Assessing the content validity of patient reported outcome measure item pool for rheumatoid arthritis disease activity using cognitive interviews

Tim Pickles[1], Rhiannon Phillips[2], Mike Horton[3], David Gillespie[1], Janice Davies[4], Susan Campbell[4], Ernest Choy[5]

[1]Centre for Trials Research, Cardiff University, Cardiff, UK;

[2]Cardiff School of Sport and Health Sciences, Cardiff Metropolitan University, Cardiff, UK;

[3]Psychometric Laboratory for Health Sciences, University of Leeds, Leeds, UK;

[4]Patient and Public Involvement Stakeholders, UK;

[5]Division of Infection and Immunity, Cardiff University, Cardiff, UK;

Corresponding author details: Tim Pickles

Centre for Trials Research, Cardiff University, 5th Floor, Neuadd Meirionnydd, Heath Park, Cardiff, CF14 4YS, UK

Email: PicklesTE@cardiff.ac.uk

Abstract

Objectives. Using Rasch measurement theory, an item pool of 12 questions has been identified, covering tenderness and swelling, disease activity, pain, physical functioning and stiffness for assessing the construct of RA disease activity. This study aimed to assess the content validity of this item pool using cognitive interviews.

Methods. Participants were randomly sampled across varying age, sex and education level categories from respondents to a survey containing RA disease activity Patient Reported Outcome Measures. The 'think aloud' technique was used to understand how individual items were interpreted and answered the 12 items of the item pool plus 12 additional relevant items that could potentially be included. Participants were asked about relevance, comprehensiveness, comprehensibility of the set of items overall. Reflexive thematic analysis was used to identify these broader aspects of content validity. Participants were also asked whether they could distinguish RA symptoms from those of other conditions, whether tenderness and swelling should be assessed together or separately, and whether they felt that fatigue and general health were separate issues to their disease activity. Content analysis was used for this section of the interviews.

Results. Twenty participants completed one-to-one cognitive interviews between November 2022 to February 2023. No participants raised concerns relating to comprehensiveness, comprehensibility or relevance aspects of items. There was a lack of consensus on the ease of distinguishing RA symptoms from other conditions, or whether tenderness and swelling should be asked about in a single item or as separate symptoms. The majority view was that fatigue and general health were not specific to RA disease activity.

Conclusion. The findings indicate that the 12-item pool adequately captures relevant concepts of RA disease activity, with no additional items required and thus provide evidence of its content validity. Future research of content validity will be needed for any new RA disease activity items.

Keywords: RA Disease Activity, Patient Reported Outcome Measures, Content Validity

Key messages:

- There were no items or concepts missing that should be covered.
- There was some initial evidence of content validity for these items.
- General health and Fatigue are separate to the construct of RA disease activity.

Introduction

RA is chronic autoimmune condition that causes multiple symptoms, such as pain, tenderness, swelling and stiffness. The current standard of care in RA is 'Treat-to-Target' (T2T), which aims for sustained remission or failing this, low disease activity score. Regular assessment of DA and adjustment of treatment accordingly is an integral part of T2T. Clinical outcome measures for assessing disease activity are Disease Activity Score (DAS) with 28-joint count (DAS28), (1) Simple Disease Activity Index (SDAI) and Clinical Disease Activity Index (CDAI). (2) Adjunct to these clinical outcome measures are Patient Reported Outcome Measures (PROMs) for the construct of RA disease activity. These PROMs, by their nature, can only include items that people living with RA can report on. Therefore, legacy RA disease activity PROMs (e.g., RADAI5, RADAI, RADAI-SF, PDAS2, PRO-CLARA, GAS, PAS, PAS-II, RAPID3, RAPID4 and PROM-score) contain items that cover the tenderness and swelling, disease activity, general health, pain, fatigue, physical functioning and stiffness domains.

A previous study (3) analysed data from surveys contain legacy RA disease activity PROMs along with foot-specific RADAI-F5, RA flare PROMs FLARE-RA and RA-FQ. Assessing structural validity through Rasch measurement theory and confirmatory factor analysis bifactor models, the study found that 12 items covering the domains of *Pain*, *Tenderness and swelling*, *Stiffness*, *Disease activity* and *Physical functioning* can be used to form an item pool to measure the construct of RA disease activity. The finding also showed that *Fatigue* and *General health* domain items measured a separate construct to the construct of RA disease activity.

Applying these results onto Wilson and Cleary's conceptual model (4) RA can be viewed as the biological and physiological variable that causes fluctuation of symptom status in pain, tenderness, swelling and stiffness, and these in turn have an impact on physical status. The construct of RA disease activity therefore sits between the symptom status and functional status boxes of this model and is therefore modified by an individual's symptom amplification and personality motivation and has psychological and social and economic support environmental characteristics. Whilst the previous study was able to provide evidence for the structural validity and internal consistency of the item pool, other important measurement properties were not assessed. One of these was content validity, which is "the degree to which the content of an HR-PRO instrument is an adequate reflection of the construct to be measured." (5) The aim of this research was to determine the content validity of the 12 items of the item pool.

Methods

Study design

Cognitive interviewing (6-8) and the think aloud technique (9) were used to assess comprehensiveness, comprehensibility and relevance and identify other issues relating to content validity. The COSMIN guidelines (10) were used as a framework to design the topic guides and the approach to analyses. This research is reported in line with the Consolidated Criteria for Reporting Qualitative Studies (COREQ) framework (see Supplementary Table S1). (11) The study was approved by the North West—Preston Research Ethics Committee (20/NW/0039).

Participants and recruitment

Twenty interviews were planned. This number of participants was within recommended boundaries (8) and gave information power sufficiency. (12) A sampling framework (Supplementary Table S2) (13) with gender, age group and education level categories was devised to create maximal variation, with number in each category based on the characteristics of the population from which participants could be sampled from (3) to allow for random stratified sampling strategy. Participants who took part in previous study (3) and consented to be re-contacted for future interviews and could provide their contact details formed the pool of participants. Participants registered in Cardiff and Vale University Health Board who responded thus were contacted in accordance with the sampling framework. Recruitment for the interviews took place between November 2022 and February 2023. Contact was made with the potential participants by mobile or landline telephone call to organise the interviews.

Items

Participants completed the 12 items identified based on previous application of Rasch measurement theory and confirmatory factor analysis bifactor models (3): *Tenderness and swelling, Disease activity, Pain, Physical functioning and Stiffness* domain items in the item pool for the construct of RA disease activity. In addition, participants completed six further items from the *General health* and *Fatigue* domain items and six *Pain* and *General health* domain items that were close to inclusion in the item pool based on the prior statistical analyses (Table 1 and Supplementary Data S1) to discover whether these items might affect face validity of the item pool and/or provide additional contextual information.

Topic guide

A topic guide was developed with prompts for the think aloud section, questions that asked about general feedback, comprehensiveness, comprehensibility and relevance (Supplementary Data S2)(10). Specific questions were asked about:

- response option format preference;
- how easy participants found it to tease out RA symptoms when answering the items;
- whether it is sensible to ask about *Tenderness and swelling* in a single item, rather than as separate items;
- whether they thought that *General health* and *Fatigue* were separate from their disease activity.

Participants were asked about response option format preference. This was because the 24 items were drawn from multiple legacy PROMs, which used multiple different response option formats (Supplementary Data S1).

Participants were asked about whether it is sensible to ask about *Tenderness and swelling* in a single item because tenderness and swelling are separate domains defined by Outcomes Measures in Rheumatology (OMERACT) (14) and assessed separately in clinic using the homunculi (minus the knees, ankles and feet).

Cognitive interview

Each participant provided written informed consent prior to their interview.

To evaluate how people formed responses to the 24 items and where problems may arise in any aspect of the items, cognitive interviewing (6-8) and think aloud techniques (9) were used to encourage the participants to provide their thoughts as they answered the items. Participants worked through the 24 items at their own pace and were invited to provide their opinions as they did so, with T.P. using the topic guide prompts as necessary to investigate issues relating to the items and their content validity. Following this, there was a broader discussion based on the questions in the topic guide. Interviews were offered to take place either in-person at the participants' residence or online over Zoom. (15) It has been shown that in-person and online interviews are equally valid and appropriate methods, (16) so there are no implications for data quality. Field notes were taken during interviews to remind of salient points made by participants and aid with the analyses.

Reflexivity

Interviews were conducted by T. P., a research fellow in statistics and PhD student at the Centre for Trials Research, Cardiff University. T.P. undertook the Introduction to Qualitative Interviewing and Introduction to Analysing Qualitative Interviews courses organised by the Medical Sociology & Health Experiences Research Group at the Nuffield Department of Primary Care Health Sciences, University of Oxford in June 2021. T.P. was a 35- and 36-year-old white heterosexual cis-gender male, with no lived experience of RA or any of the medicines involved in the treatment of RA. T. P. was involved in the recruitment of all participants. T.P. had no previous relationship with any participants ahead of the study. Given the wide range of ages of participants, and the majority of females, and the lack of lived experience of RA by the interviewer, there was a possibility that T.P.'s perspectives might have influenced how he related to participants and interpreted the data. To minimise this, patient and public involvement (PPI) stakeholders (who were involved with whole process of this research and have lived experience of RA and the treatments that it entails), plus a broader supervisory team with knowledge of qualitative research, rheumatology and measurement research, were involved in the interpretation of the results.

Analyses

Reflexive thematic analysis, using the approach described by Braun and Clarke (17) was used for the analyses of data from the think aloud section, with themes devised inductively. For the broader discussion, COSMIN define deductive themes of comprehensiveness, comprehensibility and relevance, so we looked for data that fell into those themes. Beyond this, further inductive sub-themes were created within the themes using Braun and Clarke's approach. (17)

Content analysis (18) was used for the specific questions listed above (response option format preference, how easy participants found it to tease out RA symptoms when answering the items, whether it is sensible to ask about Tenderness and swelling in a single item and whether they thought that General health and Fatigue were separate from their disease activity). Responses were categorised such that counts of these could be tabulated. This approach, rather than the thematic approach above, was used as these questions were much more structured and likely to result in simple "Yes" or "No" responses. This approach allowed for the frequency of responses to be assessed, which was useful in this context.

Interviews were audio-recorded and transcribed verbatim. T.P. independently coded the data: there was no dual coding of the data. Instead we used regular team meetings to discuss data production, the development of the coding framework and data analysis, with all authors adding their own unique perspective to the analysis through these meetings. The authors had a mix of skills and contained methodologists, a health psychologist, experts by lived experience and a consultant

rheumatologist. This approach has been identified as appropriate in qualitative research. (19) NVivo software was used to facilitate analysis. (20) Transcripts were not returned to participants for comment or correction, nor did participants provide feedback on findings.

Results

Scheduling

Twenty interviews with people living with RA took place between November 2022 and February 2023. Five interviews took place online at the participants' request (due to illness and work schedules) and 15 took place in participants' residences. This required contact with 31 potential participants (see Supplementary Table S3).

Demographics

The demographics of the sample followed the sampling framework (Table 2 and Supplementary Table S4). The minimum interview time was 17 minutes, and the maximum was 71 minutes, with a median time of 31 minutes 30 seconds. In three cases, partners of the participant were also present for the interview.

Think aloud

There were four themes created from the issues raised by participants during the think aloud part of the interviews: **Variability**, **Difficulty with instructions**, **Difficulty with response options** and **Teasing out RA symptoms** (Table 3 and Supplementary Table S5). Under the **Variability** theme, one participant thought that the wording of items `P07` and `F01` suggested that the pain was at a constant level throughout the past week, whilst, to them, it fluctuates throughout the day and across the week. A separate participant stated that responding to `PS3` was dependent on time of day. **Difficulty with instructions** applied across the *Fatigue*, *Disease Activity*, *Physical functioning* and *General health* domains. For *Fatigue* domain item `RF1`, the word 'unusual' caused confusion. For *Disease activity* domain item `PS1`, there was uncertainty about the term 'how active'. For *Physical functioning* domain item `F02`, there was uncertainty over which activities to consider. For *General health* domain item `R05`, the use of 'at this time' was confusing. For **difficulty with response options**, there was some difficulty deciding which number to pick, especially with 21 categories in *Pain* domain item `R04` and *General health* domain item `P01`. For *Fatigue* domain item `F03`, one participant found it difficult to narrow down their response. For **teasing out RA symptoms**, four items (`F01`, `F03`, `F04` and `F05`) all specify "due to your rheumatoid arthritis" in the instructions and lead to some uncertainty. Participant 13905 did not give a response to *Fatigue* domain item `F03`.

This was the only occasion when the participant did not give a response to any item. The participant was uncertain if their fatigue was due to RA as they also had breast cancer. They wanted to be able provide a 'I don't know' response and therefore did not answer F03.

Comprehensiveness

Within the **comprehensiveness** theme, there were two sub-themes identified: further things to ask beyond the construct of RA disease activity and affects response, both of which have sub-sub-themes within them (Table 4). It is important to note that the further things to ask beyond the construct of RA disease activity sub-theme was centred entirely around participants wanting to provide further information. No participants suggested that there were any concepts or items missing.

Comprehensibility

Within the **comprehensibility** theme, there were sub-themes focussing on instructions and response options, with a further general sub-theme. Only the instructions sub-theme had multiple sub-sub-themes within it (Table 4). For instructions, there were comments about *General health* domain item R05, the use of the term disease activity and whether general health includes mental health. For response options, a participant felt the anchors were the wrong way round. There was also a desire for bigger print in a general sub-theme.

Relevance

There were sub-themes relating to recall period and response options, along with a further things to ask sub-theme. There were multiple sub-sub-themes only within the further things to ask sub-theme (Table 4).There were multiple comments that suggested using longer recall periods, but none shorter than those used in the items of a week, a day and now. Reflecting back to the **teasing out RA** theme from the think aloud sections, there was a further comment from the same participant about the need for a 'don't know' option. Participants also mentioned the need to cover more on day-to-day changes and stiffness.

Other cognitive aspects

There was a need for an **other cognitive aspects** theme to group together some further sub-themes. These sub-themes had topics of instructions, ease of completion and how to decide on a response. The first two were devised from single participant quotes but there were multiple examples of participants struggling to decide on a response for various different reasons (Table 4). In thinking

about the instructions of a disease activity item, one participant stated that, to them disease activity and pain are the same. Another participant noted that they found it straightforward to complete items due having inactive disease, and that it would be harder for those with more active disease to decide how to answer. Linked to this, there were comments about finding it hard to decide how to rank on a scale, and to make decisions, especially regarding pain.

Content analysis

There was no clear response option preference or on the ability to tease out RA symptoms from those caused by other conditions (Table 5). For asking about *Tenderness and swelling* in a single item, the most common response was "Yes." On asking participants whether they felt that *General health* and *Fatigue* were separate from their disease activity, the most common response was "Yes, separate."

Discussion

Whilst there were issues with some items, such as variability of response relating to the time of day and teasing out RA symptoms, all items of the item pool could be understood and completed. One participant could not answer one *Fatigue* domain item (not in the item pool) and requested a don't know option. This indicates that the item pool has adequate comprehensiveness, comprehensibility and relevance for the construct of RA disease activity, with no additional items required.

There was a lack of understanding towards a *General health* domain item and questions from participants about whether to include mental health in response to a *General health* domain item. Additionally, there was difficulty answering a *Fatigue* domain item. Furthermore, the responses given to the specific question on *General health* and *Fatigue* indicated that *General health* and *Fatigue* were separate from their RA disease activity. This backs up the findings of the previous study (3) that showed that *General health* and *Fatigue* domain items should not be used to measure the construct of RA disease activity.

Whilst the items were deemed comprehensive in terms of measuring the construct of RA disease activity, participants suggested that they are keen to provide more detail so that nuances of their individual experiences to be communicated with their healthcare teams. There was also considerable variability between participants in how they responded to some of the items. Participants formed complex representations about what their symptoms were, what was causing them, what the timeframe and pattern of symptoms were, and their broader impact on aspects like mental health and side effects. The notion that time of day could affect response was mentioned multiple times and links to the **variability** theme. For any person living with RA, there will be varying,

generally diurnal, variations in their disease activity that will repeat from day-to-day. This is a well-known pattern for stiffness, especially regarding early morning stiffness, and pain. (21) The notion of sedentary and mobile is linked to early morning stiffness, as that stiffness comes from being sedentary through the night. If a participant is sedentary for periods of the day, then there is a possibility of further periods of stiffness. We can't know what bias this may introduce into the cognitive process of providing a response to such items. A consideration for the future would be suggesting that, if there is a situation of completing such items regularly over time, the respondent does so at roughly the same time of the day.

The suggestions by participants could be optional items outside of the item pool, which could include disease activity-specific concepts, such as addenda to *Pain*, *Tenderness and swelling*, *Physical functioning* and *Stiffness* domain items to provide more specific information. Participants also suggested adding a question on whether people living with RA felt like their RA disease activity had improved or deteriorated over the last year. They also suggested collecting information relating to co-morbidities like mental health conditions, and also information relating to the treatment of RA in terms of side effects.

Comparison with previous literature

There is limited literature to compare these results to. A systematic review of RA disease activity PROMs (22) identified just two content validity studies: one for PRO-CLARA (23) and one for PDAS2. (24) Both assessed content validity using quantitative methods, collecting data via a survey, and singularly assessed the comprehensibility aspect of content validity.

This is the first time any qualitative research on content validity has been done in the field of RA disease activity PROMs. Past attempts to elicit the opinions of people living with RA on the content validity of legacy RA disease activity PROMs has been from surveys. (23, 24) For the purposes of assessing content validity, interviewing and qualitative research methods allow for more detailed exploration of participants' perspectives and experiences, providing richer insights than surveys alone.

Patient and Public Involvement

J.D. and S.C., both people living with RA, were integral throughout and, along with the other authors' range of knowledge and experience, improved the quality and interpretation of the analyses.

Limitations

It was possible that including the *General health* and *Fatigue* domain items as part of the 24 items was a limitation. Participants may have confused the comprehensiveness questions that were later asked. if they thought that those items were also included for the measurement of the construct of RA disease activity.

A further limitation was not discussing the planned future use of these items as part of the interview. This could potentially have had an effect on participants' responses and would have added to the relevance discussion. Wales is 96% white ethnically, which is similar to the 95% of participants in this sample. Despite this similarity, the population of people living with RA is far more ethnically diverse and the views across that diversity are not available here.

That T.P. did all the interviews, coding and analyses in this study means that there was inherent consistency in how these were done. The clear drawbacks were a narrow perspective across the analyses, plus T.P.'s limited experience in qualitative researcher and non-similar experiences to those of the participants. This was balanced by the involvement of all authors to the development of the coding framework and interpretation of the data. We therefore consider that this limitation would only have a minimal influence.

Conclusion

There was evidence of content validity for this item pool, even though there were some limitations. Future research of content validity will be needed for any new RA disease activity items and address the limitations noted above by increasing diversity, discussing content and ensuring a broad group are involved in coding, analysing and interpreting the data.

Acknowledgements: Sarah Rawlinson, Laura Tilley and Penny Farthing from CTR for help regarding recording, transcription and compensation.

Funding: T.P. was supported by a National Institute of Health Research Doctoral Fellowship, funded by the Welsh Government through Health and Care Research Wales (NIHR-FS-19).

Conflict of interest statement: E.C. has received research grants from Bio-Cancer, Biogen, Pfizer and Sanofi, and honoraria/or served as member of speakers' bureaus from Abbvie, Bio-Cancer, Biocon, Biogen, Eli Lilly, Fresenius Kai, Galapagos, Janssen, Pfizer, Sanofi, UCB and Viatris. EC also receives a stipend as Editor-in-chief of *Rheumatology* (Oxford); he has not peer reviewed or made any editorial decisions for this article. The remaining authors have declared no conflicts of interest.

Data availability statement: Data can be made available on request to the Centre for Trials Research https://www.cardiff.ac.uk/centre-for-trials-research/collaborate-with-us/data-requests

References

1. Prevoo ML, van't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. Arthritis Rheum. 1995;38:44-8. https://doi.org/10.1002/art.1780380107
2. Aletaha D, Smolen J. The Simplified Disease Activity Index (SDAI) and the Clinical Disease Activity Index (CDAI): a review of their usefulness and validity in rheumatoid arthritis. Clin Exp Rheumatol. 2005;23(5 Suppl 39):S100-8. https://www.clinexprheumatol.org/abstract.asp?a=2694
3. Pickles T, Horton M, Christensen KB, Phillips R, Gillespie D, Mo N, et al. Patient Reported Outcome Measures for Rheumatoid Arthritis Disease Activity: Rasch measurement theory to identify items and domains. Rheumatology. 2025;, keaf189. https://doi.org/10.1093/rheumatology/keaf189
4. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life: A conceptual model of patient outcomes. JAMA. 1995;273(1):59-65. https://doi.org/10.1001/jama.1995.03520250075037
5. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737-45. https://doi.org/10.1016/j.jclinepi.2010.02.006
6. Willis GB. Cognitive Interviewing: A how to guide. [online] Research Triangle Institute: Raleigh, NC; 1999 [Accessed 17 September 2018]. Available from: https://www.hkr.se/contentassets/9ed7b1b3997e4bf4baa8d4eceed5cd87/gordonwillis.pdf
7. Willis GB. Cognitive interviewing: a tool for improving questionnaire design. Thousand Oaks, CA: SAGE Publications Inc.; 2004. https://doi.org/10.4135/9781412983655
8. Willis GB, Artino AR, Jr. What Do Our Respondents Think We're Asking? Using Cognitive Interviewing to Improve Medical Education Surveys. J Grad Med Educ. 2013;5(3):353-6. https://doi.org/10.4300/JGME-D-13-00154.1
9. Ericsson KA, Simon HA. Protocol Analysis: Verbal Reports as Data. Cambridge, MA: The MIT Press; 1993. https://doi.org/10.7551/mitpress/5657.001.0001

10. Mokkink LB, Elsman EBM, Terwee CB. COSMIN guideline for systematic reviews of patient-reported outcome measures version 2.0. Qual Life Res. 2024;33(11):2929-39. https://doi.org/10.1007/s11136-024-03761-6
11. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care. 2007;19(6):349-57. https://doi.org/10.1093/intqhc/mzm042
12. Malterud K, Siersma VD, Guassora AD. Sample Size in Qualitative Interview Studies: Guided by Information Power. Qual Health Res. 2016;26(13):1753-60. https://doi.org/10.1177/1049732315617444
13. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. Adm Policy Ment Health. 2015;42(5):533-44. https://doi.org/10.1007/s10488-013-0528-y
14. OMERACT. OMERACT Endorsed Core Domains. [online]. OMERACT: Ottawa, ON; 2023 November [Accessed 30 November 2023]. Available from: https://omeract.org/domain-selection/
15. Zoom Video Communications. Zoom (version 5.12.2 (9281)) [Computer software]. https://zoom.us; 2022
16. Peasgood T, Bourke M, Devlin N, Rowen D, Yang Y, Dalziel K. Randomised comparison of online interviews versus face-to-face interviews to value health states. Soc Sci Med. 2023;323:115818. https://doi.org/10.1016/j.socscimed.2023.115818
17. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol. 2006;3(2):77-101. https://doi.org/10.1191/1478088706qp063oa
18. Berelson B. Content Analysis in Communication Research. Glencoe, IL: The Free Press; 1952.
19. Barbour R S. Checklists for improving rigour in qualitative research: a case of the tail wagging the dog? BMJ 2001; 322:1115. https://doi.org/10.1136/bmj.322.7294.1115
20. QSR International. NVivo qualitative data analysis Software 12 [computer programme]. Burlington, MA: QSR International; 2017.
21. Gibbs JE, Ray DW. The role of the circadian clock in rheumatoid arthritis. Arthritis Res Ther. 2013;15(1):205. https://doi.org/10.1186/ar4146
22. Pickles T, Macefield R, Aiyegbusi OL, Beecher C, Horton M, Christensen KB, et al. Patient Reported Outcome Measures for Rheumatoid Arthritis Disease Activity: a systematic review following COSMIN guidelines. RMD Open. 2022;8:e002093. https://doi.org/10.1136/rmdopen-2021-002093

23. Salaffi F, Migliore A, Scarpellini M, Corsaro SM, Lagana B, Mozzani F, et al. Psychometric properties of an index of three patient reported outcome (PRO) measures, termed the CLinical ARthritis Activity (PRO-CLARA) in patients with rheumatoid arthritis. The NEW INDICES study. Clin Exp Rheumatol. 2010;28(2):186-200. https://www.clinexprheumatol.org/abstract.asp?a=543

24. Choy EH, Khoshaba B, Cooper D, MacGregor A, Scott DL. Development and validation of a patient-based disease activity score in rheumatoid arthritis that can be used in clinical trials and routine practice. Arthritis Rheum. 2008;59(2):192-9. https://doi.org/10.1002/art.23342

Table 1: Items used in cognitive interviews

Domain	Item pool	Additional items
Tenderness and swelling	D02, T02, Q04	
Disease activity	PS1, A01	
Pain	F01, R04, P07, Q05	PS3, D03, HE1, T03
Physical functioning	F02, F05	
Stiffness	F04	
	Included items	Additional items
General health	R05, P01, C01	PS2, T04
Fatigue	F03, PF1, RF1	

Table 2: Demographics

		n	%
Gender	Male	6	30
	Female	14	70
Age group	18-54	5	25
	55-74	10	50
	75+	5	25
Education level	Qualifications below university graduate	10	50
	University graduate qualification as minimum	10	50
Ethnicity	White	19	95
	Mixed/multiple ethnic groups	1	5
Other conditions in addition to RA	No	2	10
	Yes	18	90

Table 3: Exemplar quotations for think aloud

Theme	Referring to item	Quotation
Variability	F01	""Put the number that best describes the pain you felt during the last week." (Sighs) You see, the thing with this, it's very bad in the mornings. … But it adjusts in the day. … That's what happens to it. So, er, how can ... How do I answer that, you know?" (11928)
	PS3	"I guess it would probably depend on the time of day, in which I was doing it, as lots of things do." (13054)
Difficulty with instructions	R05	"I've got to ask, read that question again. … Considering all the ways in which your illness and health conditions may affect you at that time, please indicate how you're doing? I don't really understand the question, but I'm very well (laughs) (pause). Yes." (11498)
	R05	"Um, may affect you, affect you at this time. So, I'm not sure what at this time means in this context." (13905)
Difficulty with response options	R04	"I always have these problems in questionnaires like this, you know, is something two? Is it two point five? Three? I mean it's, depends how you, you know, (chuckling) where the pen falls. So, yeah, you, you, I appreciate the nature of such scales." (13054)
	P01	"Right. Oh, gosh, let's stick it randomly in the middle." (11994)
Teasing out RA symptoms	F03	"Um, fatigue. And then, it's interesting now for me because now you have got due to your rheumatoid arthritis, and what I would want to be able to put is I don't know." (13905)
	F05	"Um, I'm not working, because I've had surgery. Um, it's really hard for me to ... I don't know how to answer that one. "The difficulty you've had when taking part in activities." Um, do I do that as activities round the home, as in trying to cook, or, or ... Because I can't answer it as related to work/family life, because I haven't been going out with my knee." (11531)

Table 4: Exemplar quotations for the **comprehensiveness, comprehensibility, relevance** and **other cognitive aspects** themes

Theme**: comprehensiveness**	
Sub theme: <u>further things to ask beyond the construct of RA disease activity</u>	
Addendum to physical functioning to specify activities	"more questions about the physical activities … and your social life and things like that " and "Yeah, listing" (both 13350)
Addenda to pain, stiffness and tenderness and swelling items to state areas affected	"I've got to answer on the whole thing … … which I would've thought was confusing for you … So, if you, if you're working out how to do a questionnaire I'd say make it more specific for each question … … and also can we have a box to add a little addendum." (11017) "Yeah, which joints are affected … " (11531)
Difficulty with diagnosis	"about having difficulty with diagnosis" (11994)
Mental health	"well about mental health" (13350)
Medication	"I think I would have liked to have been asked about medication." (13905)
Day of the week that's worse	"Is there any days of the week where you find' ... Um, because sometimes I find that I get worse towards the end of the week" (11531)
Improved or deteriorated over previous year	"If you compared yourself to this time last year' ... You know: 'Has your condition improved or deteriorated?'" (11531)
How quickly it comes on	"Yeah. I did say about how quickly it comes on. Perhaps that should be in there …" (12073)
Side effects	"Like the side effects kind of" (13315)
Sub theme: <u>affects response</u>	
Time of day	"I suppose with the arthritis, first thing in the morning, I mean problems tend, stiffness tends to be worse first thing in the morning, pain, perhaps later on in the day. So, there could be a time-of-day affect." (13054)

Whether sedentary or mobile	"They all need a lot of thought, don't they? Because ... Maybe they should put sedentary and mobile? Maybe they should split it a little bit that way? … So, yeah ... No, I think there's a big difference between sedentary and mobility." (11928)
Theme: **comprehensibility**	
Sub theme: <u>instructions</u>	
Don't understand *General health* item	"And, I didn't understand the last question. … I still don't understand the last question" (11498)
Disease 'activity' is confusing	"I find it confusing, but like, when you say, Activity, you mean just in general? Like, how it is overall?" (13853)
Does general health include mental health?	"are you counting mental health under general health?" (13853)
Sub theme: <u>response options</u>	
Anchors wrong way round	"actually I think I'd put them the other way round. … I think I'd put bad on the left and good on the right." (11504)
Sub theme: <u>general</u>	
Bigger print/font	"the only thing I would say is that the print could be a bit bigger" (11994) "…the font, a little bit sort of bigger maybe." (11410)
Theme: **relevance**	
Sub theme: <u>recall period</u>	
Longer (e.g., 3 months, 1 month, 2-3 weeks and 2 weeks)	"I'm not sure if a week is enough … And, three months is probably a good one " (11498) "I would've said the parameters need to be say a month. … Yeah, I'd, I'd do it monthly." (11017)

	"I probably would ask as well maybe two or three weeks you know?" (12073) "I think even a fortnight …" (11994)
Sub theme: response options	
Don't know option	"it's difficult to say you don't know to, to a question" (13905)
Sub theme: further things to ask	
More on day-to-day	"I suppose they could've asked more of your day to day and … I mean, today is a bad day for me but another day could've been better." (13359)
More on stiffness	"I didn't feel there was enough to do with how stiff you are I'd say. … Cos I find out of all of it I would say that [pause] it comes in my mind more than pain. … I'd say the stiffness is more important to me than …" (13315)
Theme: **other cognitive aspects**	
Sub theme: instructions	
Activity and pain are the same	"Well, arthritis activity to me is pain … I can't imagine any activity without pain." (11017)
Sub theme: ease of completion	
Easier to complete when disease inactive	"if things are more active, then you've got more things to consider, if they're relatively inactive, then um, you know, it's quite easy to fill and the actual wording the questions, probably doesn't matter so much." (13054)
Sub theme: how to decide on a response	
Difficult to rank	"personally, I never know how to sort of rank it on a numerical scale, um, cos, to be honest I've had some level of chronic pain, since I was fifteen and I'm twenty-eight now … I find it very

	hard to sort of rank it, purely because, I, I don't really remember a baseline to go off" (13853)
Hard to think how you feel	"Sometimes it's hard to think how you're feeling" (13359)
Difficult to decide on pain	"I find it difficult to decide how, what the severity of pain" (13924)

Table 5: Content analysis results

		n	%
Response option format preference	No preference	4	20.0
	21-category numeric rating scale with a number on each category	3.5	17.5
	Any visual analogue scale (with or without marks)	3.5	17.5
	Any 11-category numeric rating scale	3	15.0
	Any numeric rating scale (11 or 21 categories)	3	15.0
	Visual analogue scale with marks	2	10.0
	Any 21-category numeric rating scale	1	5.0
Can you tease out RA symptoms when answering the items?	Yes	8	40.0
	No	8	40.0
	Didn't answer question	3	15.0
	Mixed*	1	5.0
Is it sensible for Tenderness and swelling be asked about in a single item?	Yes	10	50.0
	No	7	35.0
	Mixed	2	10.0
	Didn't answer question	1	5.0
Are *General health* and *Fatigue* separate from your disease activity?	Yes, separate	8	40.0
	No, linked	4	20.0
	Didn't answer question	4	20.0
	General health is linked but fatigue is separate	2	10.0
	Fatigue is separate (didn't mention general health)	1	5.0
	Mixed	1	5.0

*Participant responded Yes, but discussed COPD/emphysema in relation to a RA-specific *Physical functioning* domain item (F02)