RMD Open
Rheumatic & Musculoskeletal Diseases

ORIGINAL RESEARCH

Patient Reported Outcome Measures for Rheumatoid Arthritis Disease Activity: a systematic review following COSMIN guidelines

Tim Pickles ,[1] Rhiannon Macefield ,[2] Olalekan Lee Aiyegbusi ,[3,4,5,6,7] Claire Beecher ,[8,9,10] Mike Horton ,[11] Karl Bang Christensen ,[12] Rhiannon Phillips ,[13] David Gillespie ,[1] Ernest Choy [14]

To cite: Pickles T, Macefield R, Aiyegbusi OL, *et al*. Patient Reported Outcome Measures for Rheumatoid Arthritis Disease Activity: a systematic review following COSMIN guidelines. *RMD Open* 2022;**8**:e002093. doi:10.1136/rmdopen-2021-002093

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/rmdopen-2021-002093).

Received 10 November 2021
Accepted 3 March 2022

For numbered affiliations see end of article.

Correspondence to
Tim Pickles;
PicklesTE@cardiff.ac.uk

ABSTRACT

Background The current standard of care in rheumatoid arthritis (RA) requires regular assessment of disease activity (DA). All standard RA DA measurement instruments require joint counts to be undertaken by a healthcare professional with/without a blood test. Few healthcare providers have the capacity to assess patients as frequently as stipulated by guidelines. Patient Reported Outcome Measures (PROMs) could be an efficient and informative way to assess RA DA, which is highlighted by the SARS-COV-2 pandemic, as most consultations are remote rather than face-to-face. We aimed to assess all PROMs for RA DA against the internationally recognised COSMIN guidelines to provide evidence-based recommendations to select the most suitable PROMs.

Methods Review registered on PROSPERO as CRD42020176176. The search strategy was based on a previous similar systematic review and expanded to include all articles up to January 2019. All identified articles were rated by two independent assessors following the COSMIN guidelines.

Results 668 abstracts were identified, with 10 articles included. A further 21 were identified from a previous review. Ten PROMs were identified. There was insufficient evidence to place any of the identified PROMs into recommendation for use category A due to lack of evidence for content validity, as stipulated by the COSMIN guidelines.

Conclusion Lack of evidence of content validity limits suitable PROM selection, therefore none can be recommended for use. It is acknowledged that all included PROMs were developed before the COSMIN guidelines were published. Future research on PROMs for RA DA must provide evidence of content validity.

Key messages

What is already known about this subject?

- Ten Patient Reported Outcome Measures (PROMs) have been developed to measure in rheumatoid arthritis (RA) disease activity (DA). Previous reviews have suggested the use of RADAI, RADAI5, PAS-II and RAPID3.

What does this study add?

- This is the first systematic review of PROMs for RA DA that follows the recent COSMIN guidelines.
- There was insufficient evidence to recommend any of the identified PROMs for use due to lack of evidence for content validity.

How might this impact on clinical practice or further developments?

- Care should be taken when making use, or interpreting the results, of any of the PROMs for RA DA identified in this review.
- Future research on the PROMs identified here, or any future developed PROMs for RA DA, must provide evidence of content validity.

INTRODUCTION

The standard measurement instrument for assessing disease activity (DA) for patients with rheumatoid arthritis (RA) has, for many years, been the Disease Activity Score (DAS) with 28-joint count (DAS28),[1] and more recently Simple Disease Activity Index (SDAI) and Clinical Disease Activity Index (CDAI).[2] DAS28 has four variants[3] but all require a laboratory test of either erythrocyte sedimentation rate or C reactive protein (CRP), and a formal tender and swollen joint count assessment (of shoulders, elbows, wrists, hands and knees) undertaken by a healthcare professional. Some of the DAS28 variants also factor in a patient global assessment on a 10 cm visual analogue scale, which adds a level of patient involvement. In common with DAS28, SDAI and CDAI require tender and swollen joint counts and a patient global assessment. In addition, CRP and a physician global assessment are also required for SDAI

and CDAI, respectively. Between joint counts and laboratory tests, these assessments are very time-consuming and resource-intensive and can only be undertaken when a patient comes in for a scheduled consultation.

The current standard of care in RA is 'Treat-to-Target' (T2T), which aims for sustained remission or failing this, low DA score.[4 5] Regular assessment of DA and adjustment of treatment accordingly is an integral part of T2T. National Institute for Health and Care Excellence (NICE)[6] and the European Alliance of Associations for Rheumatology (EULAR)[5] recommend DA is monitored every one to 3 months when disease is uncontrolled, and every 6–12 months when treatment target has been reached. Few healthcare providers have the capacity to assess patients as frequently as stipulated by NICE or EULAR guidelines: every 6 months is typically the best that is currently managed.[7] The SARS-COV-2 pandemic has made the problem more conspicuous with remote rather than face-to-face consultations. With infrequent monitoring, treatment is not adjusted sufficiently to keep pace with fluctuation in DA. It can also be the case that those in RA remission are seen more often than necessary while opportunities to treat RA flares are often provided too late.[8]

Alongside this, and with the advent of patient-centred care and value-based healthcare, Patient Reported Outcome Measures (PROMs) have become the pertinent options for monitoring disease progression and quality of life in numerous fields.[9] Unlike paper-based PROMs, electronic PROMs and computer adaptive test (CAT) platforms provide efficient, patient-friendly and location-independent methods of collecting such data, which can also satisfy the necessary properties required to enable useful measurement.[9] These CAT platforms are developed under item response theory or Rasch measurement theory methodologies, and allow for patients to respond to a minimal set of items while still calculating an accurate estimate.[10] Such examples are seen in the patient-reported outcomes measurement information system initiative.[11] The research around RA DA has suggested PROMs might prove preferable to measures requiring biomarkers.[12 13] Further, electronic versions of PROMs could be the future of measurement in rheumatology.[7 14 15]

To best understand the currently available psychometric evidence for these PROMs, a first step is to undertake a systematic review and assess the identified PROMs. Our review builds on the work of a 2016 systematic review in the same area,[16] which concluded that three PROMS: Rheumatoid Arthritis Disease Activity Index (RADAI), Rheumatoid Arthritis Disease Activity Index-Five (RADAI5) and Routine Assessment of Patient Index Data 3 (RAPID3), plus another measurement instrument called Patient-derived Disease Activity Score with 28-joint counts (Pt-DAS28, which is DAS28 but with the 28-joint count completed by the patient) had the strongest and most extensive validation. This systematic review identified articles describing, and assessed the properties of, measurement instruments that are not PROMs. Here though, a tighter lens is applied to focus solely on PROMs in the justification for the inclusion of articles in this review. Furthermore, the accepted guidelines concerning these systematic reviews for assessing PROMs from COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) have been updated since the 2016 systematic review[16] and now have a major focus on content validity.[17–19] This inevitably influences the assessment of legacy PROMs like those for RA DA, which were developed prior to these guidelines. The recommendations for use, which are the endpoint of the application of the COSMIN guidelines, are based largely on content validity as well.

Given this clear and definitive gap, our objective was to systematically review all PROMs for RA DA against the internationally recognised 2018 COSMIN guidelines[17–19] to provide evidence-based recommendations for use of the most suitable PROMs in research and clinical practice.

METHODS

This systematic review of all PROMs for RA DA was registered with the International Prospective Register of Systematic Reviews (PROSPERO)[20] as CRD42020176176, where a protocol is available,[21] and written in compliance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines,[22] with the PRISMA checklist provided (online supplemental appendix 1).

Guidelines

COSMIN guidelines were applied throughout this systematic review.[17–19] The review of Hendrikx *et al*[16] used the original COSMIN guidelines.[23–27] These have since been updated and it is these guidelines[17–19] and the methodology within them that were implemented here.

Search strategy

A search strategy was required to identify relevant articles (online supplemental appendix 2). Hendrikx *et al* published their search strategy,[16] which was based on a COSMIN guideline.[28] This search strategy was tested and refined to ensure certain articles were identified. An adapted version of that strategy was used to search the PubMED and EMBASE databases up to January 2019. The search was undertaken by one reviewer (TP).

Article selection

Following the implementation of the search strategy, a single assessor (TP) undertook the work of reviewing the articles for relevance, through screening of titles and abstracts to assessing eligibility for the review. The following inclusion and exclusion criteria were used:

Inclusion criteria:

1. The study population described in the article is of adult patients with RA.

2. The article describes details on a PROM specifically for DA in RA that can be reviewed against the COSMIN guidelines[17–19]
3. The article is in the English language.
4. The article is published in a peer-reviewed journal.

Exclusion criteria

1. The study population described in the article includes diseases other than RA (unless the details pertaining the patients with RA are presented separately).
2. The study population described in the article includes children (unless the details pertaining the adult patients are presented separately);
3. The article describes a measurement instrument that requires healthcare professional assessment;
4. The article describes a measurement instrument that requires a biomarker level determined through laboratory test.

Data extraction: study population characteristics

Characteristics of the study population, including number of participants, age, gender, rheumatoid factor (per cent positive), disease duration and DA (at baseline if reported at multiple timepoints) were extracted by one assessor (TP). Where multiple study populations were described within a single article, these were pooled together and described as one population if the statistics presented allowed for this. Where study population characteristics were described in a separate article to that reviewed here, that separate article was sought out and characteristics extracted as necessary.

Data extraction: risk of bias, content validity and quality of measurement properties

Data on the relevant measurement properties were extracted and summarised in the 'COSMIN checklist' Microsoft Excel (2018) spreadsheet (online supplemental appendix 3) available for download from the COSMIN website.[29] The spreadsheet contains the necessary risk of bias questions, which require the assessor to complete with categories: very good (V), adequate (A), doubtful (D), inadequate (I) or N/A (N). It also requires the assessor to rate the content validity and quality of measurement properties with ratings: sufficient (+), insufficient (–), inconsistent (±) or indeterminate (?). Decisions were made across the COSMIN domains of PROM development, content validity, structural validity, internal consistency, cross-cultural validity/measurement invariance, reliability, measurement error, criterion validity, hypothesis testing for construct validity and responsiveness. Two assessors (TP, and one of RM, OLA or CB) independently assessed the articles for risk of bias and quality of measurement property. Where ratings differed, a consensus was agreed by the two assessors.

For the purpose of the COSMIN criterion validity and responsiveness domains, a gold standard measurement instrument needed to be identified. DAS28,[1] SDAI and CDAI[2] are the most widely used and are accepted measurement instruments for this purpose and were therefore used as gold standard measurement instruments for this review.

For hypothesis testing for the COSMIN construct validity and responsiveness domains, subgroups within a rheumatoid arthritis population were assessed.

Where statistics, such as effect sizes, to be reviewed for COSMIN domains were not provided in an article, but could be readily calculated from other values given in the article, such as means and SD, the relevant statistics were calculated.

For all correlations, Spearman's ρ and Kendall's τ were accepted as appropriate statistical methods. Pearson's r could also be considered as an appropriate statistical method if some form of distributional information of the PROM was provided, such as mean and SD or a histogram.

In the case where multiple statistics were used to assess for quality of a PROM's measurement property (that is, one COSMIN domain) within an article, the lowest rating was used.

The COSMIN team were consulted for advice on the following two methodological points to confirm the best course of action.

Where risk of bias items required disease-specific knowledge, the independent assessors consulted with EC who, as a rheumatologist, is a clinical expert in RA to ensure consensus.

Assessment of some measurement properties according to the COSMIN guidelines requires statistical tests that can only be performed with specific software. In the case that the article stated that an analysis had been undertaken but the software stated and/or the outputs reported in the article were not feasible for that analysis, then the assessors followed the rating for risk of bias and quality of measurement properties for the actual analysis, rather than what it was stated as.

Determination of overall rating, quality of evidence and recommendation for use

The quality of the evidence for each summarised measurement property of each PROM was determined by assessors using the modified GRADE[30] approach defined in the COSMIN guidelines.[17–19] Quality of evidence has categories: high (H), moderate (M), low (L) and very low (VL). These were determined for each COSMIN domain for each PROM on the basis of risk of bias, inconsistency, imprecision and indirectness as specified in the COSMIN guidelines.[17–19] The rules for downgrading levels are well set for risk of bias, imprecision and indirectness, but require some formulation for inconsistency. We decided the following:

- If there was a ≥75% majority for a quality of measurement property rating across the studies, then there was no inconsistency.
- If there was a majority towards one such quality of measurement property rating but that majority was 60% to <75%, then we noted the inconsistency as serious.

- If there was no majority (50%) or if there was a majority towards one such quality of measurement property rating but that majority was 50% to <60%, then we noted the inconsistency as very serious.

The overall rating was the majority content validity or quality of measurement property rating for each domain for each PROM, so we similarly used categories sufficient (+), insufficient (–), inconsistent (±) or indeterminate (?). In the case of no majority, the lower quality of measurement property rating was used.

Once all summarised ratings for each COSMIN domain of each PROM were decided, recommendations for use of the reviewed PROMS for RA DA were applied according to the categories stipulated in the COSMIN guidelines.[17–19] These are:

1. PROM has evidence for sufficient content validity (any level) and at least low-quality evidence for sufficient internal consistency. Therefore, the PROM can be recommended for use and results obtained with these PROMS can be trusted;
2. PROM cannot be categorised into A or C. Therefore, the PROM has potential to be recommended for use, but requires further research to assess its quality.
3. PROM has high quality evidence for an insufficient measurement property. Therefore, the PROM should not be recommended for use.

Assessor agreement

Percentage agreement was calculated for all content validity and related risk of bias, plus all quality of measurement properties and related risk of bias combined. While there are three individuals acting as independent assessors, these are all combined here.

RESULTS

Article identification

A PRISMA Flow Diagram[22 31] presents the results of the search strategy and the proceeding reviewing undertaken to reach the 31 articles in this review (figure 1).

Of the 34 articles from the Hendrikx *et al* review[16] (left side of figure 1), 13 articles[32–44] were excluded. The 21 remaining articles[45–65] from the Hendrikx *et al* review[16] were included in this review.

After the deletion of 8 duplicates (identified twice by the same source), 668 articles were identified (right side of figure 1). Ten[66–75] of these were included in the review, so a total of 31 articles were included. These 31 articles described ten PROMs: RADAI, Rheumatoid Arthritis Disease Activity Index-Short Form (RADAI-SF), RADAI5, RAPID3, Routine Assessment of Patient Index Data 4 (RAPID4), Patient-based Disease Activity Score 2 (PDAS2), Patient Activity Score (PAS), Patient Activity Score-II (PAS-II), Patient Reported Outcome CLinical ARthritis Activity (PRO-CLARA) and Global Arthritis Score (GAS). Of the 10 articles published since the Hendrikx *et al* review,[16] 8 described RAPID3 and RAPID4 under the COSMIN criterion validity domain, 1 described

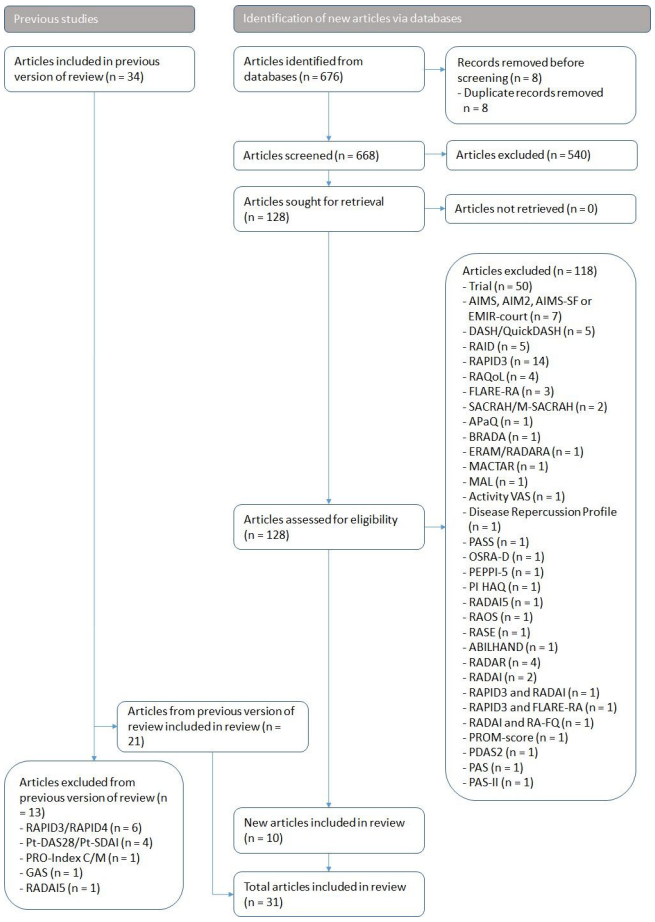

Figure 1 PRISMA flow diagram of article selection.

PRO-CLARA under the COSMIN hypotheses testing for construct validity domain and another one described RADAI under the COSMIN responsiveness comparison between subgroups subdomain.

Study characteristics

The characteristics of the study populations of the 31 articles are given in table 1. The majority of these articles (n=19) described a single PROM, nine described two PROMs, two described three PROMs and one article described four PROMs.

Only 1 clinical trial[68] was in the was included in this review of the 51 clinical trial articles sought for retrieval. It was notable that many excluded trials describe PROMs but not for RA DA, and where PROMs for RA DA were described, the statistical detail provided is only available to be assessed against the COSMIN guidelines in this single article.

Additional methodological requirements defined after articles were identified

The majority of PROMs described in the articles found in this review have a scoring system involving precalculation of a variable before summing up to create a total score, rather than just summing up the items they include. This was problematic for the assessment of the COSMIN structural validity and internal consistency domains, as, to be

Table 1 Characteristics of the study populations of the 31 articles

Article	PROMs described	n	Age (years)	Female (%)	RF+ (%)	Disease duration	Baseline disease activity
Boone 2015 SpringerPlus[66]	RAPID3	150	59.7 (10.8)	67	NR	12.2 (9.1) years	3.4 (1.4)
Che 2016 Clin Exp Rheum[67]	RAPID3	677	48.6 (12.3)	77.4	54.4	3.41 (1.64) months	5.0 (1.2)*
Kim 2014 Rheum Int[69]	RAPID3	400	55.4 (10.9)	87.0	91.7	7.6 (7.0) years	3.4 (1.1)
Kumar 2017 Ind J Med Res[70]	RAPID3	100	42.1 (11.6)	82	NR	6.0 (6.0–12.00)† months	7 (6–7)†
Munoz 2016 Clin Rheum[71]	RAPID3	119	61 (14)	77.3	NR	14 (5–21)† years	3.8 (2.7–5.1)†
Qorolli 2017 Clin Rheum[72]	RAPID3	68	53.5 (11.6)	85.3	NR	9.6 (7.9) years	5.9 (1.3)
Walker 2017 Rheum[74]	RAPID3, RAPID4	80	57.3 (44.9–65.9)†	58.8	NR	4.5 (1.1–9.5)† years	2.4 (1.9–3.5)†‡
Yokogawa 2015 Mod Rheum[75]	RAPID3	348	65.2 (12.1)	78.7	NR	11.4 (10.5) years	3.09 (1.22)
Salaffi 2015 Biomed Res Int[73]	PRO-CLARA	303	56.9 (9.9)	80.8	NR	5.9 (3.9) years	18.407 (7.2233)§
Katz 2018 ACR[68]	RADAI	96	54.8 (13.4)	87.5	NR	14.8 (12.3) years	NR
Bossert JBSPIN 2011[45]	RADAI5, RAPID3	200	56.9 (11.5)	75.5	78	12.8 (8.3) years	3.61 (1.43)
Castrejon J Clin Rheum 2013[46]	RAPID3	39	56.8 (13.9)	90	NR	3.5 (NR) years	NR
Choy A&R 2008[47]¶	PDAS2	582	59.9 (20-90)**	78	80	9.3 (0–54)** years	NR
Fransen Rheum 2000[49]	RADAI	584	59 (12)	71.9	69	8 (3–15)† years	4.3 (1.4)*
Fransen A&R 2001[48]††	RADAI	92	52 (13)	83	87	9 (4–14)† years	4.3 (1.4)*
Fujiwara Mod Rheum 2013[50]	RAPID3	250	59.3 (14.0)	78.4	NR	10.35 (9.83) years	NR
Houssien Rheum 1999[51]	RADAI	100	57.7 (12.2)	78	NR	11.5 (8.3) years	4.24 (1.3)*
Leeb J Rheum 2008[52]	RADAI, RADAI5	169	57 (19-78)**	79.9	50	7.2 (0.2–46.0)** years	3.51 (0.28-6.67)** and 3.19 (1.21–5.83)‡**
Leeb JBSPIN 2009[53]	RAPID3, RADAI5	108	59.5 (24-87)‡‡	77.8	54	NR	2.95 (0.43-6.24)‡‡
Pincus ACR 2011[54]§§	RAPID3	982	52.0 (11.6)	83.2	81.9	6.2 (4.3) years	6.9 (4.3-9.1)‡‡
Salaffi CER 2010[59]¶¶	PRO-CLARA, RAPID3, RADAI	443	57.5 (11.6)	81.5	77	5.7 (6.3) years	3.94 (2.04)
Rintelen J Rheum 2013[57]	RADAI5	705	62.7 (13.4)	75.9	54.4	97.3 (98.0) months	3.31 (1.37)***
Wolfe J Rheum 2005[65]	PAS, PAS-II, RADAI	9078	62.2 (12.6)	78.2	NR	16.2 (10.9) years	NR
Pincus J Rheum 2008[55]	RAPID3, RAPID4	285	57.4 (14.6)	73.0	NR	9.7 (9.0) years	3.4 (1.7)
Salaffi CER 2012[58]	PRO-CLARA, RAPID3, RADAI, PAS	191	56.6 (12.2)	82.7	NR	5.1 (5.5) years	6.02 (1.15) and 4.38 (0.92)*
Stucki A&R 1995[61]	RADAI, RADAI-SF	55	60 (14.6)	62	NR	5.1 (1.3–10.7)† years	3.1 (1.6–7.8)††††
Sullivan J Rheum 2010[62]	RADAI, GAS	740	57.06 (13.7)	83.0	63.8	14.25 (12.3) years	4.05 (1.5)‡
Uhlig ARD 2009[63]	RAPID3, RADAI	28	61.1 (6.2)	64	64	16.6 (10.4) years	3.12 (1.27)

Continued

Table 1 Continued

Article	PROMs described	n	Age (years)	Female (%)	RF+ (%)	Disease duration	Baseline disease activity
Veehof ARD 2008[64]	RADAI, RADAI-SF	191	54.4 (13.3)	71	NR	7.0 (3.0–17.0)† years	5.42 (1.07)
Pincus ACR 2010[56]	RAPID3	200	53.4 (16.2)	81.0	NR	11.6 (10.8) years	3.7 (1.5)
Singh Clin Rheum 2012[60]	RAPID3	200	42.2 (NR)	83	NR	4.9 (NR) years	5.2 (1.6)

For age, disease duration and baseline disease activity, values are mean (SD) unless noted otherwise. Baseline disease activity is DAS28-ESR-4v unless noted otherwise.
*DAS28-ESR-3v.
†Median (IQR).
‡DAS28-CRP-4v.
§CDAI.
¶Details on age, percentage female, percentage RF+ and disease duration pooled across the three populations listed separately.
**Mean (range).
††Details on baseline disease activity pooled across the Nonflare and Flare groups listed separately.
‡‡Median (range).
§§Details on age, percentage female, percentage RF+, disease duration and baseline disease activity pooled across the three populations listed separately defined in separate article.[83]
¶¶Details on age, percentage female, percentage RF+ and disease duration pooled across the two populations listed separately.
***DAS28, no further information provided.
†††Modified DAS, defined in separate article.[84]
NR, not reported; RF+, rheumatoid factor positive.

undertaken correctly and to provide relevant meaningful results, these require the individual items to be inputted into the analysis, rather than a combination of precalculated variable and individual items. Therefore, in the case where results that would be assessed under the COSMIN structural validity and internal consistency domains were given but the PROM had such a scoring structure, the result was ignored and a note added to the relevant cell or cells of the 'COSMIN checklist' Microsoft Excel (2018) spreadsheet stating this. All other COSMIN domains were still assessed, as those numeric COSMIN domains focus on analyses requiring the total score, rather than that of the individual items. A list of PROMs, the reasons why there was a problem and whether any structural validity or internal consistency analyses were undertaken are provided (online supplemental appendix 4).

For the assessment of the COSMIN measurement error domain, the minimally important change (MIC) must be defined for comparison against smallest detectable change (SDC) or limits of agreement (LoA). SDCs were calculated for RADAI and RAPID3 in one reviewed article[63] and these were compared against values (labelled as minimally important difference (MID)) from an article not reviewed here.[76] It was notable that in article not reviewed here,[76] RAPID3 was correctly scored 0–30, which was the case in the article providing the MIC,[76] but in the reviewed article,[63] RAPID3 was scored 0–10, so for comparison, the MIC in Pope was divided by 3; 3.6 became 1.2.

Within the COSMIN hypothesis testing for construct validity domain, and specifically the COSMIN comparison with other outcome measurement instruments (convergent validity) subdomain, the second risk of bias item asked: 'Were the measurement properties of the comparator instrument(s) adequate?' and to answer this, there was a need to determine whether sufficient measurement properties of the comparator instrument were available and which study population they applied to. In all cases but one, there were 'sufficient measurement properties of the comparator instrument(s) in a population similar to the study population' except for pulp-to-palm distance, which only had measurement properties described in an orthopaedics population.[77] For this reason, the risk of bias rating for RAPID3 in one of the reviewed articles[72] was doubtful (D), regardless of the other comparator instruments.

Relevant specifically to quality of measurement properties ratings of the COSMIN hypothesis testing for construct validity and responsiveness domains, COSMIN recommend that the review team formulate a set of hypotheses.[78] Therefore, hypotheses (online supplemental appendix 5) were agreed in consultation between all assessors and with clinical expertise from EC.

Content validity and related risk of bias

Only two articles described content validity, both under the heading of Cognitive interview study or other pilot test for the COSMIN comprehensibility domain (table 2, Content validity—PROM Development columns). These covered the PROMs PDAS2 (47) and PRO-CLARA.[59]

For PDAS2, very little detail about the process was available other than the number of participants interviewed, which was 20, so a D risk of bias rating was given under cognitive interview study or other pilot test comprehensibility study. For the content validity comprehensibility, patients were not asked about the comprehensibility of item instructions and it was not clear if patients were asked about the comprehensibility of all of the items and

Table 2 Risk of bias rating, content validity rating and quality of measurement property rating of the 31 articles

		Content validity—PROM Development			Structural validity		Internal consistency		Reliability		Measurement error		Criterion validity		Hypotheses testing for construct validity			Responsiveness				
		1by	Cty	RR											9a	9b		10a	10b	10c	10d	
PROM	Article	R	C	C	R	Q	R	Q	R	Q	R	Q	R	Q	R	R	Q	R	R	R	R	Q
RADAI	Salaffi CER 2010[59]												V	+		V	+					
	Salaffi CER 2012[58]																			V	V	+
	Fransen A&R 2001[48]												V	–		V	–	V			V	+
	Fransen Rheum 2000[49]						API	API					V	–	V		–					
	Uhlig ARD 2009[63]								V	+	V	–										
	Veehof ARD 2008[64]				API	API	API	API					V	–						V	I	+
	Stucki A&R 1995[61]						API	API							V		–					
	Sullivan J Rheum 2010[62]												V	–				V				–
	Houssien Rheum 1999[51]												V	–	V		–					
	Leeb J Rheum 2008[52]						API	API														
	Wolfe J Rheum 2005[65]														V		–					
	Katz 2018 ACR[68]																			V		–
RADAI-SF	Veehof ARD 2008[64]				V	+	V	+					V	–						V	I	+
	Stucki A&R 1995[61]						V	+														
RADAI5	Bossert JBSPIN 2011[45]												V	–		V	+					
	Leeb J Rheum 2008[52]				D	?	V	?					V	–	V		?	V				–
	Leeb JBSPIN 2009[53]				D	?	V	?					V	–								
	Rintelen J Rheum 2013[57]												V	–								
PDAS2	Choy A&R 2008[47]	D	–	+			API	API	D	+			V	+	V		–		V	V		–
PAS	Salaffi CER 2012[58]																			V	V	+
	Wolfe J Rheum 2005[65]														V		–					
PAS-II	Wolfe J Rheum 2005[65]														V		–					

Continued

Table 2 Continued

PROM	Article	Content validity—PROM Development: 1by R	Cty C	RR C	Structural validity: R	Q	Internal consistency: R	Q	Reliability: R	Q	Measurement error: R	Q	Criterion validity: R	Q	Hypotheses testing for construct validity: 9a R	9b R	Q	Responsiveness: 10a R	10b R	10c R	10d R	Q
RAPID3	Salaffi CER 2010[59]												V	+		V	+					
	Salaffi CER 2012[58]																			V	V	+
	Pincus ACR 2011[54]												V	+	V		+					
	Pincus J Rheum 2008[55]												V	–								
	Castrejon J Clin Rheum 2013[46]																				I	–
	Fujiwara Mod Rheum 2013[50]												V	–								
	Pincus ACR 2010[56]												V	–								
	Singh Clin Rheum 2012[60]												V	+								
	Boone 2015 SpringerPlus[66]												V	–								
	Kim 2014 Rheum Int[69]												V	–	V		–					
	Kumar 2017 Ind J Med Res[70]												V	–								
	Walker 2017 Rheum[74]												V	–	V		–					
	Yokogawa 2015 Mod Rheum[75]						API	API	I	?			V	–	V		–	V				–
	Uhlig ARD 2009[63]								V	+	V	–										
	Bossert JBSPIN 2011[45]												V	–		V	+					
	Leeb JBSPIN 2009[53]				API	API	API	API					V	–								
	Che 2016 Clin Exp Rheum[67]												V	–								
	Munoz 2016 Clin Rheum[71]												V	+								
	Qorolli 2017 Clin Rheum[72]												V	+	D		–					
RAPID4	Pincus J Rheum 2008[55]												V	–								
	Walker 2017 Rheum[74]												V	–	V		–					

Continued

Table 2 Continued

PROM	Article	Content validity—PROM Development			Structural validity		Internal consistency		Reliability		Measurement error		Criterion validity		Hypotheses testing for construct validity		Responsiveness			
		1by	Cty	RR											9a	9b	10a	10b	10c	10d
		R	C	C	R	Q	R	Q	R	Q	R	Q	R	Q	R	R	R	R	R	R
PRO-CLARA	Salaffi CER 2010[59]	D	?	+	API	API	API	API					V	+	V	V	–			Q
	Salaffi CER 2012[58]																		V	V +
	Salaffi 2015 Biomed Res Int[73]															V	+			
GAS	Sullivan J Rheum 2010[62]												V	–			V			–

Modified COSMIN. Only relevant columns shown so specific content validity columns, plus no cross-cultural validity/measurement invariance columns. Designed for risk of bias ratings but adapted to also show content validity ratings and quality of measurement property ratings.
1by: Cognitive interview study or other pilot test comprehensibility study.
9a: Convergent validity.
9b: Known groups validity.
10a: Comparison with gold standard.
10b: Comparison with other instruments.
10c: Comparison between subgroups.
±, inconsistent; +, sufficient; -, insufficient ; ?, indeterminate; A, adequate; API, analyses for these COSMIN domains are presented in these articles but have to be ignored; C, content validity rating; Cty, comprehensibility; D, doubtful; I, inadequate; N, N/A; PROMs, Patient Reported Outcome Measures; Q, quality of measurement property rating; R, risk of bias rating; RR, rating of reviewers; V, very good.

their response options (some items were mentioned but others were not), so the overall content validity rating was –. The rating of reviewers was + because it was clear that all items and response options were appropriately worded and response options matched the items.

For PRO-CLARA, 72 patients were surveyed but a qualitative method should be used to assess content validity, so a D risk of bias rating was given under Cognitive interview study or other pilot test Comprehensibility study. There was no detail about patients being asked about the comprehensibility of item instructions and of the items and their response options, so the overall content validity rating was ?. The rating of reviewers was + because it was clear that all items and response options were appropriately worded and response options matched the items.

Content validity, downgrading, overall rating and quality of evidence

There were no Relevance or Comprehensiveness ratings, so there were no overall content validity ratings; therefore, the quality of evidence rating for both of these was Very low. As there was no majority for content validity overall rating, the lowest was used and was thus insufficient (–) for PDAS2 and indeterminate (?) for PRO-CLARA (table 3, Content validity Comprehensibility row).

Quality of measurement properties and related risk of bias

Table 2 columns to the right of content validity summarise the quality of measurement properties and related risk of bias ratings.

Quality of measurement properties, downgrading, overall rating and quality of evidence

The evidence in table 2, allowed the overall rating and quality of evidence to be determined, as presented in table 3.

Recommendations for use

Using the overall Rating and quality of evidence for each COSMIN domain within each PROM, recommendations for use in research and clinical practice, the main result of this systematic review, were attributed (table 3, final row) as follows:

- Category B: RADAI-SF, PDAS2, PAS, PRO-CLARA and RAPID3.
- Category C: RADAI, RADAI5, PAS-II, RAPID4 and GAS.

There were no PROMs attributed to Category A, as none had sufficient evidence of content validity. All Category C PROMs had at least one COSMIN domain with High quality evidence for an insufficient (-) measurement property and all Category B PROMs had at least one COSMIN domain with High quality evidence for a sufficient (+) measurement property, except RAPID3, which, at best, had Moderate quality evidence for an insufficient (-) measurement property. Despite this, it fitted into neither Category A nor C and was therefore attributed to Category B.

Assessor agreement

From a total of 435 ratings, 399 were in agreement, giving an overall agreement of 91.7%.

DISCUSSION

The lack of sufficient evidence for content validity means that no PROMs identified in this review can be recommended for use (attributed to Category A) in research and clinical practice. PROMs RADAI-SF, PDAS2, PAS, PRO-CLARA and RAPID3 are attributed to Category B and therefore have potential to be recommended for use, but require further research to assess their quality. PROMs RADAI, RADAI5, PAS-II, RAPID4 and GAS are attributed to Category C and therefore should not be recommended for use.

RAPID3 is attributed to Category B despite, at best, having Moderate quality evidence for an insufficient (-) measurement property for the COSMIN responsiveness domain. This is a lower level of evidence than all PROMs in Category C, which have at least one COSMIN domain with High quality evidence for an insufficient (-) measurement property. It would appear as a limitation of the COSMIN guidelines that there is no Category D for PROMs like RAPID3.

While not possible to excuse the research community for not having undertaken the necessary research, it is notable that all identified PROMs were first described before or in the same year as the first set of COSMIN guidelines in 2010,[23–27] and therefore all before the updated COSMIN guidelines[17–19] in which content validity was prioritised. This research must be done before any of these PROMs can be recommended and is also true of any new PROMs that are developed for the measurement of RA DA.

It is also important to note the fact that many of the PROMs identified here have a scoring system involving precalculation of a variable before summing up to create a total score, rather than just summing up the items they include, and that this causes an issue with the assessment of the COSMIN structural validity domain and the COSMIN internal consistency domain. That this is the case for 8 of the 10 PROMs (RADAI, PDAS2, PAS, PAS-II, RAPID3, RAPID4, PRO-CLARA and GAS) identified here suggests that there is a systematic reason behind this for PROMs for RA DA. Five of these precalculate a joint count variation and six precalculate a functional ability variation joint count variations are used in all three gold standard measurement instruments defined here (DAS28, CDAI and SDAI),[1–3] while joint count and functional ability variations are key within the American College of Rheumatology (ACR) criteria.[79] A desire to continue the use of known instruments with PROMs may have contributed to this issue and can be moved away from, as is seen in RADAI-SF and RADAI5.

Assessor agreement was considerably lower for content validity related risk of bias and content validity ratings. This is due to the paucity of information available in

Table 3 Overall rating and quality of the evidence for measurement properties of the PROMs, plus recommendation for use category

	RADAI		RADAI-SF		RADAI5		PDAS2		PAS		PAS-II		RAPID3		RAPID4		PRO-CLARA		GAS	
COSMIN domain	**OR**	**QoE**	**OR**	**QoE**	**OR**	**QoE**	**OR**	**QoE**	**OR**	**QoE**	**OR**	**QoE**	**OR**	**QoE**	**OR**	**QoE**	**OR**	**QoE**	**OR**	**QoE**
Content validity																				
Relevance																				
Comprehensiveness																				
Comprehensibility							–	VL									?	VL		
Structural validity	API	API	+	H	?	M	API	API	AWI	AWI	AWI	AWI	API	API	AWI	AWI	API	API	AWI	AWI
Internal consistency	API	API	+	H	?	H	API	API	AWI	AWI	AWI	AWI	API	API	AWI	AWI	API	API	AWI	AWI
Reliability	+	L					+	L					?	L						
Measurement error	–	L											–	L						
Hypothesis testing for construct validity	–	H			?	L	–	H	–	H	–	H	–	L	–	M	–	L		
Cross-cultural validity/measurement invariance																				
Criterion validity	–	H	–	H	–	H	+	H					–	M	–	H	+	H	–	H
Responsiveness	+	M	+	H	–	H	–	H	+	H			–	M			+	H	–	H
Recommendation for use category	C		B		C		B		B		C		B*		C		B		C	

*In Recommendation Category B despite having no high quality evidence.
?, indeterminate; +, sufficient; -, insufficient; API, analyses for these COSMIN domains are presented in these articles but have to be ignored; AWI, analyses would have to be ignored but aren't presented; H, high; L, low; M, moderate; OR, overall rating; PROMs, Patient Reported Outcome Measures; QoE, quality of evidence; VL, very low.

the two articles,[47 59] and difficulty in interpreting what the authors actually undertook. Assessor agreement was much higher for quality of measurement property related risk of bias and quality of measurement property, and the overall agreement is also high. This is largely dominated by 231 quality of measurement property related risk of bias agreements on a Very good rating.

The ACR Rheumatoid Arthritis Disease Activity Measure Workgroup have published a systematic review of all RA DA measurement instruments[80] and recommend the following two PROMs identified here as preferred measures for regular use: RAPID3 and PAS-II. DAS28, CDAI and SDAI are also recommended as preferred measures for regular use. Additionally, in this ACR review, two PROMs identified here reached the minimum standard for regular use: RADAI and RADAI5, and this was also the case for DAS, Patient Derived DAS28, Hospital Universitario La Princesa Index, Multi-Biomarker Disease Activity Score and Routine Assessment of Patient Index Data 5 (RAPID5).

The previous systematic review undertaken by Hendrikx *et al* in 2016[16] stated that, of the PROMs identified here, RADAI, RADAI5 and RAPID3 had the most extension validations and the strongest level of evidence. It also stated the same of a measurement instrument labelled as Pt-DAS28, which is not a PROM.

There are therefore recommendations for PROMs RADAI, RADAI5 and RAPID3 from both sources[16 80] and PROM PAS-II from the ACR review,[80] while here we cannot recommend any identified PROMs.

The ACR review,[80] an update from 2012,[81] includes all possible measurement instruments for assessing RA DA. It is therefore difficult to fully implement the COSMIN guidelines[17–19] for this review as these relate solely to PROMs. Furthermore, their methods included a Delphi survey to aid with determining if measurement instrument should be recommended.

In the Hendrikx *et al* review,[16] a previous set of COSMIN guidelines was employed[23–27] that did not prioritise content validity. Also, measurement instruments reviewed included biomarkers and/or healthcare professional assessments, included articles contained information on the evolution of PROMs not yet in their finalised state, and other included articles described measurement instruments as PROMs when they fulfilled a different role. This review applies a tighter lens focusing solely on PROMs and makes use of the most recent COSMIN guidelines,[17–19] which provides some reasoning behind the discrepancies noted above.

The set of assessors were not experts in RA and therefore did not have the knowledge to complete with certainty some risk of bias items. As mentioned, where this was the case, TP discussed the matter with EC, and then with the independent assessors. This is a limitation of the independence of the assessors on these few risk of bias items, as there was essentially only one opinion. Further limitations are that only one assessor (TP) undertook the search strategy, article selection and data extraction of study population characteristics.

We state the necessary hypotheses required for this review in the *Additional methodological requirements defined after articles were identified*. These were written in consultation with the review team and EC. For comparison, we searched PROSPERO for the term 'COSMIN' and limited the review to those registered in musculoskeletal Health area of review. A total of 184 records were returned but found only one published article that defines hypotheses.[82] The hypotheses in this article relate solely to correlations. As we have stated, this articles also sets 0.5 as a lower bound for convergent correlations, but then uses 0.3–0.5 for semiconvergent correlations, where we use 0.4 as a lower bound, and also sets 0.3 as an upper bound for divergent correlations, where we use 0.3–0.5. There is of course no set guideline on where to place these bounds, and we see here that only correlations are set, where we also defined hypotheses for effect sizes and areas under the curve. The article found through PROSPERO[82] makes reference to some of the updated 2018 COSMIN articles[17 18] and also the online user manual, which does provide very similarly written generic hypotheses, which are reproduced from Measurement in Medicine.[78] There is a case to attempt more detailed research into this area to provide guidelines on how review teams, or indeed researchers attempting the original research, could define these hypotheses.

None of the identified articles made use of item response theory or Rasch measurement theory to evidence the psychometric properties of these PROMs. These are defined for use in the COSMIN structural validity domain and can also provide evidence for the COSMIN cross-cultural validity/measurement invariance domain, for which there was no evidence. All evidence in the COSMIN structural validity domain was provided through confirmatory or explanatory factor analyses.[52 53 64]

In conclusion, no PROMs identified in this review can be recommended for use according to COSMIN guidelines[17–19] due to lack of sufficient evidence for content validity. This is despite previous reviews[16 80] suggesting the use of RADAI, RADAI5, PAS-II and RAPID3. All PROMs identified here were first described before initial COSMIN guidelines were published and thus also before the updated guidelines that prioritised content validity. The majority of identified PROMs have scoring systems that preclude evidence in the COSMIN structural validity and internal consistency domains. Care should be taken when making use, or interpreting the results, of any of the PROMs for RA DA identified in this review. Future research on the PROMs identified here, or any future developed PROMs for RA DA, must look to evidence content validity. Future developed PROMs should implement scorings systems without precalculation of variations entered into the scoring system with other items. These could also look to item response theory or Rasch measurement theory to evidence their psychometric

properties. This would allow for any such PROMs to be developed in computer adaptive tests.

Author affiliations
[1]Centre for Trials Research, Cardiff University, Cardiff, UK
[2]Bristol Centre for Surgical Research, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK
[3]Institute of Applied Health Research, University of Birmingham, Birmingham, UK
[4]Centre for Patient Reported Outcomes Research, Institute of Applied Health Research, University of Birmingham, Birmingham, UK
[5]NIHR Applied Research Collaboration West Midlands, Birmingham, UK
[6]Birmingham Health Partners Centre for Regulatory Science and Innovation, University of Birmingham, Brmingham, UK
[7]NIHR Birmingham Biomedical Research Centre, NIHR Surgical Reconstruction and Microbiology Research Centre, University of Birmingham, Birmingham, UK
[8]School of Nursing and Midwifery, National University of Ireland Galway, Galway, Ireland
[9]Evidence Synthesis Ireland and Cochrane Ireland, National University of Ireland Galway, Galway, Ireland
[10]Health Research Board - Trials Methodology Research Network, National University of Ireland, Galway, Ireland
[11]Psychometric Laboratory for Health Sciences, University of Leeds, Leeds, UK
[12]Section of Biostatistics, University of Copenhagen, Copenhagen, Denmark
[13]Cardiff School of Sport and Health Sciences, Cardiff Metropolitan University, Cardiff, UK
[14]Department of Infection and Immunity, Cardiff University, Cardiff, UK

Twitter Tim Pickles @picklestim, Olalekan Lee Aiyegbusi @lee_1881, Claire Beecher @beecherclaire1, Rhiannon Phillips @Nannon_Phillips, David Gillespie @DaveGuk87 and Ernest Choy @ErnestChoy1

Contributors TP and EC drafted the initial research idea. TP undertook the search strategy, reviewed all titles and then abstracts for inclusion, reviewed all full-text articles, extracted results and assessed them against COSMIN guideline criteria. RM, OLA and CB extracted results and assessed them against COSMIN guideline criteria for the articles assigned by TP. All authors declare having read and made a substantial contribution to the final manuscript. TP is the guarantor.

Funding TP is supported by a National Institute of Health Research (NIHR) Doctoral Fellowship, funded by the Welsh Government through Health and Care Research Wales (NIHR-FS-19). The systematic review was funded via a Health and Care Research Wales Pathway to Portfolio Grant. This work was supported by the MRC-NIHR Trials Methodology Research Partnership (MR/S014357/1). OLA receives funding from the NIHR Birmingham Biomedical Research Centre (BRC), NIHR Applied Research Collaboration (ARC), West Midlands at the University of Birmingham and University Hospitals Birmingham NHS Foundation, Innovate UK (part of UK Research and Innovation), Gilead Sciences Ltd and Janssen Pharmaceuticals, Inc. OLA declares personal fees from Gilead Sciences Ltd, GlaxoSmithKline (GSK) and Merck outside the submitted work. RM is funded by the NIHR Bristol Biomedical Research Centre (BRC).

Disclaimer The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Competing interests None declared.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data sharing not applicable as no datasets generated and/or analysed for this study.

ORCID iDs
Tim Pickles http://orcid.org/0000-0001-7743-0234
Rhiannon Macefield http://orcid.org/0000-0002-6606-5427
Olalekan Lee Aiyegbusi http://orcid.org/0000-0001-9122-8251
Claire Beecher http://orcid.org/0000-0001-6581-3756
Mike Horton http://orcid.org/0000-0002-6675-7335
Karl Bang Christensen http://orcid.org/0000-0003-4518-5187
Rhiannon Phillips http://orcid.org/0000-0002-4256-4598
David Gillespie http://orcid.org/0000-0002-6934-2928
Ernest Choy http://orcid.org/0000-0003-4459-8609

REFERENCES

1 Prevoo ML, van 't Hof MA, Kuper HH, *et al*. 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.
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:S100–8.
3 Tamhane A, Redden DT, McGwin G, *et al*. Comparison of the disease activity score using erythrocyte sedimentation rate and C-reactive protein in African Americans with rheumatoid arthritis. *J Rheumatol* 2013;40:1812–22.
4 Molenaar ETH, Voskuyl AE, Dinant HJ, *et al*. Progression of radiologic damage in patients with rheumatoid arthritis in clinical remission. *Arthritis Rheum* 2004;50:36–42.
5 Smolen JS, Landewé R, Bijlsma J, *et al*. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2016 update. *Ann Rheum Dis* 2017;76:960–77.
6 NICE. Rheumatoid arthritis in adults: management, 2009. Available: https://www.nice.org.uk/guidance/cg79/chapter/recommendations
7 Austin L, Sharp CA, van der Veer SN, *et al*. Providing 'the bigger picture': benefits and feasibility of integrating remote monitoring from smartphones into the electronic health record. *Rheumatology* 2020;59:367–78.
8 Martin M. The role of remote monitoring in the future of the NHS, 2020. Available: https://www.nhsx.nhs.uk/blogs/role-remote-monitoring-future-nhs/
9 Black N. Patient reported outcome measures could help transform healthcare. *BMJ* 2013;346:f167.
10 Harrison CJ, Geerards D, Ottenhof MJ, *et al*. Computerised adaptive testing accurately predicts CLEFT-Q scores by selecting fewer, more patient-focused questions. *J Plast Reconstr Aesthet Surg* 2019;72:1819–24.
11 Gershon RC, Rothrock N, Hanrahan R, *et al*. The use of PROMIS and assessment center to deliver patient-reported outcome measures in clinical research. *J Appl Meas* 2010;11:304–14.
12 El Miedany Y, El Gaafary M, Youssef S, *et al*. AB0325 Biomarkers and Patient Tailored Approach in Rheumatoid Arthritis: Can Proms be the Missing Biomarker? *Ann Rheum Dis* 2015;74:1002.2–1002.
13 Fautrel B, Alten R, Kirkham B, *et al*. Call for action: how to improve use of patient-reported outcomes to guide clinical decision making in rheumatoid arthritis. *Rheumatol Int* 2018;38:935–47.
14 El Miedany Y. PROMs for Rheumatoid Arthritis. In: El Miedany Y, ed. *Patient reported outcome measures in rheumatic diseases*. Springer, 2016.
15 Hamann PDH, Minaur N, Tobias JH, *et al*. Capturing remote disease activity - results of a 12-month clinical pilot of a smartphone app in NHS rheumatology clinics in Bristol. *Rheumatology* 2020;59:2158–61.
16 Hendrikx J, de Jonge MJ, Fransen J, *et al*. Systematic review of patient-reported outcome measures (PROMs) for assessing disease activity in rheumatoid arthritis. *RMD Open* 2016;2:e000202.
17 Mokkink LB, de Vet HCW, Prinsen CAC, *et al*. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;27:1171–9.
18 Prinsen CAC, Mokkink LB, Bouter LM, *et al*. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;27:1147–57.
19 Terwee CB, Prinsen CAC, Chiarotto A, *et al*. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res* 2018;27:1159–70.
20 Page MJ, Shamseer L, Tricco AC. Registration of systematic reviews in prospero: 30,000 records and counting. *Syst Rev* 2018;7:32.

21 Pickles T, Macefield R, Olalekan OL, *et al*. Patient-reported outcome measures for rheumatoid arthritis disease activity: a systematic review. PROSPERO. 2020;CRD42020176176. Available: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020176176
22 Page MJ, McKenzie JE, Bossuyt PM, *et al*. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71.
23 Mokkink LB, Terwee CB, Gibbons E, *et al*. Inter-Rater agreement and reliability of the COSMIN (consensus-based standards for the selection of health status measurement instruments) checklist. *BMC Med Res Methodol* 2010;10:82.
24 Mokkink LB, Terwee CB, Knol DL, *et al*. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol* 2010;10:22.
25 Mokkink LB, Terwee CB, Patrick DL, *et al*. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737–45.
26 Mokkink LB, Terwee CB, Patrick DL, *et al*. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010;19:539–49.
27 Terwee CB, Mokkink LB, Knol DL, *et al*. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 2012;21:651–7.
28 Terwee CB, Jansma EP, Riphagen II, *et al*. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res* 2009;18:1115–23.
29 COSMIN. Guideline for Systematic Reviews of Outcome Measurement Instruments [Available from. Available: https://www.cosmin.nl/tools/guideline-conducting-systematic-review-outcome-measures/
30 Schünemann H, Brożek J, Guyatt G, *et al*. GRADE Handbook - Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach, 2013. Available: https://gdt.gradepro.org/app/handbook/handbook.html
31 Moher D, Liberati A, Tetzlaff J, *et al*. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009;339:b2535.
32 Blanchais A, Berthelot J-M, Fontenoy A-M, *et al*. Weekly home self-assessment of RAPID-4/3 scores in rheumatoid arthritis: a 6-month study in 26 patients. *Joint Bone Spine* 2010;77:582–7.
33 Castrejón I, Dougados M, Combe B, *et al*. Can remission in rheumatoid arthritis be assessed without laboratory tests or a formal joint count? possible remission criteria based on a self-report RAPID3 score and careful joint examination in the ESPOIR cohort. *J Rheumatol* 2013;40:386–93.
34 Harrington JT. The uses of disease activity scoring and the physician global assessment of disease activity for managing rheumatoid arthritis in rheumatology practice. *J Rheumatol* 2009;36:925–9.
35 Heegaard C, Dreyer L, Egsmose C, *et al*. Test-Retest reliability of the disease activity score 28 CRP (DAS28-CRP), the simplified disease activity index (SDAI) and the clinical disease activity index (CDAI) in rheumatoid arthritis when based on patient self-assessment of tender and swollen joints. *Clin Rheumatol* 2013;32:1493–500.
36 Janta I, Naredo E, Martínez-Estupiñán L, *et al*. Patient self-assessment and physician's assessment of rheumatoid arthritis activity: which is more realistic in remission status? A comparison with ultrasonography. *Rheumatology* 2013;52:2243–50.
37 Kavanaugh A, Lee SJ, Weng HH, *et al*. Patient-Derived joint counts are a potential alternative for determining disease activity score. *J Rheumatol* 2010;37:1035–41.
38 Pincus T, . RAPID3, an index of only 3 patient self-report core data set measures, but not ESR, recognizes incomplete responses to methotrexate in usual care of patients with rheumatoid arthritis. *Bull Hosp Jt Dis* 2013;71:117–20.
39 Pincus T, Bergman MJ, Yazici Y, *et al*. An index of only patient-reported outcome measures, routine assessment of patient index data 3 (RAPID3), in two abatacept clinical trials: similar results to disease activity score (DAS28) and other rapid indices that include physician-reported measures. *Rheumatology* 2008;47:345–9.
40 Pincus T, Chung C, Segurado OG, *et al*. An index of patient reported outcomes (PRO-Index) discriminates effectively between active and control treatment in 4 clinical trials of adalimumab in rheumatoid arthritis. *J Rheumatol* 2006;33:2146–52.
41 Pincus T, Hines P, Bergman MJ, *et al*. Proposed severity and response criteria for routine assessment of patient index data (RAPID3): results for categories of disease activity and response criteria in abatacept clinical trials. *J Rheumatol* 2011;38:2565–71.
42 Pincus T, Strand V, Koch G, *et al*. An index of the three core data set patient questionnaire measures distinguishes efficacy of active treatment from that of placebo as effectively as the American College of Rheumatology 20% response criteria (ACR20) or the Disease Activity Score (DAS) in a rheumatoid arthritis clinical trial. *Arthritis Rheum* 2003;48:625–30.
43 Riazzoli J, Nilsson Jan-Åke, Teleman A, *et al*. Patient-Reported 28 swollen and tender joint counts accurately represent RA disease activity and can be used to assess therapy responses at the group level. *Rheumatology* 2010;49:2098–103.
44 Rintelen B, Haindl PM, Sautner J, *et al*. The rheumatoid arthritis disease activity index-5 in daily use. proposal for disease activity categories. *J Rheumatol* 2009;36:918–24.
45 Bossert M, Prati C, Vidal C, *et al*. Evaluation of self-report questionnaires for assessing rheumatoid arthritis activity: a cross-sectional study of RAPID3 and RADAI5 and flare detection in 200 patients. *Joint Bone Spine* 2012;79:57–62.
46 Castrejón I, Bergman MJ, Pincus T. MDHAQ/RAPID3 to recognize improvement over 2 months in usual care of patients with osteoarthritis, systemic lupus erythematosus, spondyloarthropathy, and gout, as well as rheumatoid arthritis. *J Clin Rheumatol* 2013;19:169–74.
47 Choy EH, Khoshaba B, Cooper D, *et al*. Development and validation of a patient-based disease activity score in rheumatoid arthritis that can be used in clinical trials and routine practice. *Arthritis Rheum* 2008;59:192–9.
48 Fransen J, Häuselmann H, Michel BA, *et al*. Responsiveness of the self-assessed rheumatoid arthritis disease activity index to a flare of disease activity. *Arthritis Rheum* 2001;44:53–60.
49 Fransen J, Langenegger T, Michel BA, *et al*. Feasibility and validity of the radai, a self-administered rheumatoid arthritis disease activity index. *Rheumatology* 2000;39:321–7.
50 Fujiwara M, Kita Y. Reexamination of the assessment criteria for rheumatoid arthritis disease activity based on comparison of the disease activity score 28 with other simpler assessment methods. *Mod Rheumatol* 2013;23:260–8.
51 Houssien DA, Stucki G, Scott DL. A patient-derived disease activity score can substitute for a physician-derived disease activity score in clinical research. *Rheumatology* 1999;38:48–52.
52 Leeb BF, Haindl PM, Maktari A, *et al*. Patient-Centered rheumatoid arthritis disease activity assessment by a modified radai. *J Rheumatol* 2008;35:1294–9.
53 Leeb BF, Sautner J, Mai HTH, *et al*. A comparison of patient questionnaires and composite indexes in routine care of rheumatoid arthritis patients. *Joint Bone Spine* 2009;76:658–64.
54 Pincus T, Furer V, Keystone E, *et al*. RAPID3 (routine assessment of patient index data 3) severity categories and response criteria: similar results to DAS28 (disease activity score) and CDAI (clinical disease activity index) in the rapid 1 (rheumatoid arthritis prevention of structural damage) clinical trial of certolizumab pegol. *Arthritis Care Res* 2011;63:1142–9.
55 Pincus T, Swearingen CJ, Bergman M, *et al*. RAPID3 (routine assessment of patient index data 3), a rheumatoid arthritis index without formal joint counts for routine care: proposed severity categories compared to disease activity score and clinical disease activity index categories. *J Rheumatol* 2008;35:2136–47.
56 Pincus T, Swearingen CJ, Bergman MJ, *et al*. RAPID3 (routine assessment of patient index data) on an MDHAQ (multidimensional health assessment questionnaire): agreement with DAS28 (disease activity score) and CDAI (clinical disease activity index) activity categories, scored in five versus more than ninety seconds. *Arthritis Care Res* 2010;62:181–9.
57 Rintelen B, Sautner J, Haindl P, *et al*. Remission in rheumatoid arthritis: a comparison of the 2 newly proposed ACR/EULAR remission criteria with the rheumatoid arthritis disease activity index-5, a patient self-report disease activity index. *J Rheumatol* 2013;40:394–400.
58 Salaffi F, Ciapetti A, Gasparini S, *et al*. The comparative responsiveness of the patient self-report questionnaires and composite disease indices for assessing rheumatoid arthritis activity in routine care. *Clin Exp Rheumatol* 2012;30:912–21.
59 Salaffi F, Migliore A, Scarpellini M, *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;28:186–200.
60 Singh H, Gupta V, Ray S, *et al*. Evaluation of disease activity in rheumatoid arthritis by routine assessment of patient index data 3 (RAPID3) and its correlation to disease activity score 28 (DAS28)

and clinical disease activity index (CDAI): an Indian experience. *Clin Rheumatol* 2012;31:1663–9.
61 Stucki G, Liang MH, Stucki S, *et al*. A self-administered rheumatoid arthritis disease activity index (radai) for epidemiologic research. psychometric properties and correlation with parameters of disease activity. *Arthritis Rheum* 1995;38:795–8.
62 Sullivan MB, Iannaccone C, Cui J, *et al*. Evaluation of selected rheumatoid arthritis activity scores for office-based assessment. *J Rheumatol* 2010;37:2466–8.
63 Uhlig T, Kvien TK, Pincus T. Test-Retest reliability of disease activity core set measures and indices in rheumatoid arthritis. *Ann Rheum Dis* 2009;68:972–5.
64 Veehof MM, ten Klooster PM, Taal E, *et al*. Psychometric properties of the rheumatoid arthritis disease activity index (radai) in a cohort of consecutive Dutch patients with RA starting anti-tumour necrosis factor treatment. *Ann Rheum Dis* 2008;67:789–93.
65 Wolfe F, Michaud K, Pincus T. A composite disease activity scale for clinical practice, observational studies, and clinical trials: the patient activity scale (PAS/PAS-II). *J Rheumatol* 2005;32:2410–5.
66 Boone NW, Teeuwisse P, van der Kuy P-H, *et al*. Evaluating patient reported outcomes in routine practice of patients with rheumatoid arthritis treated with biological disease modifying anti rheumatic drugs (b-DMARDs). *Springerplus* 2015;4:462.
67 Che H, Combe B, Morel J, *et al*. Performance of patient-reported outcomes in the assessment of rheumatoid arthritis disease activity:the experience of the ESPOIR cohort. *Clin Exp Rheumatol* 2016;34:646–54.
68 Katz P, Margaretten M, Gregorich S, *et al*. Physical activity to reduce fatigue in rheumatoid arthritis: a randomized controlled trial. *Arthritis Care Res* 2018;70:1–10.
69 Kim S-K, Park S-H, Bae J, *et al*. Performance of routine assessment of patient index data 3 (RAPID3) for assessment of rheumatoid arthritis in clinical practice: differential agreement of RAPID3 according to disease activity categories. *Rheumatol Int* 2014;34:1311–8.
70 Kumar BS, Suneetha P, Mohan A, *et al*. Comparison of Disease Activity Score in 28 joints with ESR (DAS28), Clinical Disease Activity Index (CDAI), Health Assessment Questionnaire Disability Index (HAQ-DI) & Routine Assessment of Patient Index Data with 3 measures (RAPID3) for assessing disease activity in patients with rheumatoid arthritis at initial presentation. *Indian J Med Res* 2017;146:S57–62.
71 Muñoz JGB, Giraldo RB, Santos AM, *et al*. Correlation between rapid-3, DAS28, CDAI and SDAI as a measure of disease activity in a cohort of Colombian patients with rheumatoid arthritis. *Clin Rheumatol* 2017;36:1143–8.
72 Qorolli M, Hundozi-Hysenaj H, Rexhepi S, *et al*. RAPID3 scores and hand outcome measurements in RA patients: a preliminary study. *Clin Rheumatol* 2017;36:1379–85.
73 Salaffi F, Carotti M, Gutierrez M, *et al*. Patient acceptable symptom state in self-report questionnaires and composite clinical disease index for assessing rheumatoid arthritis activity: identification of cut-off points for routine care. *Biomed Res Int* 2015;2015:930756.
74 Walker UA, Mueller RB, Jaeger VK, *et al*. Disease activity dynamics in rheumatoid arthritis: patients' self-assessment of disease activity via WebApp. *Rheumatology* 2017;56:1707–12.
75 Yokogawa N, Kaneko T, Nagai Y, *et al*. Validation of RAPID3 using a Japanese version of multidimensional health assessment questionnaire with Japanese rheumatoid arthritis patients: characteristics of RAPID3 compared to DAS28 and CDAI. *Mod Rheumatol* 2015;25:264–9.
76 Pope J, Bingham CO, Fleischmann RM, *et al*. Impact of certolizumab pegol on patient-reported outcomes in rheumatoid arthritis and correlation with clinical measures of disease activity. *Arthritis Res Ther* 2015;17:343.
77 Macdermid JC, Fox E, Richards RS, *et al*. Validity of pulp-to-palm distance as a measure of finger flexion. *J Hand Surg Br* 2001;26:432–5.
78 de Vet HCW, Terwee CB, Mokkink LB, *et al*. *Measurement in medicine: a practical guide*. Cambridge University Press, 2011.
79 American College of Rheumatology Committee to Reevaluate Improvement Criteria. A proposed revision to the ACR20: the hybrid measure of American College of rheumatology response. *Arthritis Rheum* 2007;57:193–202.
80 England BR, Tiong BK, Bergman MJ, *et al*. 2019 update of the American College of rheumatology recommended rheumatoid arthritis disease activity measures. *Arthritis Care Res* 2019;71:1540–55.
81 Anderson J, Caplan L, Yazdany J, *et al*. Rheumatoid arthritis disease activity measures: American College of rheumatology recommendations for use in clinical practice. *Arthritis Care Res* 2012;64:640–7.
82 Braaksma C, Wolterbeek N, Veen MR, *et al*. Systematic review and meta-analysis of measurement properties of the Hip disability and Osteoarthritis Outcome Score - Physical Function Shortform (HOOS-PS) and the Knee Injury and Osteoarthritis Outcome Score - Physical Function Shortform (KOOS-PS). *Osteoarthritis Cartilage* 2020;28:1525–38.
83 Keystone E, Heijde Dvander, Mason D, *et al*. Certolizumab pegol plus methotrexate is significantly more effective than placebo plus methotrexate in active rheumatoid arthritis: findings of a fifty-two-week, phase III, multicenter, randomized, double-blind, placebo-controlled, parallel-group study. *Arthritis Rheum* 2008;58:3319–29.
84 van der Heijde DM, van't Hof MA, van Riel PL, *et al*. Validity of single variables and composite indices for measuring disease activity in rheumatoid arthritis. *Ann Rheum Dis* 1992;51:177–81.