possible items (23 items x 5 languages) and RAID 42 items (7 x 6 languages).

Ethics: UK ethics approval for the study was obtained (London, City Road and Hampstead Ethics Committee, 12/LO/1198), then the Principal Investigator (PI) in each country obtained local approvals as required. Patients completed written, informed consent as required by the Declaration of Helsinki.

RESULTS

10 patients per country participated, comprising 72% female with a range of disease activity and severity and importantly, varied educational level (Table 1).

(Put Table 1 here)

Analysis demonstrated that the cognitive processes patients made hinged very precisely on the detailed phrasing of the questions. For example Q5 in the RAID specifies 'physical well-being' while Q6 specifies 'emotional well-being':

[UK Pt 701]: "I understand 'physical well-being'. I'm just telling my brain it's physical well-being, because that's how you feel and how things are acting, as opposed to maybe fatigue or..."

[Int] "Right, so you're sort of clarifying to yourself?"

[UK 701]: "Yeah, physical means actually doing, rather than actually feeling" (UK 701)

[UK 701]: "Considering your arthritis overall, how would you rate your level of emotional well-being during the last week?' Um probably 4, that would go down because I felt so relaxed, I didn't have to worry about work or anything" [...] "Um 'cos although it says 'considering your arthritis overall', emotional is still how you feel, isn't it"

Deductive analysis identified a range of concerns that patients raised (Table 2, shown by deductive categories). Apart from 3 questions, wherever the total number of concerns were raised by $\geq 4/10$

patients/country, these were diverse, not consistent. In France, for example RAID Q1 had 4 concerns overall, but two were issues with recall, and two related to response options.

(Put Table 2 here)

BRAF consistently problematic questions: The response option of one of the 23 BRAF items was identified as ‘consistently problematic’ ($1/115 = 0.9\%$). The BRAF-NRS Coping item was scored by 7/10 Dutch patients in a way that conflicted with their verbal comments, suggesting that the response option layout was confusing (Table 3 presents examples of different manifestations of this problem). The BRAF-NRS Coping response options have anchors ‘Not at all well’ (0) on the left, to ‘Very well’ (10) on the right. None of the other 4 countries raised concerns about scoring direction (operational equivalence). A potential revision was created with anchors reversed.

(Put Table 3 here)

RAID consistently problematic questions: Two of the RAID’s 7 questions were consistently problematic ($2/42 = 4.8\%$). In the original source version (English), RAID Q5 asks about physical well-being: *Considering your arthritis overall, how would you rate your level of physical well-being during the past week?* Four of the 10 Dutch patients reported this as confusing (analysis category Understanding). In comparison with the original English Q5, the Dutch version alone contained an additional phrase in brackets: *physical well-being (without taking pain, inflammation and fatigue into account)*. Examples shown in Table 4 demonstrate that some patients understood this to mean they should rate life as if they did not have RA at all (302), others purely rated physical disability or non-RA pain (303) or rated the pain, inflammation and fatigue they were instructed to ignore (307). This issue appears to be a literal Dutch translation error and the phrase in brackets was therefore removed.

(Put Table 4 here)

RAID Q7 was found confusing by 6/10 Spanish patients, under the analysis category of Understanding in relation to the term ‘coping’ (conceptual equivalence). The original English question was phrased: *“Considering your arthritis overall, how well did you cope (manage, deal, make do) with your disease during the last week?”* The Spanish version was translated as *“afrentar (lidiar, sobrellevar, hacer pasar)”* or cope with, deal with, overcome, make go away (Table 5). Patients commented that *“afrentar”* was a very formal phrase for coping, related to adjustment; they disliked the phrase *“Hacer pasar”* (to make it go away), believing this was not necessarily within their control.

(Put Table 5 here)

A potential revision of the Spanish RAID Q7 using Spanish terms that were more colloquial, including some suggested by these CI participants was piloted in another 5 cognitive interviews with fresh patients, along with the original. Participants with higher education levels or long disease duration understood both versions. However, those with lower education levels struggled with “*sobrellevado*” and “*lidiado*”. This led to a further potential revision, to be tested in a future study: “*Teniendo en cuenta su artritis en general, ¿Qué tal se ha arreglado (ha llevado, afrontado, se ha apañado) con su enfermedad durante la última semana?*” (Considering your arthritis overall, how well did you sort out (deal with, cope with, manage) your disease during the last week?)

DISCUSSION

This study demonstrates the value of cognitive interviewing during translation of PROMs. Formal evaluation identified how closely patients pay attention to specific wording of an item (eg physical versus emotional well-being), highlighting the care with which developers must formulate their questions. This was demonstrated during development of the original UK BRAFs, where formal cognitive interviewing with 15 patients resulted in small but crucial changes to wording.[9] Similarly cognitive interviewing applied to these translated PROMs identified only a few but they are potentially important. The addition of extra descriptors for physical well-being (Dutch RAID Q2) by the clinician translator was intended to help patients understand the global concept, but it caused confusion by offering alternatives to consider. For this reason, and to ensure cross-cultural comparability, the added phrase has been removed.

During development of the BRAF and RAID, the developers acknowledged challenges in constructing their ‘coping’ items.[9,14] This current study also suggests that the concept of coping may be difficult to capture. Coping as presented in the RAID (Q7) includes several terms.[12-14] Some Spanish patients considered these difficult to interpret (semantic equivalence). The conceptual differences between ‘coping’ and ‘management’ were highlighted during the BRAF development: some patients conceptualized ‘coping’ as the emotional strategies to deal with RA but ‘management’ as practical strategies; others considered these concepts to be the opposite way round, or interchangeable.[9] The challenge of *terminology* for coping may be due to the elusive nature of the underlying concept, or to the difficulty of finding terms to capture this concept in different cultures.

The challenge of *rating* coping is demonstrated in the BRAF-NRS Coping, where the direction of scoring led to misinterpretation (operational equivalence). As with all the RAID items, the first two BRAF-NRS (Severity, Effect) are scored using traditional approaches (worse status on the right). The reverse direction of the BRAF-NRS Coping anchors (worse status on the left) was

recommended by patients in cognitive interviewing during development [22] but is here called into question by Dutch patients. Furthermore, the different rating direction makes it harder to intuitively compare scores with the other BRAF-NRS where high scores reflect worse health. The final arbiter of this difference in patient preference must lie in unravelling the validation strengths of the different direction versions. Thus the significance (or otherwise) of the current study's findings relating to coping terminology (Spanish RAID Q7) and rating (BRAF-NRS coping) are being explored in a large ongoing study of construct validity for the original and revised wording.

This study suggests that whilst both full (professional) and simplified (clinician/developer) translation protocols are acceptable, full protocols might be preferable, based on the percentage of consistently problematic items identified (0.9% vs 4.8%). The major differences between the procedures lay in using the initial concept elaboration (defining the intended conceptual meaning of questions) as a reference, the use of independent translators and in-country investigators who were qualified PROM translators, continual involvement of the PROM developers, and harmonization across multiple languages in combined meetings. The rationale for using multiple translators at forward and backward translation in a full protocol, followed by harmonization meetings with the developers, is to ensure that the proposed phraseology captures the original concepts as intended by the developers. Harmonization is particularly important where some cultures do not have a direct term for a concept, for example during the translation of the first 25 BRAF versions, the translations of being 'embarrassed' by fatigue initially included 'ashamed,' 'uncomfortable,' 'awkward,' and 'bewildered' (Canadian French, Dutch, Japanese, Korean, Russian and US Spanish) prior to harmonization.[23]

It is now well accepted that development of PROMs should adhere to recommended methodology to ensure relevance, and identify the underlying concepts and appropriate phraseology.[24,25] They should be grounded in patient experience through collaboration with patients,[26,27] and evaluated for face, content, construct and criterion validity, reliability and sensitivity to change.[24,25] Whilst cross-cultural translation of PROMs should also follow recommended standards[18], this is often not done. For example a systematic review of health-related PROMs translated into Turkish, Arab, and Surinamese found that only 50% followed recommended procedures.[28] In the literature, a recommended translation protocol is often described in the methods but the findings reported in a single sentence, with the paper concentrating on the psychometric testing.[29,30]. Ideally, equal weighting to both cross-cultural translation and psychometric testing should be given as in the Danish translation of RA Self-Efficacy Scale.[31] Cross-cultural translation and subsequent psychometric testing in the new language should be an iterative process, with each informing the other.[32] The translation of the English Beck Depression Inventory into Welsh followed recommended ISPOR methods[4] but subsequent psychometric testing identified poorer construct validity, suggesting that early psychometric testing should be

added to the ISPOR guidelines.[33] Recommended practice is for cognitive debriefing with patients during translation but no formal methodology is referenced.[4] Formal cognitive interviewing methodology[17-19] as used here, should perhaps replace cognitive debriefing.

This study was conducted in 6 countries with a relatively homogenous European culture, therefore conceptual equivalence (eg the meaning of fatigue) and item equivalence (eg relevance of stairs) were unlikely to be problematic.[3] In more diverse cultures conceptual and item relevance issues are likely to need considerable attention. In India for example, the concept of scales (0-10) is not widespread, and global well-being would likely be evaluated as purely pain, the score for which is sometimes reported by family consensus or 'whatever the doctor considers to be right'.[33] When translating the Health Assessment Questionnaire into Bengali, as a significant proportion of the population do not use a car, that question had to be rephrased as using a rickshaw.[34] Patient participation in the development and translation of PROMs is crucial, recommended as best practice[24-26] and was integral to the development of concepts and items in the BRAFs and RAID.[9,12]. Partnership with patients is perceived as vital and normal practice in these 6 EU countries, but in other cultures would be harder to establish, as patients continue to be viewed primarily as research subjects.[35]

This study would have been more robust with two independent coders per country, the inclusion of more diverse cultures, and more patients with less education (>70% had a degree). The study's strengths include the input of patient research partners, and involvement of the PROM developers who understand the original concepts. Ten cognitive interviews/country was greater than the ISPOR recommendation (5-8)[4] and consistent concerns were pre-defined as raised by $\geq 4/10$ patients. This study was led by the clinician/developers, thus any bias would have been for supporting simplified protocols.

The usefulness of formal evaluation of PROM translations through cognitive interviewing has clearly been shown as crucial in this study. Furthermore, professional full protocols might be preferable to clinician/developer simplified protocols, although those are still acceptable and there is a trade-off to be made between simple, inexpensive translations that might facilitate uptake and more detailed (and therefore expensive) professional translations that might potentially be more accurate. The importance of the coping measurement issues raised in this study is being evaluated by further statistical testing not only of the translations, but also of the original source PROM (a level of detailed review that is rarely performed).

ACKNOWLEDGMENTS AND AFFILIATIONS

This study was funded by a European League Against Rheumatism Grant (EULAR, EP1009). The authors would like to thank the patients for participating, and the wider steering group for their opinions on the study: Dr Fiona Cramp (UK), Professor Mart Van de Laar and Dr Jaap Fransen (the Netherlands), Professor Georg Schett (Germany), Professor Ingemar Petersson (Sweden). Bev Davis (UK, patient research partner) and Gabrielle Von Krause (France, patient research partner). We would like to thank local patient research partners Sabine Stolper (Germany), Lyn Packwood (The Netherlands), Cristina Fernández (Spain) and Anna Schéle (Sweden).

The authors have no competing interests to declare.

Contributions

S Hewlett, JN, CB, LC, EM, JK and LG contributed to study design. JN, CB, ME, S Hagel, MA and MR conducted the interviews and primary analysis, with JN checking coding. Data interpretation and discussions were led by S Hewlett, JN JK and LG, with all other authors contributing. All authors contributed to the paper and approved the final submission.

REFERENCES

- 1 Sanderson T, Hewlett S, Flurey C et al. The impact triad (severity, importance and self-management) as a method of enhancing the measurement of the personal life impact of rheumatic diseases. *J Rheumatol* 2011;38:2:191-4.
- 2 Le Gal M, Mainguy Y, Le Lay K et al. Linguistic validation of six patient-reported outcome instruments into 12 languages for patients with fibromyalgia. *Joint, Bone, Spine* 2010;77:165-70.
- 3 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 UK. Oxford University Press: 2008:29-33.
- 4 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 in Health* 2005;8:94-104.
- 5 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 (Oxford)* 2000;39:603-11.
- 6 Katz P. The impact of rheumatoid arthritis on life activities. *Arthritis Care Res (Hoboken)* 1995;8:4:272-8.
- 7 Scott DL, Lempp HK. Outcomes associated with early rheumatoid arthritis. *Expert Rev Pharmacoecon Outcomes Res* 2006;6:5:495-508.
- 8 Pollard L, Choy EH, Scott DL. The consequences of rheumatoid arthritis: quality of life measures in the individual patient. *Clin Exp Rheumatol* 2005;23:5(Supp39):43-52.
- 9 Nicklin J, Cramp F, Kirwan J, Urban M, Hewlett S. Collaboration with patients in the design of patient reported outcome measures: Capturing the experience of fatigue in rheumatoid arthritis. *Arthritis Care Res (Hoboken)* 2010;62:1552-8.
- 10 Nicklin J, Cramp F, Kirwan J, Urban M, Hewlett S. Measuring fatigue in RA: A cross-sectional study to evaluate the BRAF Multi-Dimensional Questionnaire, Visual Analogue and Numerical Rating Scales. *Arthritis Care Res (Hoboken)* 2010;62:1559-6.
- 11 Dures E, Cramp F, Greenwood R, Kirwan J, Hewlett S. Reliability and sensitivity of the Bristol RA Fatigue Multidimensional Questionnaire (BRAFM-DQ) and BRAF short scales. *Rheumatology (Oxford)* 2013;52:1832-39.
- 12 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.
- 13 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.

- 14 Heiberg T, Austad C, Kvien TK, Uhlig T. 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.
- 15 Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: Literature review and proposed guidelines. *J Clin Epidemiol* 1993;46:1417-32.
- 16 Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 2000;25:3186-91.
- 17 Drennan J. Cognitive interviewing: Verbal data in the design and pretesting of questionnaires. *J Adv Nurs* 2003;42:57-63.
- 18 Collins D. Pretesting survey instruments: An overview of cognitive methods. *Qual Life Res* 2003;12:229-38.
- 19 Tourangeau R. Cognitive aspects of survey measurement and mismeasurement. *Int J Public Opin Res* 2003;15:3-7.
- 20 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.
- 21 Fries J, Spitz P, Kraines G, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;23:2:137-45.
- 22 Nicklin J. The development of scales to measure fatigue in people with rheumatoid arthritis [dissertation]. Bristol (UK): University of the West of England; 2009.
- 23 Nicklin JK, Hewlett S, Two R, Kirwan JR. Have you felt embarrassed or upset? An example of the care needed in translating patient reported outcome measures. *Ann Rheum Dis* 2014;73(Supp2):OP0087-HPR.
- 24 Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcomes measures for use in clinical trials. *Health Technology Assessment (UK)* 1998;2:1-74.
- 25 US Dept Health and Human Services Food and Drug Administration. Guidance for Industry: Patient-reported outcome measures: Use in medical product development to support labelling claims. 2009
<http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf> (Accessed 9 February 2015)
- 26 Boers M, Kirwan JR, Tugwell P et al. The OMERACT Handbook. 2014.
http://www.omeract.org/pdf/OMERACT_Handbook.pdf (Accessed 9 February 2015)
- 27 De Wit M, Berlo S, Aanerud G et al. EULAR Recommendations for the inclusion of patient representatives in scientific projects – A EULAR initiative. *Ann Rheum Dis* 2011;70:722-6.
- 28 Uysal-Bozkir O, Parlevliet J, Rooij S. Insufficient cross-cultural adaptations and psychometric properties for many translated health assessment scales: A systematic review. *J Clin Epidemiol* 2013;66:608-18.

- 29 Ornetti P, Parratte S, Gossec L et al. Cross-cultural adaptation and validation of the French version of the Knee injury and Osteoarthritis Outcome Score (KOOS) in knee osteoarthritis patients. *Osteoarthritis Cartilage* 2008;16:423-8.
- 30 Katchamart W, Ussavasodhi P, Chiowchanwesawakit P, Chanapai W. Cross-cultural adaptation and validation of the Thai version of the Rheumatoid Arthritis Disease Activity Index (RADAI). *Rheumatol Int* 2013;33:2503–8.
- 31 Primdahl J, Wagner L, Hørslev-Petersen K. Self-efficacy in Rheumatoid Arthritis: Translation and test of validity, reliability and sensitivity of the Danish Version of the Rheumatoid Arthritis Self-Efficacy Questionnaire (RASE). *Musculoskeletal Care* 2010;8:123-35
- 32 Roberts G, Roberts S, Tranter R et al. Enhancing rigour in the validation of patient reported outcome measures (PROMs): Bridging linguistic and psychometric testing. *Health Qual Life Outcomes* 2012;10:64
- 33 Malaviay A. Outcome measures in rheumatoid arthritis: Indian perspective. *Indian Journal of Rheumatology* 2013;8:S36-7.
- 34 Islam N, Baron Basak T, Oude Voshaar MAH et al. Cross-cultural adaptation and validation of the Bengali health Assessment Questionnaire for use in rheumatoid arthritis patients. *Int J Rheum Dis* 2013;16:413-7.
- 35 Dharmanand BG. Patient perspective in rheumatology outcome measures: Indian perspective. *Indian Journal of Rheumatology* 2013;8:S23.

Table 1: Demographic data of patients participating in cognitive interviewing (n=60)

	UK 10	France 10	Germany 10	Netherlands 10	Spain 10	Sweden 10	Total 60 (%)
Disease duration							
<5 yrs	4	2	5	3	2	3	19 (32%)
Age							
<39 yrs	1	1	2	3	1	2	10 (17%)
40-60	3	4	5	4	9	5	30 (50%)
>60 yrs	6	5	3	3	0	3	20 (33%)
Gender							
Female	7	8	8	7	8	5	43 (72%)
Education							
Standard school age	8	2	0	3	2	1	16 (27%)
1 st degree/prof qual	2	5	7	3	3	6	26 (43%)
Higher degree	0	3	3	4	5	3	18 (30%)
Disease activity (VAS)*							
0-3.9	5	2	6	6	2	1	22 (37%)
4.0-6.9	2	7	2	2	6	4	23 (38%)
7.0-10	3	1	2	2	2	5	15 (25%)
Fatigue (NRS)*							
0-3	3	2	5	3	1	1	15 (25%)
4-6	4	6	1	4	3	6	24 (40%)
7-10	3	2	4	3	6	3	21 (35%)
Disability (HAQ)*							
≤1.00	4	4	6	7	3	3	27 (45%)
1.125-2.00	3	5	3	2	5	6	24 (40%)
2.125-3.00	3	1	1	1	2	1	9 (15%)

* High = more severe

Table 2: Issues raised for BRAFs and RAID in cognitive interviews (10 patients/country)

	France	Germany	Netherlands	Spain	Sweden	UK*
BRAF-MDQ						
Q1	1R	1J		3U	1R	
Q2			1R	1U, 1Rs	1R, 1Rs	
Q3	2U	1J, 1Rs		1R, 1J	2J, 1Rs	
Q4		1J, 1Rs	1U, 1J	2U, 1Rs	2U, 1J, 1Rs	
Q5	1J				1J	
Q6	1J, 1Rs					
Q7	2J				1J	
Q8	1J		2U		1Rs	
Q9			1R		1U	
Q10	1J	1U	1U	3U	1U, 1Rs	
Q11	1J				2J	
Q12		1U, 1R		3U, 1J	1U, 1Rs	
Q13	1J	1J		1U	1Rs	
Q14			1U	1U		
Q15					1U, 1R, 1J	
Q16				1J, 1Rs		
Q17	1U, 1R	2U	1U	2U	3U	
Q18	3U				2U	
Q19		2U	1U		1J	
Q20			1U			
BRAF-NRS						
Severity	1J			3U, 1J, 2Rs	2R	
Effect				3U, 1J, 1Rs	1U, 1J	
Coping			1U, 1J, 7Rs	1U	2U, 1J	
RAID						
Q1	2R, 2Rs		1Rs	1J, 3Rs		
Q2	1Rs	1J	1U, 1R	3U, 3J, 2Rs	2U, 2J	
Q3			1U, 1R, 1J	2U, 1J	1J, 1Rs	1J
Q4	1U, 1R		1U	2J, 1Rs		3J
Q5	1U, 2R, 2Rs		4U	1U, 3Rs	3U, 1J	1U, 1J
Q6	1J		1U, 1Rs	2U, 3Rs	3U, 1Rs	1J
Q7	3U			6U, 1Rs	1Rs	1U

U=Understanding; R=Retrieval; J=Judgement; Rs=Response option

*UK did not complete BRAFs as these were extensively evaluated during development

Table 3: Netherland comments on BRAF-NRS Coping (n=10)

Please circle the number which shows how well you have <u>coped</u> with fatigue over the past 7 days : (0 = Not at all well, to 10 = Very well)		
ID	Patient comment	Interviewer comment
301	"Pretty good"	Patient circles 4/10, which reflects pretty poor
302	"One day I take timeout. I can deal with it: [score] a 1"	Patient circles 5/10 not 1
304	"Effect" is I think something more abstract than "dealing with." "Dealing" - you did something or you did not do anything. You can hang them on activities or something. What I said about "I have played sport less this week" that you should maybe indicate as "cope". But you can also say it is an "effect" of your fatigue. It can be both, so I would put the same here"	Patient wishes to rate effect and coping the same, but circles 9 for effect (poor) and 9 for coping (good)
305	[Patient scored immediately without thinking aloud]	Patient circled coping 0/10 (poor) but scored his fatigue as being low on all other BRAF items
310	"Oh, I can handle it well, I put a 3"	Patient circles 3/10 which reflects poor coping

Table 4: Netherland comments on RAID Q5 (n=10)

Considering your arthritis overall, how would you rate your level of physical well-being (without taking pain, inflammation and fatigue into account) during the past week?

ID	Patient comment	Interviewer comment
302	“ Outside of all the pain and inflammation and fatigue Just as if you do not have RA, I think. Well then I really feel well. Pretty good”	Patient understood the question as if she had to rate her physical well-being as if she had no RA, instead of rating her physical well-being due to RA without taking pain, inflammations and fatigue into account. Scored 1/10
303	“Well, not so much trouble, but... [hesitates]” <i>Int:</i> “Does it make sense? What does your ‘But...’ mean?” <i>Pt:</i> “For me it's always a problem that I cannot go out” [wheelchair user]	The patient answered that she could not go outside and that is her problem. Eventually she chose 5/10. Further on in the interview the pt. said she could not go outside this week because of the snow, which is not wheelchair-friendly.
307	“In general. Yes fatigue weighs very heavy. Pain and inflammation are lighter, so I just make a mark in the middle”	Patient understood the question as if she had to rate her pain, inflammations and fatigue, instead of rating her physical well-being due to RA without taking these symptoms into account. Score 5/10

Table 5: Spanish comments on RAID Q7 (n=10)

Considering your arthritis overall, how well did you cope (manage, deal, make do) with your disease during the last week?		
ID	Patient comment	Interviewer comment
502	“‘ <i>Hacer pasar.</i> ’ What is that? Sounds funny”	Patient did not like this phrase
504	“‘ <i>Hacer pasar.</i> ’ Is this Spanish?” (<i>mocking facial expression</i>)	Patient did not like this phrase
505	“‘ <i>Hacer pasar.</i> ’ I don’t like this”	Patient suggested ‘ <i>apañarse</i> ’ (muddle through)
506	“‘ <i>Afrontar.</i> ’ You need a degree to understand that! What does it mean?”	Patient considers this an academic term for coping
508	“‘ <i>Afrontar....Hacer pasar.</i> ’ Oh my God! What is all this? Oh, I see, ‘ <i>sobrellevar</i> ’ overcome. Yes, I overcame the diagnosis well, with time”	Patient did not like this, understands this as only relating to the time of diagnosis (adjusting to the diagnosis).
509	“Yes, I understand ‘ <i>afrontar</i> ’ - if I can go to work”	When it was suggested this question was about more than going to work, patient replied “Then tell me what?”