

Title Page

Sanderson T¹, Kirwan J², Almeida C³, Morris M³, Noddings R², Hewlett S³.

Item development and face validity of the Rheumatoid Arthritis Patient Priorities in Pharmacological Interventions (RAPP-PI) outcome measures

Running title: Item development and face validity of the RAPP-PI outcome measures

¹ Durham University, Department of Anthropology, Durham University, Durham, DH1 3LE

² University of Bristol, Academic Rheumatology, Bristol Royal Infirmary, Bristol, BS2 8HW

³ University of the West of England, Glenside Campus, Blackberry Hill, Bristol, BS16 1DD

Corresponding author: t.c.sanderson@durham.ac.uk
Tel.: 0118 947 2277
Fax: 0117 342 3841 (c/o office)

Item development and face validity of the Rheumatoid Arthritis Patient Priorities in Pharmacological Interventions (RAPP-PI) outcome measures

Abstract

Background

The assessment of rheumatoid arthritis (RA) is dominated by core sets and indices that have been developed by RA professionals. Previous research developed a set of eight priority treatment outcomes generated by patients to complement the professionally developed core sets for RA.

Objective

This study aimed to facilitate quantitative measurement of these outcomes.

Methods

Two consultation meetings with patient research partners diagnosed with RA (N=18) were held to identify face validity in existing instruments (Phase 1) at the Bristol Royal Infirmary (BRI). Where validated measures did not exist, new numerical rating scales (NRS) were constructed and discussed at two focus groups with patients diagnosed with RA (N=8) at the BRI and Royal National Hospital for Rheumatic Diseases (RNHRD) (Phase 2). Feedback on the stem question, time frame, anchors and layout was recorded and transcribed verbatim.

Results

Of the 8 priorities, existing NRS for pain, activities of daily living (ADL) and fatigue were voted as acceptable (Phase 1), but new NRS were required for 5 priorities. The partners strongly recommended that the three separate domains of *severity*, *effect* and *ability to cope* in each measurement area be assessed, as in the existing validated fatigue NRS. Focus group participants (Phase 2) made significant contributions to the phrasing of questions, for example how to ensure 'mobility' could be uniformly understood and how changes in valued activities be judged appropriately.

Conclusion

Through extensive patient feedback 24 numerical rating scales were constructed based on priorities identified by patients and encompassing domains where existing questionnaires contain many more items and do not address three important concepts endorsed by patients: severity, effect and coping. The RAPP-PI PROMs are now ready for the evaluation of comprehension, construct validity and sensitivity through an observational study.

Key Points for Decision Makers

- Patient research partners can effectively contribute in the identification of face validity in instruments;

- A clear consensus was reached among RA patients that the each of 8 outcomes should be measured according to three distinct domains: *severity, effect and ability to cope*.

1. Introduction

Rheumatoid arthritis (RA) is an unpredictable auto-inflammatory disease characterised by painful and swollen joints, fatigue and joint damage. Traditionally, rheumatologists and other health professionals have decided which outcomes should be measured and these have been standardised using core sets and composite indices to allow comparison across studies. The OMERACT / American College of Rheumatology (ACR) core set [1,2] and the Disease Activity Score (DAS) [3], developed by RA professionals, have become dominant in clinical practice and in research. Although these include patient-reported outcomes, patients were not involved in their development or selection. The OMERACT/ACR variables are tender and swollen joint counts, physician's global assessment of disease activity, inflammatory markers, and the patient's assessment of pain, disease activity and physical function. The patient's assessment of pain and disease activity are commonly measured by visual analogue scales (VAS) although the wording of the VAS have not been standardised and has been shown to vary across studies [4]. Physical function is commonly measured by the Health Assessment Questionnaire (HAQ) [5], These variables were chosen through a process of literature review and nominal groups with experts because they sample the broad range of improvement in RA (have content validity) and are all at least moderately sensitive to change [1,3].

The DAS is a composite index used to assess disease activity in RA and provides evidence of clinical need and governs decisions on the appropriateness of biologic therapies in the UK [6]. The DAS is an algorithm of tender and swollen joint counts, blood inflammatory markers, and patient opinion of global health or disease [3]. These variables were chosen by rheumatologists from the literature and tested for their relative contribution to changes in inflammation. Patient opinion of global health or disease is measured by a VAS, the wording of which has also been found to vary across studies and clinical practice with a potential impact on prescribing biologic therapy [4].

The International Classification of Functioning, Disability and Health (ICF) core set for RA [7], is the most comprehensive but with 96 items (26% body functions, 19% body structures, 33% activities and participation, and 22% environmental factors) it is not parsimonious and has not been widely used even in its brief form. This was developed using Delphi techniques with RA professionals, systematic review and empirical data collection. It was validated with RA patients after its publication, using the core set as a structure for the focus groups [8], and hence limiting the potential for new items to emerge. All these methods have focused on obtaining professional priorities for assessing disease activity, although OMERACT has since endorsed fatigue as an essential measure in RA, following substantial input from patient research partners [9].

The USA Food and Drug Administration (FDA) has recognised the need to incorporate a wider patient perspective in drug evaluation programs. It has called for (and issued draft guidance on) standardisation of methods for developing patient-reported outcome measures (PROMs), which incorporate the patient perspective [10]. PROMs have become standard measures of pain, functional disability and global health/disease activity in RA due to their inclusion in OMERACT/ACR criteria and the DAS [1-3]. They have been shown to be as effective as the traditional physician- or laboratory- reported outcomes in reflecting changes in disease activity over time, and predicting long term-morbidity and mortality [11].

When developing instruments to measure PROMs it is essential to use appropriate methodology to involve patients, such as focus groups to develop appropriate wording for numerical rating scales (NRS), and cognitive interviewing to explore understanding of candidate questions, as recommended by the FDA [12,13].

The growing rheumatology literature on the patient perspective has revealed three important issues. First, differences have been shown between physician and patient assessments of outcomes such as pain, overall health, and physical and mental function [14]. Second, RA patients identify treatment outcomes that are not routinely measured, such as sleep, well-being, and normality [15-17], which may relate to disease activity. Feeling well and returning to normal were consistently described as important outcomes by RA patients in focus groups and interviews in the UK [15,16,18] and in Sweden [19]. Third, RA patients have both similar and different priorities to clinicians. The RA Impact of Disease (RAID) study developed an index of seven patient priority domains with patients and clinicians across ten European countries: pain, functional capacity, fatigue, emotional well-being, quality of sleep, coping, and physical well-being [16].

Previously, some authors of this paper (TS, MM and SH) had developed the RA Patient Priorities in Pharmacological Interventions (RAPP-PI) patient core set by collaboration with patient research partners. This identified eight priority outcomes: pain, activities of daily living (ADL), visible joint damage, mobility, life enjoyment, independence, fatigue, and valued activities [20]. Although pain, functional disability, fatigue and joint damage are routinely measured, the other outcomes in these patient-centred studies are not, despite being linked to experience of disease by patients. Feasibility is an important requirement of outcome measures [21], and the responder burden of completing many multi-item questionnaires to measure each of these domains may be unfeasible in many studies. Therefore we wanted to develop a single measure that allows treatment outcomes to be assessed with validity and reliability in all domains of disease regarded as important to patient and clinicians.

Comparison of the RAPP-PI patient core set with commonly used PROMs in RA clinical trials (Short Form 36 Health Survey questionnaire (SF-36), RAQoL questionnaire, Nottingham Health Profile (NHP) and Arthritis Impact Measurement Scale (AIMS2)) showed that none of them covered all eight RAPP-PI outcomes [22]. Therefore, the addition of an existing validated questionnaire to the RA core set would not be sufficient to assess patient priorities highlighted by the RAPP-PI. For this, the identification of eight validated NRS with sufficient face validity from the literature or a new PROM would be required. The RAPP-PI patient core set was developed through three phases. First, in-depth interviews were conducted with RA patients to identify outcomes and the data were analysed in collaboration with patient research partners [16]. Second, nominal

groups were held to begin to prioritise more than 60 outcomes with different RA patients. Third, a national survey was conducted to prioritise 32 outcomes in a manageable number for a core set, and which was representative of the RA population [20]. A patient research partner contributed to the steering committee throughout the study. The patient-centred method for outcome selection provides a foundation for involvement of patients in the identification of existing NRS or the development of a new outcome measure. This paper reports the process to enable the RAPP-PI patient core set outcomes to be measured, by identifying whether existing NRS have sufficient face validity from a patient perspective and developing a practical tool to capture changes in priorities known to be important to patients. While there is a substantial choice of instruments to assess RA (see Methods for examples), the aim of this project was to develop a comprehensive and feasible outcome measure including all of the patient priorities previously identified [20], constructed with a high level of patient involvement at every stage of the process. After validation studies, this will produce a multidimensional instrument to assess the patient perspective of impact,

2. Methods

There were two stages in the development of the RAPP-PI outcome measures: consultation meetings with patient research partners (Phase 1), and focus groups (Phase 2). Subsequent cognitive interviews and pilot testing of the measures (Phase 3) is reported separately [23]. Ethics approval was granted by the London Dulwich Research Ethics Committee (11/LO/1524). Patient information was provided to eligible participants for each phase, and written informed consent was taken. Two patient research partners were part of the steering committee (BN and AW).

2.1 Identification of instruments to measure priority outcomes (Phase 1)

Patient research partners with RA at the Bristol Royal Infirmary (BRI) attended two consultation meetings in November and December 2011. All had definite RA according to ACR criteria [24] and varying educational and professional experience (Table 1). Their mean (SD) age was 53.3 (12.2) years, disease duration 19.2 (10.5) years 15 were female and 7 were currently in part time or full time employment. Fifteen had previously been involved in research projects as patient partners in a variety of roles (Table 2). For three others this was their first contribution as a patient research partner, although they had attended a training day in preparation. The partners were sent a booklet containing visual analogue scales (VAS) or NRS found in the literature review [25-41] (Table 3) selected by the research team, or items or subscales from questionnaires where no VAS/NRS could be identified that assessed a similar notion in to that in the original interview data [15]. Not all scales or items identified in the literature were included. For example, items from the Valued Life Activities (VLA) were not used because they were too specific, but were discussed in the consultation meeting as extra examples of wording [42]. Limited existing scales were identified for the eight priorities (Table 3): six for pain; ADL six; mobility two; life enjoyment three; independence four; fatigue five; and valued activities five. For joint damage three NRS were drafted from previous interview data [15] because no instruments were identified that assessed patient perceived joint damage. Partners were asked to complete the booklet as if they were a participant completing the questionnaire, and then to write comments about their preferences for the scales listed, thereby

identifying whether existing instruments were considered to effectively assess the eight priorities from a patient perspective.

Four priorities were discussed at each meeting. Partners shared their opinion about the face validity of the scales and then voted for their preference for a specific existing scale or the need for a new scale to be developed. Emphasis was put on a scale being 'good enough', rather than perfect, given the resources involved in constructing new instruments. A consensus of greater than 50% was required for an existing scale to be used. The meetings were digitally recorded and transcribed verbatim. Content analysis [43] was used to identify reasons for supporting or rejecting the use of an instrument. These qualitative data were used to inform the draft NRS in Phase 2.

2.2 Development of new RAPP-PI NRS (Phase 2)

Two focus groups were held with patients with a diagnosis of RA [20] at Bristol Royal Infirmary and the Royal National Hospital for Rheumatic Diseases in January and February 2012, who had not participated in Phase 1 (Table 1). Prior to the groups, draft scales were constructed where partners in Phase 1 had voted for new instruments. Data from Phase 1 on the wording of questions, timeframe, response 'anchors' at each end of the scale and layout were utilised to construct the drafts by the research team, including the two patient research partners on the team.

The draft NRS were presented to the focus group participants for feedback on the stem question, time frame, anchors and layout. The groups were digitally recorded and transcribed verbatim. The transcripts were analysed primarily by TS and CA using content analysis [43], with the research team (including the two patient research partners) contributing to the finalisation of the questions.

3. Results

3.1 Consultation meetings (Phase 1): Identification of instruments to measure priority outcomes

In total 18 patient research partners attended at least one consultation meetings, with 14 attending both meetings. Existing NRS for pain [16] (12/17 votes), ADL [16] (12/17 votes) and fatigue [37] (13/15 votes) were voted as acceptable. The other five priority outcomes received between 12-17 votes for a new scale to be developed. The selected fatigue scale consists of three components: level of fatigue, effect of fatigue, and coping with fatigue. The partners argued that these three components were measuring important different aspects of fatigue and that the three components should be applied to all of the priority outcomes. 'Level' was interpreted as the severity of the priority outcome experienced, and was understood as distinct from 'effect' and 'ability to cope' (see Table 3 for example quotations). 'Effect' was understood as how a priority outcome affected their life (that is, the impact). 'Coping' was interpreted as a patient's own ability to change how the outcome was experienced. A vote was taken on the inclusion of the three components at the second consultation meeting and

all 15 participants (100%) agreed that each priority should be represented by three questions on severity, effect and ability to cope. This resulted in the need for 24 questions (19 new).

The qualitative data from the discussions and comments written on the instrument booklets regarding support for or rejection of existing instruments provided the content for the draft NRS (Table 3). For example, participants consistently argued that it was important to put 'rheumatoid arthritis' into the wording of each question. The draft NRS were piloted with the two patient research partners on the steering committee, before proceeding to Phase 2.

3.2 Development of new RAPP-PI NRS with the three components Severity, Effect and Coping (Phase 2)

Although 14 patients had agreed to participate in the two focus groups only 8 took part. Their mean (SD) age was 64.7 (8.2) years, disease duration 12.8 (11.8) years, their disease severity measured on a patient global visual analogue scale (0= well, 10=very bad) was 4.0 (2.5), 6 were female, 2 were in part time or full time employment, 3 had university qualifications and a further 3 had school qualifications at age 16. The reasons given by other patients for lack of attendance were an RA flare (2), transport difficulties on the day (2) and unknown (2). Focus group discussions initially discussed the wording of the NRS (stem questions), anchors, timeframe, and layout (Table 4), as specified in the topic guide (Figure 1). For example, different understandings of the word '**mobility**' were apparent: "Moving around", "Get from room to room", "Doing your chores", "Movement in your arm" and "Dexterity". There was a consensus from participants that the phrase 'getting around' should be used to convey the original meaning of the priority from the previous interview data [17].

In the consultation meeting (Phase 1) four participants (of 17) worried about the effect of the term '**joint damage**' on newly diagnosed patients who completed the questions. However, the focus group participants suggested that joint change could be understood by patients to include improvements of joints, as well as worsening, and this would cause scoring problems for the NRS. The consensus from the focus groups was to use 'joint damage' as patients should be "realistic" about the condition. In relation to the '**valued activities**' priority, participants agreed that as their condition had progressed, the activities that they enjoyed had to change: "I suppose I've modified things I've tried to do down the years... It should ask about doing activities you *currently* enjoy" (his emphasis). The group decided that the comparison to pre-RA activities could create a distorted answer to the scales. The feedback on layout included using boxes around each set of priority questions for visual separation, importance of consistency in the format of questions, keeping questions as short as possible, and underlining key words to differentiate between the questions. Participants also spontaneously discussed their support for the eight priorities (see Table 4 for examples).

The draft NRS were edited in light of the focus group analysis and finalised through consensus with the research team, including the two patient research partners. The final NRS are provided in Figure 2, including the instructions for completing the questions as also presented to the focus group participants for feedback (e.g. see layout quotation number 3).

4. Discussion

This research has developed patient-centred outcome measures that have strong face validity, through involvement of patient research partners in consultation meetings and on the steering committee, and patients in focus groups. Not only were patients able to significantly contribute to this process, they also understood the process of instrument validation (in Phase 1, to decide whether existing instruments were appropriate) and the measurement properties of wording, timeframe and anchors (Phase 2). For example, the discussion on how to phrase the priority on Visible joint damage included the impact of wording on patients (damage vs change), the effect of anchors on scoring, and the need for an extended timeframe.

Existing validated measures were identified by participants for three of the priorities (pain, activities of daily living (ADL) and fatigue), and only one of these (fatigue) addressed all three components of severity, effect and coping [44]. There was a complete consensus amongst patient research partners that the three components should be addressed for each of the priorities. Thus where there are existing measures, many are multi-item questionnaires, and they do not address these different components. The development of a single item for each priority allows the inclusion of all three components without expanding the questionnaire to such an extent that it becomes too burdensome for responders to complete. However, there is limited evidence that these components measure significantly different constructs. Nicklin and colleagues [37] found satisfactory criterion and construct validity of the Bristol RA Fatigue (BRAf) short scales for severity, effect, and coping, and that the data also demonstrated a ‘potential disconnect between the ability to cope with fatigue and its severity, which offers the possibility that patients might be able to improve their ability to cope with fatigue and thus reduce its effect, even if fatigue severity per se cannot be changed’ (p.1566). It is unknown whether similar effects may be found for the seven other RAPP-PI priorities and further research is required to determine how these three components are related within the Impact Triad [45].

NRS were constructed for each of the priorities where existing validated instruments did not have sufficient face validity, and all of the 24 NRS formatted to have a consistent layout. The priority of Visible joint damage provided the greatest challenge because no existing instrument could be found that measured patients’ self-reporting of joint damage. The three NRS were developed from previous interview data [17], and discussed during the consultation meetings and focus groups. A minority of Phase 1 participants were concerned about the effect of the term ‘joint damage’ on newly diagnosed patients, but the overall consensus was that RA patients should be realistic about the potential consequences of progression of their disease and the implications for measurement were more important. However, testing of these NRS through an observational study may show that self-reported joint damage may not be an effective method for measuring this patient priority. It may be more reliable to use clinicians’ assessment of tender and swollen joints, and PROMS for self-reported anxiety about joint damage instead.

Two of the existing scales that were voted as having acceptable face validity were taken from the RA Impact of Disease (RAID) index [16]. The similarities in domains between the RAID and RAPP-PI may derive from the common focus on patient perspective and it is reassuring that patient priorities have been identified. However,

the diverse approaches in constructing the NRS may result in a different final selection of domains for the RAPP-PI. Furthermore, one noticeable difference between the RAID approach and the one adopted here is that patients clearly recognised the difference between the three components of severity, effect and coping. It would be important in an observational study to compare how these RAPP-PI components relate to the RAID NRS, that is, to establish with which component the RAID's 'Impact' NRS correlates most. These data may provide an indication of the necessity of including questions about ability to cope in rheumatology instruments.

A limitation of this research is that only eight (of 14 confirmed) participants attended the focus groups (Phase 2) across two sites. Although relevant data were also generated in the consultation meetings (Phase 1), these participants were limited to one site. However, the large number of patient research partners involved, with diverse educational and professional experience as well as of user involvement, enabled the process of item development and face validity to benefit from lively discussion. In addition to their diverse experiences of living with the condition, this discussion was generated from their involvement in other studies, including other outcome measure development, and the latest research partner training day that focused on methodological issues. Therefore, we argue that the effect of involving patient research partners in addition to patient participants was to fully engage with conceptual issues such as the different aspects of domains.

Further testing of the NRS wording will take place through future cognitive interviewing to address this, covering comprehension of the questions, ability to recall information and make a judgement [46]. A strength is that these new outcome measures are based on the development of a conceptual framework grounded in patient data from a previous study [17], in collaboration with patients as recommended by the FDA [10], and benefiting from a higher level of user involvement [47]. Further research aims to construct a final tool, which will assess priorities important to patients in one instrument, constructing an algorithm from observational data to produce a single score.

5. Conclusions

It is effective to collaborate with patients when developing new measures, and a range of methods can be successfully used, including consultation meetings and focus groups. This study, through extensive patient feedback, constructed an instrument with 24 NRS based on priorities identified by patients and encompassing domains where existing questionnaires contain many more items and do not address three important concepts endorsed by patients: severity, effect and coping. Further research is required to pilot the RAPP-PI PROMs for the evaluation of feasibility, construct validity and sensitivity through an observational study.

Compliance with Ethical Standards

Funding: Supported by Arthritis Research UK feasibility study grant 19434.

Conflicts of interest: Dr Sanderson, Professor Kirwan, Ms Almeida, Dr Morris, Mr Noddings and Professor Hewlett have no conflicts of interest.

Ethical approval: Approval was granted by the London Dulwich Research Ethics Committee (11/LO/1524).
Informed consent: Written informed consent was obtained for participating in the consultation meetings and focus groups, the recording of discussion, and the use of anonymised quotations.

Acknowledgements

We appreciate the support with recruitment at the RNHRD by Dr Madeleine Piper, Nina Griffith and specialist nurses Sue Brown, Debbie Bond and Nicola Waldron; the support with recruitment at the BRI by Jan Richardson and specialist nurses Tracy French and Sue Taylor; Anne Watts, patient research partner. All authors contributed to the design, funding application, analysis and authorship of this paper. TS and CA conducted the data collection. TS is the guarantor for the study.

References

1. Boers M, Tugwell P, Felson DT, van Reil PLCM, Kirwan JR, et al. World Health Organisation (WHO) and International League Against Rheumatism (ILAR) core endpoints for symptom-modifying antirheumatic drugs in rheumatoid arthritis clinical trials. *J Rheumatol.* 1994; 21(41): 86-9.
2. Felson D, Anderson J, Boers M. The American College of Rheumatology preliminary cores set of disease activity measures for rheumatoid arthritis clinical trials. *Arthritis Rheum.* 1993;36(6):729-40.
3. Van der Heijde D, Hof M, Piet L, Putte L. Development of a disease activity score based on judgement in clinical practice by rheumatologists. *J Rheumatol.* 1993;20(3):579-81.
4. French T, Hewlett S, Kirwan J, Sanderson T. Different wording of the Patient Global visual analogue scale (PG-VAS) affects Rheumatoid Arthritis patients' scoring and the overall Disease Activity Score (DAS28): A cross-sectional study. *Musculoskeletal Care.* (early release) DOI: 10.1002/msc.1046.
5. Fries JF, Spitz PW, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum.* 1980;23:137-45.
6. NICE. Adalimumab, etanercept and infliximab for the treatment of rheumatoid arthritis after failure of a TNF-alpha inhibitor. Appraisal consultation document. London: NICE; 2008.
7. Stucki G, Cieza A, Geyh S, Battistella L, Lloyd J, Symmons D, Kostanjsek N, Schouten J. ICF Core Sets for rheumatoid arthritis. *J Rehabil Med.* 2004;Supp(44):87-93.
8. Coenen M, Stamm T, Cieza A, Amann E, Kollerits B, Stucki G. Validation of the comprehensive ICF core set for rheumatoid arthritis from the patients' perspective using focus groups. *Arthritis Res Ther.* 2006;8:R84.

9. Kirwan J, Minnock P, Adebajo A, Bresnihan B, Choy E, De Witt M, et al. Patient Perspective Workshop: Fatigue as a recommended patient-centred outcome measure in rheumatoid arthritis. *J Rheumatol* 2007; 34: 1174-82.
10. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Center for Devices and Radiological Health. Guidance for industry: patient-reported outcome measures. Use in medical product development to support labeling claims; 2009.
<http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf>
. Accessed 4 Jan 2014.
11. Linde L, Sorensen J, Ostergaard M, Horslev-Pedersen K, Lund Hetland M. Health-related quality of life: validity, reliability, and responsiveness of SF-36, 15D, EQ-5D, RAQoL, and HAQ in patients with rheumatoid arthritis. *J Rheumatol*. 2008; 35:1528-37.
12. Patrick DL, Burke LB, Powers JH, Scott JA, Rock EP, Dawisha S, et al. Patient-reported outcomes to support medical product labeling claims: FDA perspective. *Value Health*. 2007;10(Suppl 2):S125–37.
13. Frost MF, Reeve BB, Liepa AM, Stauffer JW, Hays RD. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? *Value Health*. 2007;2:S94–105.
14. Hewlett S. Patients and clinicians have different perspectives on outcomes in arthritis. *J Rheumatol*, 2003;30(4):877-9.
15. Hewlett S, Carr M, Ryan S, Kirwan J, Richards P, Carr A et al. Outcomes generated by patients with rheumatoid arthritis: how important are they? *Musculoskeletal Care*. 2005;3(3):131-42.
16. Gossec L, Patternotte S, Aanerud GJ, Balanescu A, Boumpas DT, Carmona L 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;70:935-942.
17. Sanderson T, Morris M, Calnan M, Richards P, Hewlett S. What outcomes from pharmacological treatments are important to people with rheumatoid arthritis? Creating the basis of a patient core set. *Arthritis Care Res*. 2010;62(5):640-6.
18. Marshall NJ, Wilson G, Lapworth K and Kay LJ. Patients' perceptions of treatment with anti-TNF therapy for rheumatoid arthritis: a qualitative study. *Rheumatology*. 2004;43:1034-8.
19. Ahlmen M, Nordenskiöld U, Archenholtz B, Thyberg I, Ronnqvist R, Linden L et al. Rheumatology outcomes: the patient's perspective. A multicentre focus group interview study of Swedish rheumatoid arthritis patients. *Rheumatology*. 2005;44(1):105-10.

20. Sanderson T, Morris M, Calnan M, Richards P, Hewlett S. Patient perspective of measuring treatment efficacy: the Rheumatoid Arthritis Patient Priorities for Pharmacological Interventions (RAPP-PI) outcomes. *Arthritis Care Res.* 2010;62(5):647-656.
21. Boers M, Kirwan JR, Wells G, Beaton D, Gossec L, D'agostino M-A, et al. Developing Core Outcome Measurement Sets for Clinical Trials: OMERACT Filter 2.0. *J Clin Epidemiol.* 2014; 67: 745-53.
22. Sanderson T. Well-being and the prioritisation of treatment outcomes in rheumatoid arthritis. Doctoral thesis; 2009.
23. Sanderson T, Kirwan J, Almeida C, Morris M, Watts A, Noddings R, Hewlett S. Feasibility, and preliminary validity and sensitivity of the Rheumatoid Arthritis Patient Priorities in Pharmacological Interventions (RAPP-PI) outcome measures (*awaiting submission*)
24. Arnett FC, Edworthy SM, Bloch DA, McShane, DJF, Cooper NS, Healey LA, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum.* 1988;31:315-24.
25. Herr KA, Spratt K, Mobily PR, Richardson G. Pain intensity assessment in older adults: use of experimental pain to compare psychometric properties and usability of selected pain scales with younger adults. *Clin J Pain.* 2004;20(4):207-19.
26. Anderson D. Development of an instrument to measure pain in rheumatoid arthritis: Rheumatoid arthritis pain scale (RAPS). *Arthritis Care Res.* 2001;45;317-23.
27. Pincus T, Swearingen C, Wolfe F. Toward a multidimensional Health Assessment Questionnaire (MDHAQ): assessment of advanced activities of daily living and psychological status in the patient-friendly health assessment questionnaire format. *Arthritis Rheum.* 1999;42:2220–30.
28. Wolfe F, Michaud K, Pincus T. Preliminary Evaluation of a Visual Analog Function Scale for Use in Rheumatoid Arthritis. *J Rheumatol.* 2005;32:1261-6.
29. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol.* 1988;15:1833–40.
30. WHOQOL. The World Health Organisation Quality of Life (WHOQOL)-BREF UK version. WHO: Geneva; 2004.
31. WHOQOL. The World Health Organisation Quality of Life (WHOQOL) -100. WHO: Geneva; 1995.
32. Meenan RF, Mason JH, Anderson JJ, Guccione AA, Kazis LE. AIMS2: the content and properties of a revised and expanded Arthritis Impact Measurement Scales health status questionnaire. *Arthritis Rheum.* 1992;35:1-10.

33. Diener, E., Emmons, R.A., Larsen, R.J. and Griffin, S. 1985. The Satisfaction With Life Scale. *J Pers Assess.* 1985;49(1):71-5.
34. Stuijbergen AK, Seraphine A, Roberts G. An explanatory model of health promotion and quality of life in chronic disabling conditions. *Nursing Res.* 2000;49(3):122-9.
35. de Jong Z, van der Heijde D, McKenna SP, Whalley D. The reliability and construct validity of the RAQoL: A rheumatoid arthritis-specific quality of life instrument. *Brit J Rheumatol.* 1997;36:878-83.
36. Nicassio PM, Wallston KA, Callahan LF, Herbert M, Pincus P. The measurement of helplessness in rheumatoid arthritis: the development of the Arthritis Helplessness Index. *J Rheumatol.* 1985;12:462-7.
37. Nicklin J, Cramp F, Kirwan J, Greenwood R, Urban M, Hewlett S. Measuring fatigue in rheumatoid arthritis: a cross-sectional study to evaluate the Bristol Rheumatoid Arthritis Fatigue Multi-Dimensional questionnaire, visual analog scales, and numerical rating scales. *Arthritis Care Res.* 2010;62(11):1559-68.
38. Khanna D, Pope JE, Khanna PP, Maloney M, Samedi N, Norrie D, Ouimet G, Hays R. The minimally important difference for the fatigue visual analog scale in patients with rheumatoid arthritis followed in an academic clinical practice. *J Rheumatol.* 2008;35(12):2335-43.
39. Westhovens R, Bole JC, Li T, Martin M, MacLean R, Lin P, Blaisdell B, Wallenstein GV, Aranda R, Sherrer Y. Improved health-related quality of life for rheumatoid arthritis patients treated with abatacept who have inadequate response to anti-TNF therapy in a double-blind, placebo-controlled, multicentre randomized clinical trial. *Rheumatology.* 2006;45,10:1238-46.
40. Brazier JE, Harper R, Jones NMB, O'Cathain A, Thomas KJ, Usherwood T, Westlake L. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ.* 1992;305:160-4.
41. Devins GM, Binik YM, Hutchinson TA, Hollomby DJ, Barré PE & Guttman RD. The emotional impact of end-stage renal disease: Importance of patients' perceptions of intrusiveness and control. *Int J Psych Med.* 1983;13(4):327-43.
42. Katz PP, Radvanski DC, Allen D, Buyske S, Schiff S, Nadkarni A, Rosenblatt L, Maclean R, Hassett AL. Development and validation of a short form of the valued life activities disability questionnaire for rheumatoid arthritis. *Arthritis Care Res (Hoboken).* 2011;63(12):1664-71.
43. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. *Nurse Educ Today.* 2004;24(2):105-12.
44. Nicklin J, Cramp F, Kirwan J, Urban M, Hewlett S. 2010. Collaboration with patients in the design of patient-reported outcome measures: capturing the experience of fatigue in rheumatoid arthritis. *Arthritis Care Res.* 2010;62(11):1552-8.
45. Sanderson T, Hewlett S, Flurey C, Dures E, Richards P, Kirwan J. The impact triad (severity, importance and self-management) as a method of enhancing the measurement of the personal life impact of rheumatic diseases (Editorial). *J Rheumatol.* 2011;38(2):191-4.

46. Collins D. Pretesting survey instruments: An overview of cognitive methods. *Qual Life Res.* 2008;12(2):229-38.

47. Hewlett S, Wit M, Richards P, Quest E, Hughes R, Heiberg T, Kirwan J. Patients and professionals as research partners: Challenges, practicalities, and benefits. *Arthritis Rheum*, 2006; 55: 676–80.

Legends

Table 1: Participant characteristics for Phases 1 and 2

Table 2. Characteristic of 15 experienced patient partners in Phase 1. (Three additional patients were partners for the first time.)

Table 3: Existing scales presented to partners in Phase 1

Table 4: Phase 1 consensus results on existing scales and the need for new NRS, supported by example quotations

Table 5: Phase 2 focus group feedback on draft NRS

Figure 1: Topic guide for Phase 2 discussions

Figure 2. Final RAPP-PI NRS

Table 1: Participant characteristics for Phases 1 and 2

Variable	Consultation meetings (Phase I) (N=18)	Focus groups (Phase II) (N=8)
Age in years (S.D.)	55.3 (12.2)	64.7 (8.2)
Female, N (%)	15 (78.9%)	6 (66.7%)
Symptom duration in years (S.D.)	19.9 (10.5)	12.8 (11.8)
Patient global NRS* (S.D.)		4.0 (2.5)
Working (Part-time or full-time), N (%)	7 (36.8%)	2 (22.2%)
Educational attainment, N (%)		
No qualifications		2 (25.0%)
School qualifications (16 years of age)		3 (37.5%)
University qualification		3 (37.5%)

*Numerical Rating Scale (100mm, 0=very well)

Table 2. Characteristic of 15 experienced patient partners in Phase 1. (Three additional patients were partners for the first time.)

	Patient														
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Time period of involvement (Years)	12	5	11	11	5	10	5	3	2	1	4	2	1	3	7
Number of projects Contributions	15	12	7	7	3	5	4	3	1	1	2	1	1	2	3
(Number of projects with each type of involvement)															
Advice and support	5	8	4	5	2	5	3	2	1	1	1	1	1	1	1
Grant applicant	7	5	4	1	1		2		1	1	2			1	
Contributions to protocol	4	3	5	2	2	1		2	1	1	2	1	1	1	2
Steering group	2	4	2	1	2	1	1	2	1		1	1	1	1	
Analysed results	4	3	1	1	2	1	1		1	1			1		
Interpretation of results	6	3	2	2	2	1	1	2	1	1		1			
Co-author papers	4	5	2	1	2	1	1		1	1		1			
Presentation of results	3			1											
Notes*	I,P, S	I,P, S	P,S		I		I,P, S								

*I=International participant; P=Given presentations on patient partnership; S=Has had the role of supporter to new patient partners.

Table 3: Existing scales presented to partners in Phase 1

RAPP-PI priority	Scale id	Name/ Type of scale	Source
Pain	A	Pain VAS ^{1*}	Used in clinic
	B	RAID ² Pain NRS ³	[16]
	C	Pain Verbal Descriptive Scale (PVD)	[25]
	D	RA ⁴ Pain Scale (an NRS ³)	[26]
	E	MDHAQ ⁵ Pain item (an NRS ³)	[27]
	F	Pain NRS ^{3**}	Used in colleague's clinic
Activities of daily living (ADL)	A	VAS ¹ -Function	[28]
	B	RAID ² Essential activities of daily living NRS ³	[16]
	C	WOMAC ⁶ Osteoarthritis Function VAS ¹	[29]
	D	WHOQOL-BREF ⁷ daily activities item	[30]
	E	WHOQOL ⁸ routine activities item	[31]
	F	WHOQOL ⁸ everyday activities item	[31]
Mobility	A	WHOQOL-BREF ⁷ able to get around item	[30]
	B	AIMS2 ⁹ mobility subscale	[32]
Life enjoyment	A	WHOQOL-BREF ⁷ enjoy life item	[30]
	B	Overall life VAS ¹	[33]
	C	SWLS ¹⁰ (5 items)	[33]
Independence	A	WHOQOL ⁸ independence item	[31]
	B	AOI ¹¹ scale burden item	[34]
	C	RAQOL ¹² questionnaire dependence item AHI ¹³ helpless item	[35]
	D		[36]
Fatigue	A	BRAF ¹⁴ 3 NRS ³	[37]
	B	RAID ² Fatigue NRS ³	[16]
	C	WHOQOL ⁸ fatigue item	[31]
	D	Fatigue NRS ³ (past week)	[38]
	E	Fatigue VAS ¹ (past week)	[39]
Valued activities	A	WHOQOL ⁸ enjoy free time item	[31]
	B	RAQOL ¹² activities item	[35]
	C	SF-36 ¹⁵ social activities subscale	[40]

	D	AIMS2 ⁹ social activity subscale	[32]
	E	AII ¹⁶ scale – 4 items	[41]
Perceived joint damage		3 Draft NRS from interview data***	[17]

1VAS=Visual Analogue Scale; 2 RAID=Rheumatoid Arthritis Impact of Disease scale; 3 NRS=Numerical Rating Scale; 4=Rheumatoid Arthritis; 5 MDHAQ=Multidisciplinary Health Assessment Questionnaire; 6 WOMAC= Western Ontario and McMaster Universities Arthritis Index; 7 WHOQOL-BREF=World Health Organisation Quality of Life scale- Brief Version; 8= WHOQOL= World Health Organisation Quality of Life scale; 9 AIMS2=Arthritis Impact Measurement Scale version 2; 10 SWLS=Satisfaction With Life Scale; 11 AOI= Acceptance of Illness scale; 12 RAQOL=Rheumatoid Arthritis Quality of Life questionnaire; 13 AHI=Arthritis Helplessness Index; 14 BRAF=Bristol Royal Infirmary Fatigue; 15 SF-36= Short Form (36) Health Survey; 16 AII= Adapted Illness Intrusiveness scale.

* Please mark on the line your level of pain in the past week (VAS), Anchors: No pain (0) – Pain as bad as it could be (100mm); ** How would you rate your overall level of pain?, Anchors: Best possible (0) – Worst possible (10); *** A) Please circle the number which shows the change you have seen in the shape of your hands or feet due to your RA in the PAST YEAR; B) Please circle the number which shows how upset you feel about the joint damage that has happened because of your RA in the PAST YEAR; C) Please circle the number which shows how worried you are about joint damage occurring because of your RA in the NEXT YEAR

Table 4: Phase 1 consensus results on existing scales and the need for new NRS with three domains, supported by example quotations

Priority	Scales (see Table 3 for details)	Votes*	Justification for consensus decision - Example comments/ quotations from booklets and discussion (Scale id: “Quotation” (Participant id))
Pain	A-F	B=12 D=4 New=1	B: “Includes condition in the title” (Consultation Meeting (CM) 3) B: “I like this, non-complicated” (CM1) C: “Too open to different interpretations” (CM16) D: “No timescale is given” (CM4)
Activities of daily living (ADL)	A-F	B=12 E=2 New=4	B: “Straightforward wording, “daily activities” is good. “Activity” is better word than “function” as latter isn’t an everyday word” (CM6) B: “Good timescale” (CM4) B: “Easier to understand” (CM8) E: “This is a good one because it’s specific” (CM3)
Mobility	A-B	New=17	A: “Not specific enough: is it generally, on a bad day, or on a good day?” (CM7) A: “Mobility is a lot more than just walking” (CM12) B: “It’s over-complex and unclear: is this about mobility or movement or specific activity (CM2)
Life enjoyment	A-C	New=17	A: “The question is too personal / intrusive” (CM7) B: “I don’t like the wording, especially “these days”” (CM3) C: “Too complicated and visually messy” (CM1) “None of the questions mention RA” (CM16)
Independence	A-D	New=15	A: “What does “neither poor nor good mean”? You would have to change the scale to totally dependent to totally independent” (CM2) B: “It’s brutal and negative! I don’t like the word “burden”, it’s very negative” (CM4) C: “Not helpful because includes whole of normality and state of mind.” (CM 12) D: “Don’t like “helpless”. “Helpless” is very negative. I don’t like it” (CM4)
Fatigue	A-E	A=13	A: “Easily understood and consistent. Tackles several issues at once” (CM3)

		D=2	A: "I prefer this one: what it's like and how you're coping with it" (CM4) C: "Don't like "bothered"" (CM7) D: "Confusing as says "fatigue or tiredness" – they're too separate issues (CM4)
Valued activities	A-E	C=1 E=2 New=12	B: "Don't like this question – depressing!" (CM8) C: "Covers physical and emotional, but only refers to social activities" (CM15) E: "Quite like this one, but don't like the scoring options" (CM4) E: "All worded wrong, confusing and long" (CM3)
Perceived joint damage	Draft scales A-C	B=1 Edit=14	A: "But is this actual damage or feelings about damage?" (CM9) B: "Is this different to joint counts and x-rays?" (CM7) B: "Possible confusing if ask non-erosive patients this question" (CM18) C: "More important is the need/ worry of surgery and the effect on the family and carer post-surgery" (CM3)
Distinction and interpretation of the three domains (level, effect & coping)			““If making lunch or going out for the day, it [pain] will have different effects. So you could have the same level of pain, it could be as bad, but depends what I'm doing what effect it has.” (CM6) [Level as severity, differentiated from effect] “A more important issue [than level] is asking about the impact of, for example, joint damage i.e. how it has affected your life.” (CM1) [Effect interpreted as impact] “I've coped very well because I've just laid down and read a book. Sometimes you force yourself to do things, but I've really listened to my body this time.” (CM14) [Coping as self-management] “It's the difference between coping with something, a physical or mental coping, and 'having' something, the level of fatigue.” (CM7) [Coping as distinct from level]

*Pain, ADL, mobility and life enjoyment were covered in meeting 1 (N=17) and independence, fatigue, valued activities and perceived joint damage at meeting 2 (N=15).

Table 5: Phase 2 focus group feedback on draft NRS

Concept	Example quotations (Participant id)
Stem question	<p><u>Mobility NRS</u>: “Mobility means, to me, trying to get from room to room, or out to the car” (FG3)</p> <p>“So it’s moving around, it’s getting around, isn’t it?” (FG6)</p> <p>“But getting around is outside and moving around is inside” (FG3)</p> <p>“Trouble is, you have things like mobility scooters which take you from A to B” (FG2)</p> <p><u>Valued activities NRS</u>: “Before I had RA, when I cleaned the windows of the house, I used to do all the windows. Now if I did that I couldn’t lift my arms by the evening.” (FG1)</p> <p>“I think this needs to say “describe the difficulty with doing activities you currently enjoy” (FG5)</p> <p><u>Joint damage NRS</u>: “When you say “How much have your joints changed?”, is it inevitably “got worse”? You wouldn’t know whether it’s negative or positive.” (FG5)</p> <p>“Also problems with the plural. I’ve got one damaged joint” (FG2)</p> <p>“Would you be upset if you saw “joint damage”? (TS)</p> <p>“No. You have to face up to it, that’s what’s happened.” (FG10)</p> <p>“If anybody says “it’s upsetting if you call it damage”, get a life. It’s what we live with” (FG5)</p>
Anchors	<p><u>Add examples</u>: “No, the more text there is, the more I will skip” (FG5)</p> <p><u>Reverse scoring for coping questions</u>: “I had understood it, yeah” (FG7)</p> <p>“You do have to be conscious that you’re [switching]” (FG7)</p> <p>“You have to pay attention. Or you put a note at the top, the way it is scored.” (FG10)</p> <p>“I got the gist of it.” (FG9)</p>
Timeframe	<p><u>Past 7 days</u>: “Over the last 7 days you might have had a bad day, you might have had a good day...” (FG4)</p> <p>“Yeah, you’ve gotta take an average, I suppose” (FG3)</p> <p>“Cos if you do it daily, you could have had a good day, but the day before that you could have been terrible” (FG1)</p> <p><u>Valued activities</u>: “That’s really contextual. There’s loads of stuff I used to do 20 years ago. I think it needs to say “currently”, what your current expectations is, how that’s changed” (FG5)</p> <p><u>Joint damage</u>: “Hopefully you noticed it’s “over the last year” (TS)</p> <p>“Yes, ‘cos you underlined it” (FG1)</p>
Layout	<p><u>Emphasis</u>: “Yeah have it in bold or you could underline it. Pick out the key word like “effect” on number 2” (FG3)</p> <p>“Or shading sections differently. Here is the next section because its shading is different.” (FG5)</p> <p>“Why don’t you put instructions in bullet points to break up the text?” (FG6)</p> <p><u>By domain or component</u>: “I think by domain personally.” (FG9)</p>

Support for
priorities

“It was really useful having the difference between pain and daily physical activities. I’m not getting any pain as such. But my daily, the physical activities, are a bit thing.” (FG8)

“The emotional side of it all. The frustration and unhappiness really when you used to do it [fine sewing].” (FG8)

“I think that’s coming up in this Life Enjoyment” (FG7)

“And in Valued activities” (FG5)

Figure 1: Focus group topic guide

For each draft NRS (constructed from data in Phase 1):

- What do you think of the wording of the question?
- Are the anchors (the words describing each end of the scale) appropriate?
- Is this a suitable time frame for the question?
- How easily would you be able to answer this question?
- Are the three aspects of each domain severity, effect and coping necessary?

Figure 2: Final RAPP-PI NRS

In this section we are trying to understand the different aspects of your experience of RA.

There are three questions for each area:

- the severity (e.g. of pain);
- the effect it has had on your life;
- how well you have coped with it.

*****The coping question has the opposite scoring, with the worst score on the left side.*****

A) Pain in RA:

1. What level of pain did you feel on average over the past 7 days:

None 0 1 2 3 4 5 6 7 8 9 10 **Extreme**

2. What effect has the pain had on your life during the last 7 days?

No effect 0 1 2 3 4 5 6 7 8 9 10 **A great deal of effect**

3. How have you coped with the pain during the last 7 days?

Not at all 0 1 2 3 4 5 6 7 8 9 10 **Very well**

B) Daily physical activities and RA:

1. How difficult on average were daily physical activities over the past 7 days?

No difficulty 0 1 2 3 4 5 6 7 8 9 10 **Extreme difficulty**

2. What effect has your difficulty in doing daily physical activities had on your life over the past 7 days?

No effect 0 1 2 3 4 5 6 7 8 9 10 **A great deal of effect**

3. How have you coped with the difficulty in doing daily physical activities over the past 7 days?

Not at all 0 1 2 3 4 5 6 7 8 9 10 **Very well**

C) Getting around with RA (outside the house):

1. What level of difficulty did you have in getting around over the past 7 days?

No difficulty 0 1 2 3 4 5 6 7 8 9 10 **Extreme difficulty**

2. What effect has your difficulty in getting around had on your life over the past 7 days?

No effect 0 1 2 3 4 5 6 7 8 9 10 **A great deal of effect**

3. How have you coped with your difficulty in getting around over the past 7 days?

Not at all 0 1 2 3 4 5 6 7 8 9 10 Very well

D) Enjoying life with / despite RA

1. Considering your rheumatoid arthritis, how much have you enjoyed life over the past 7 days?

Not at all 0 1 2 3 4 5 6 7 8 9 10 Extremely enjoyed

2. What effect have disruptions to your enjoyment had on your life over the past 7 days?

No effect 0 1 2 3 4 5 6 7 8 9 10 **A great deal of effect**

3. How have you coped with the disruptions to your enjoyment of life over the past 7 days?

Not at all 0 1 2 3 4 5 6 7 8 9 10 Very well

E) Independence and RA:

1. How much did rheumatoid arthritis disrupt your independence over the past 7 days?

No disruption 0 1 2 3 4 5 6 7 8 9 10 **Greatly disrupted**

2. What effect have disruptions to your independence had on your life over the past 7 days?

No effect 0 1 2 3 4 5 6 7 8 9 10 **A great deal of effect**

3. How have you coped with the disruptions to your independence over the past 7 days?

Not at all 0 1 2 3 4 5 6 7 8 9 10 Very well

F) Fatigue in RA:

1. Please circle the number which shows your average level of fatigue during the past 7 days

No fatigue 0 1 2 3 4 5 6 7 8 9 10 **Totally exhausted**

2. Please circle the number which shows the effect fatigue has had on your life during the past 7 days

No effect 0 1 2 3 4 5 6 7 8 9 10 **A great deal of effect**

3. Please circle the number which shows how well you have coped with fatigue over the past 7 days

Not at all 0 1 2 3 4 5 6 7 8 9 10 Very well

G) Activities you enjoy and RA:

1. What level of difficulty did you experience with doing activities you currently enjoy over the past 7 days?

No difficulty 0 1 2 3 4 5 6 7 8 9 10 **Extreme difficulty**

2. What effect has your difficulty in doing activities you currently enjoy had on your life over the past 7 days?

No effect 0 1 2 3 4 5 6 7 8 9 10 **A great deal of effect**

3. How have you have coped with your difficulty in doing activities you currently enjoy over the past 7 days?

Not at all 0 1 2 3 4 5 6 7 8 9 10 Very well

H) Joint damage because of RA:

1. What level of new joint damage have you seen in your worst joint(s) over the last year?

No new visible 0 1 2 3 4 5 6 7 8 9 10 **Great new visible**
damage damage

2. What effect has joint damage had on your life over the past 7 days?

No effect 0 1 2 3 4 5 6 7 8 9 10 **A great deal of effect**

3. How have you coped with the effect of joint damage over the past 7 days?

Not at all 0 1 2 3 4 5 6 7 8 9 10 Very well