Measuring Pain and Efficacy of Pain Treatment in Inflammatory Arthritis: A Systematic Literature Review

MATTHIAS ENGLBRECHT, INGO H. TARNER, DÉSIRÉE M. van der HEIJDE, BERNHARD MANGER, CLAIRE BOMBARDIER, and ULF MÜLLER-LADNER

ABSTRACT. *Objective.* To systematically review the available literature on measuring pain and the efficacy of pain treatment in inflammatory arthritis (IA), as an evidence base for generating clinical practice recommendations.

Methods. A systematic literature search was performed in Medline, Embase, Cochrane Library, and the American College of Rheumatology/European League Against Rheumatism 2008/2009 meeting abstracts, searching for studies evaluating clinimetric properties of pain measurement tools in IA (convergent validity, internal consistency, retest reliability, responsiveness, feasibility, and standardization). Studies that presented information on these properties were reviewed and their data were integrated into the pool of results available for pain measures in IA.

Results. In total, 51 articles were included in the review. Validated information on pain was available for tools covering different facets such as overall pain, anatomically specific pain, or a mixture of both. Data from these studies showed that single pain-related items such as the visual analog scale (VAS), numeric rating scale (NRS), or verbal rating scale (VRS) provide sufficient clinimetric information. Similar results were obtained for the pain subscales of the Arthritis Impact Measurement Scales (AIMS/AIMS2) and the bodily pain subscale of the Medical Outcome Study Short-Form Survey 36. Most clinimetric coefficients showed acceptable results with respect to validity, reliability, and sensitivity to change, while the degree of standardization and feasibility mostly filled at least 2 of 3 predefined criteria.

Conclusion. A variety of pain measures are available to cover different aspects of pain such as intensity, frequency, or location. Single-item tools such as VAS, NRS, or VRS can be recommended to measure overall pain in clinical practice. If more specific issues need to be addressed, more sophisticated tools should be taken into account. (J Rheumatol Suppl. 2012 Sept;90:3–10; doi:10.3899/jrheum.120335)

Key Indexing Terms:
PAIN MEASUREMENT   PAIN   ARTHRITIS   VALIDITY
RELIABILITY   PSYCHOMETRIC PROPERTIES

The perception of pain is common in inflammatory conditions such as rheumatoid arthritis (RA) or ankylosing spondylitis (AS). Hence, pain is perceived as the predominant characteristic when patients are asked which symptom of their arthritis deserves most improvement[1]. Additionally, painful inflammatory conditions limit daily activities of patients, reduce their opportunity for social interaction, and significantly reduce quality of life. With pain having become part of the core sets for classifying the severity of various rheumatic diseases the question of how to measure pain effectively has also become important. In the context of inflammatory arthritis (IA), pain is a construct that is likely to comprise several characteristics such as intensity, frequency, duration, and location. During the last 3 decades, various pain measurement tools have been introduced to cover these aspects and to provide reliable assessments. This article is part of the 3e (Evidence, Expertise, Exchange) Initiative on Pain Management by Pharmacotherapy in Inflammatory Arthritis[2]; our objective was to systematically review the available literature by addressing one of 10 selected questions as an evidence base for generating recommendations: "How do we measure pain and how do we monitor effectiveness of pain treatment in inflammatory arthritis?". The answer to this question supports

From the Department of Internal Medicine 3, University of Erlangen-Nuremberg, Erlangen, Germany; Department of Internal Medicine and Rheumatology, Justus-Liebig University Giessen; Department of Rheumatology, Clinical Immunology, Osteology and Physical Medicine, Kerckhoff Klinik Bad Nauheim, Bad Nauheim, Germany; Rheumatology Department, Leiden University Medical Center, Leiden, The Netherlands; and Institute for Work and Health, Toronto, Ontario, Canada.

M. Englbrecht, Dipl-Psych Univ, Department of Internal Medicine 3, University of Erlangen-Nuremberg; I.H. Tarner, MD, Department of Internal Medicine and Rheumatology, Justus-Liebig University Giessen, Department of Rheumatology, Clinical Immunology, Osteology and Physical Medicine, Kerckhoff Klinik Bad Nauheim; D.M. van der Heijde, MD, PhD, Professor, Rheumatology Department, Leiden University Medical Center; B. Manger, MD, Professor, Department of Internal Medicine 3, University of Erlangen-Nuremberg; C. Bombardier, MD, FRCPC, Professor, Institute for Work and Health; U. Müller-Ladner, MD, Professor, Department of Internal Medicine and Rheumatology, Justus-Liebig University Giessen, Department of Rheumatology, Clinical Immunology, Osteology and Physical Medicine, Kerckhoff Klinik Bad Nauheim.

Address correspondence to M. Englbrecht, Department of Internal Medicine 3 and Institute for Clinical Immunology, University of Erlangen-Nuremberg, Ulmenweg 18, 91054 Erlangen, Germany; E-mail: matthias.englbrecht@uk-erlangen.de

the purpose of this systematic literature review (SLR) to present and compare clinimetric properties of pain measures as well as to derive a recommendation for use of these tools in clinical practice.

METHODS

This SLR was performed according to the guidelines of the Cochrane Handbook for Systematic Literature Reviews[3] and the general requirements determining clinimetric properties[4].

Rephrasing the question — applying the PICO concept. The original question was rephrased to make it correspond to the PICO (Population, Intervention, Comparison, Outcome) concept for SLR suggested by the Cochrane Collaboration. In this context, the population (P) was defined as adults (≥ 18 yrs) that had a definite diagnosis of IA [e.g., RA, spondyloarthritis (SpA), inflammatory bowel disease, or reactive arthritis], while patient-reported outcomes (i.e., pain questionnaires or pain scales) were defined as the intervention (I). Assessments of tenderness of joints or tendons as well as pain behavior observations or special imaging techniques were excluded from this SLR by steering committee decision. In contrast to SLR that assess efficacy or safety of a drug, the validation of a questionnaire for single items cannot be done by applying a randomized controlled trial design or evaluating corresponding levels of evidence. Therefore, the definition of a comparison condition (C) was not useful since that would have precluded a pain measure. In the context of this SLR, quantifiable information on clinimetric properties such as validity, reliability, or responsiveness is needed to evaluate the quality of a given tool. According to these requirements, the outcomes of interest (O) for this SLR were convergent validity, internal consistency, retest reliability, responsiveness, and determination of clinically meaningful differences of pain measures. Additionally, each instrument was evaluated with respect to its degree of feasibility and standardization, based on previously defined criteria that referred to the clinimetric concept of Feinstein[4].

Clinimetric properties and corresponding coefficients. The evaluation of a pain measurement tool involves several clinimetric and statistical criteria. First, an assessment of the degrees of feasibility and standardization of each pain measure was done by a grading method that checked the presence of previously defined criteria and resulted in a score ranging from 0 (worst) to 3 (best). The checklist on feasibility contained administration time, user friendliness, and general acceptability as criteria of relevance, whereas the checklist on standardization appraised the presence of standardized instructions for test administration, calculation of test scores, and interpretation of test results. For each of these criteria that was fulfilled one additional point was added to the grading of feasibility or standardization, respectively, resulting in a maximum score of 3 for each of these domains.

Convergent validity, as one of the important clinimetric properties of interest, is a subset of construct validity, whereas instruments measuring the same construct (e.g., pain in IA) were expected to be highly correlated. The corresponding coefficient of convergent validity is the Bravais-Pearson correlation[5,6] also known as product-moment correlation. Cronbach α[7], on the other hand, is the most common index of internal consistency. In this context, good internal consistency implies that items forming a pain measurement questionnaire to assess the same facet of pain should show consistent low, medium, or high ratings according to the individual condition of each patient. Similar to convergent validity, retest reliability (also known as test-retest or test-test reliability) can also be assessed via product-moment correlation. However, in the context of retest reliability, it is not the scores of different measures (as is the case with convergent validity) but rather, the results of each score at different points in time that are correlated to reflect score stability. The intraclass correlation coefficient (ICC)[8] is another indicator of retest reliability and reflects systematic deviations between 2 subsequent ratings of the same measure. Thus, in conditions implying no stability of treatment one would expect good retest reliability resulting in corresponding coefficients. Responsiveness is the opposite of retest reliability: its purpose is to reflect the sensitivity to change of a measure in a given condition, for instance, the change of pain levels in IA after alteration of analgesic therapy. Standardized response means (SRM; i.e., the mean change in score divided by the standard deviation of the change) and effect sizes (ES)[9] are common indicators of these changes. If more than one result was reported on the same coefficient in different articles, the range of the coefficients is stated. Information on available definitions of relevant changes in pain measurement scores was collected as an additional specification.

Literature search and inclusion criteria for relevant scientific contributions. An SLR was performed in order to identify relevant articles published from 1947 to 2010 and indexed in Medline, Embase, or the Cochrane Library. Additionally, conference abstracts submitted for the 2008 and 2009 annual scientific meetings of the European League Against Rheumatism (EULAR) and the American College of Rheumatology (ACR) were hand-searched and reviewed. The first part of the search terms defining the population to be searched was a highly sensitive term list received from an experienced librarian from the Cochrane Collaboration. The subsequent parts of the search strategy were developed in collaboration with 2 other experienced librarians from the University of Erlangen-Nuremberg (for a complete search strategy see online Appendix, available from www.3epain.com). In order to correctly identify relevant articles a 2-step literature screening process was performed. First, titles and abstracts of articles retrieved by the search term strategy were screened with respect to clinimetric properties of pain measures used in IA. Second, for those articles identified as being potentially relevant, a review of the full text was performed. Both subsequent steps implied the following inclusion criteria: (1) articles were written in one of the predefined languages (i.e., Dutch, English, French, German, Portuguese, or Spanish); (2) the cohort was restricted to adult patients (≥ 18 years of age) with definite IA; (3) clinimetric properties corresponding to the aforementioned outcomes of interest and presented for at least one distinct IA subgroup if not for the whole cohort; and (4) the pain measurement tool had to be a patient-reported outcome (PRO) without any involvement of a third party. Articles that did not fulfil all of the aforementioned criteria were excluded from the systematic review process. The selection process was done by the first author and results were double-checked by mentors.

RESULTS

In total, the search strategy retrieved 3313 records from Medline, Embase, and the Cochrane Library. Additionally, 1491 abstracts submitted for EULAR and ACR conferences in 2008 and 2009 were hand-searched and screened. After the first step of literature exclusion, 140 journal articles and 28 conference abstracts remained for detailed full-text review, in which 51 were identified as presenting relevant information on clinimetric properties (see Figure 1). None of the 117 scientific contributions excluded in the full-text review was written in a language that was not supported by the present 3e Initiative [more information on the PICO method and a list of articles excluded during the full-text review are available online as supplementary material (Appendix on www.3epain.com)].

General information on included pain measurement tools. Although some of the pain measurement tools were dedicated to a certain indication [i.e., the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) or Disease Activity Index for Reactive Arthritis (DAREA)], the majority of instruments presented clinimetric data obtained from rheumatoid arthritis cohorts. The only exception to this was the visual analog scale (VAS) for overall pain intensity, which also provided data for ankylosing spondylitis (AS), and one suggestion on how to measure the minimally important difference in spondy-
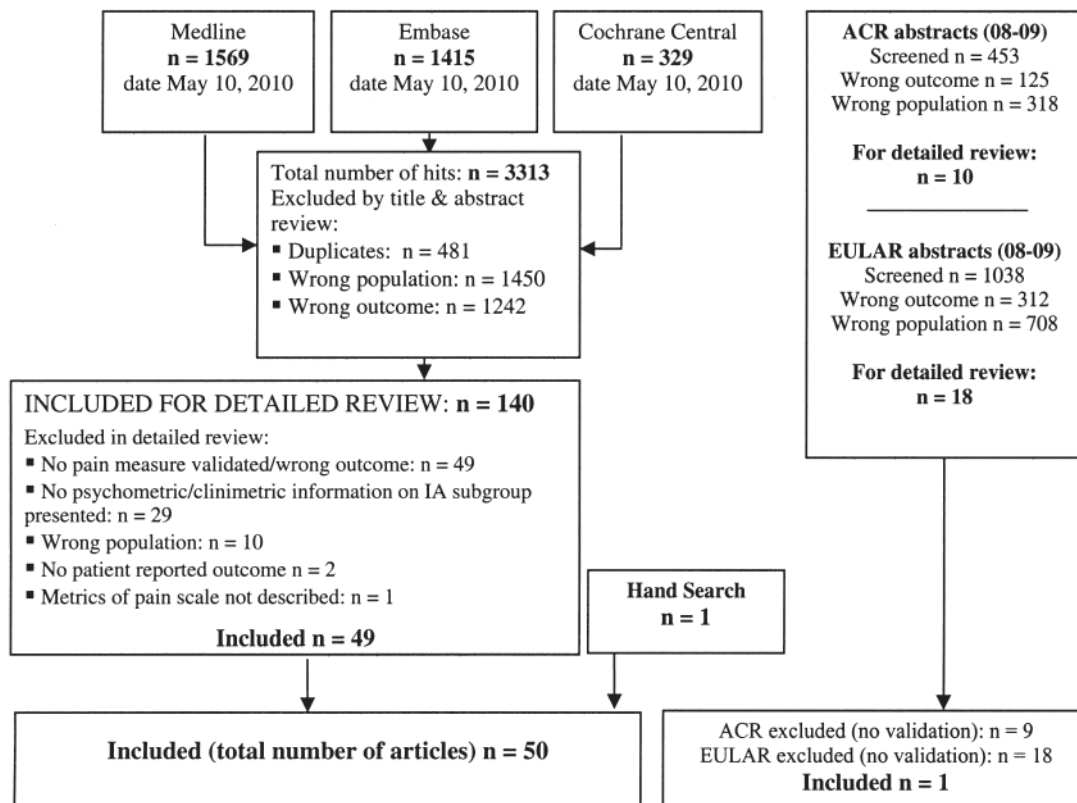
Figure 1. Literature search from which 141 articles and 28 abstracts were selected for detailed review. Fifty-one articles met the inclusion criteria.

loarthritis and the smallest detectable difference in AS. All instruments returned using the search strategy could be grouped into 3 categories: (1) overall pain measures (n = 14); (2) combination tools measuring overall pain and anatomically-specific pain in designated proximal or distal areas (n = 2); and (3) tools assessing merely anatomically-specific pain (n = 4). The results for the different categories of pain measures are indicated by subheadings in the corresponding tables (see Tables 1 and 2; clinimetric data presented below refer strictly to the scales, questionnaires, or subscales on pain as described in Tables 1 and 2).

Feasibility and standardization of included tools. The scores derived from the 2 checklists revealed that for most of the tools at least 2 out of 3 criteria of feasibility and standardization were fulfilled. The characteristic of general acceptability (as an item of feasibility) was fulfilled by all instruments. There were no reports of incomprehensibility and no statements of patients refusing to fill out pain-related items for any reason. However, the Index bidimensionnel de la douleur (IBD) and the bodily pain subscale of the Medical Outcome Study Short-Form 36 (SF-36 BP) have 2 different levels of pain rating scales, making them less user friendly. The McGill Pain Questionnaire (MPQ) and the Rheumatoid Arthritis Pain Score (RAPS) on the other hand were found to be more laborious to evaluate pain status since pain is only one of several

core measures that are suggested to be monitored in clinical routine investigation of IA. With regard to standardization, all pain measurement tools provided standardized instruction for test application and calculation of test scores, except for the publication on RAPS, which did not provide adequate instruction for calculation of test scores. Nevertheless, especially those instruments covering multiple facets of pain were found to be lacking additional information on how to interpret the resulting individual total score between minimum and maximum scores, or to provide certain information on the meaning of each pain domain with respect to the corresponding score.

Convergent validity, reliability, and responsiveness. Except for the product-moment correlation between the pain subscales of the Arthritis Impact Measurement Scales 2 (AIMS2) and the Rheumatoid Arthritis Outcome Score (RAOS), all coefficients showed an absolute value of r ≥ 0.50 indicating satisfactory convergent validity and a shared variance of at least 27%. As expected, the correlations between SF-36 BP scale and comparative pain measures were negative ($r_{SF-AIMS2}$ = –0.69 and r_{VAS} = –0.80, respectively) due to the inverse direction of the scale, with higher values showing less pain. The VAS for overall pain intensity was the most frequently comparable pain measure, leading to a variety of available coefficients for this scale. Where reported, the majority of the internal consistency coefficients were found to be ≥ 0.70,

Table 1. General information on included pain measurement tools and corresponding grading of feasibility and standardization.

Tool	Description	Score	F	S	I
		Overall Pain Measures			
VAS	Line of 10 cm/100 mm	0–10 or 0–100, no pain to worst pain	3	3	RA, SpA, AS
NRS	11-point numerical scale	0–10 no pain to worst pain	3	3	RA
VRS	Mostly 5-point Likert scale	0–4 or 1–5, no pain to worst pain	3	3	RA
MPQ	3 domains; 107 items: description of pain feeling; change in pain over time; pain intensity	Total score: 0–78 no pain to maximum pain	2	2	RA
SF-MPQ	1 domain (0–3; 15 items): description of pain feeling	Total score: 0–45, no pain to maximum pain	3	2	RA
IBD	2 domains: pain intensity (0–4); pain frequency (0–5)	Total score: 0–20, best to worst	2	2	RA
SF-36, pain subscale	2 items: pain intensity (1–6); impact on daily living (1–5)	0–100, worst to best	2	2	RA
AIMS, pain subscale	4 items (1–6): frequency of severe arthritis pain (last month); description of arthritis pain; duration of morning stiffness; pain in 2 or more joints at the same time	0–10 (standardized), best to worst	3	3	RA
AIMS2, pain subscale	4 items (1–5): frequency of severe arthritis pain; frequency of morning stiffness > 1 hour; frequency of sleep difficulty due to pain; usual pain from arthritis	0–10 (standardized), best to worst	3	3	RA
SF-AIMS, pain subscale	First 2 items of AIMS	0–10 (standardized), best to worst	3	3	RA
SF-AIMS2, pain subscale	First 3 items of AIMS2	0–10 (standardized), best to worst	3	3	RA
DAREA, pain item	Single item (0 = none to 2 = severe): patient's pain assessment	—	3	3	ReA
ADL, pain subscale	8 items (1–4): frequency of perceived pain during various activities	1–4, best to worst	3	3	RA
NHP, pain subscale	8 items (0 = yes or 1 = no): pain at night; unbearable pain; presence of pain during various activities	0–8, worst to best	3	2	RA
		Combination Measures (overall pain and anatomically-specific pain)			
RAPS	4 domains (0–6; 24 items): physiologic component; affective component; sensory component; cognitive component	Not reported	2	1	RA
RADAI, pain items	2 items: intensity of general arthritis pain (0–10); joint list (0–3 for each joint)	No pain score available	3	3	RA
		Anatomically-specific Pain Measures			
FFI, pain subscale	9 items (0–9): severity of foot pain during various activities	0–9, best to worst	3	2	RA
MHQ, pain subscale	5 items (1–5): frequency of hand pain during various activities	5–25 (unweighted), best to worst	3	2	RA
RAOS, pain subscale	9 items (0–4): frequency of hip, knee, or foot pain; intensity of hip, knee, or foot pain during last week	0–100 (standardized), worst to best	3	2	RA
BASDAI, pain items	2 items (0–10): neck, back, and hip pain; pain or swelling in other joints	0–10 (for each item) best to worst	3	2	AS

F: grading of feasibility (range 0 to 3); S: grading of standardization (range 0 to 3); I: indication of clinimetric properties; VAS: visual analog scale; NRS: numerical rating scale; VRS: verbal rating scale; MPQ: McGill Pain Questionnaire; SF-MPQ: Short-Form McGill Pain Questionnaire; IBD: Index bidimensionnel de la douleur; SF-36: Short Form 36; AIMS: Arthritis Impact Measurement Scales; SF-AIMS: Short-Form AIMS: DAREA: Disease Activity Index for Reactive Arthritis; ADL: Activities of Daily Living; NHP: Nottingham Health Profile; RAPS: Rheumatoid Arthritis Pain Scale; RADAI: Rheumatoid Arthritis Disease Activity Index; FFI: Foot Function Index; MHQ: Michigan Hand Outcomes Questionnaire; RAOS: Rheumatoid Arthritis Outcome Score; BASDAI: Bath Ankylosing Spondylitis Disease Activity Index.

Table 2. Clinimetric properties of pain measurement tools.

Tool	I	Convergent Validity	Internal Consistency (α)	Retest Reliability	Sensitivity to Change	Defined Differences	References
Overall Pain Measures							
VAS	RA	$r_{ADL} = 0.54$ $r_{IBD} = 0.56$ $r_{MPQ} = 0.54$ $r_{RAPS} = 0.68$ $r_{SF\text{-}MPQ} = 0.53$ $r_{NHP(pain)} = 0.62$ $r_{SF\text{-}36(BP)} = -0.80$ $r_{AIMS(pain)} = 0.75$ $r_{AIMS2(pain)} = 0.66$ $r_{SF\text{-}AIMS2(pain)} = 0.76$ $r_{SF\text{-}AIMS2(pain)} = 0.59\text{--}0.60$	—	$r = 0.59\text{--}0.94$	SRM 0.7–1.05 ES 0.24–1.6	CIC	12–35
	SpA	—	—	—	—	MID	
	AS	—	—	ICC[†] 0.78–0.93/0.75–0.92	—	SDD	
NRS	RA	$r_{SF\text{-}MPQ} = 0.64\text{--}0.70$	—	$r = 0.95\text{--}0.96$	SRM 0.75–1.08 ES 0.81	—	15,21,25,36
VRS	RA	—	—	$r = 0.82\text{--}0.90$	SRM 0.97 ES –0.26	—	21,28
MPQ	RA	$r_{VAS} = 0.54$	0.87	$r = 0.70$	SRM 0.68 ES 0.19	—	15,16,30
SF-MPQ	RA	$r_{VAS} = 0.53$ $r_{NRS} = 0.64\text{--}0.70$ $r_{AIMS} = 0.52$	—	ICC 0.89 (95% CI 0.84–0.93)	—	—	37,38
IBD	RA	$r_{VAS} = 0.56$	—	—	—	—	39
SF-36, bodily pain subscale	RA	$r_{SF\text{-}AIMS2} = -0.69$ $r_{VAS} = -0.80$	—	ICC 0.90 (95% CI 0.84–0.93)	SRM 0.44–1.69 ES 0.29–1.07	MIC	13,24,36,40–42
AIMS, pain subscale	RA	$r_{VAS} = 0.75$ $r_{SF\text{-}MPQ} = 0.52$	0.82–0.90	$r = 0.36\text{--}0.51$	SRM 1.2–1.8 ES 1.6–2.3	—	17,31,37,43–48
AIMS2, pain subscale	RA	$r_{VAS} = 0.66$ $r_{RAOS} = 0.36$	0.82	ICC 0.75–0.88	SRM 0.82–2.67 ES 0.32–1.66	—	12,27,36,49,50
SF-AIMS, pain subscale	RA	$r_{VAS} = 0.76$	0.84–0.91	$r = 0.19\text{--}0.38$	—	—	45,48
SF-AIMS2, pain subscale	RA	$r_{VAS} = 0.59\text{--}0.60$ $r_{SF\text{-}36} = -0.69$	0.75–0.84	ICC 0.71–0.83	—	—	13,27
DAREA, pain item	ReA	—	—	—	ES 2.72	—	20
ADL, pain subscale	RA	$r_{VAS} = 0.54$	—	$r = 0.81$	—	—	18
NHP, pain subscale	RA	$r_{VAS} = 0.62$	—	—	—	—	23,51
Combination Measures (overall pain and anatomically-specific pain)							
RAPS	RA	$r_{VAS} = 0.68$	0.64–0.86	—	—	—	52
RADAI, pain items	RA	$r_{items} = 0.67$	—	—	—	—	53
Anatomically-specific Pain Measures							
FFI, pain subscale	RA	—	0.95	ICC 0.7–0.89	—	—	54,55
MHQ, pain subscale	RA	—	—	—	SRM 0.15 ES 0.19	MCID	56,57
RAOS, pain subscale	RA	$r_{AIMS2(pain)} = 0.36$	0.90	ICC 0.88	ES –0.28	—	58
BASDAI, pain items	AS	—	—	—	SRM_{item1} 0.94–1.24 SRM_{item2} 0.55–1.18	—	59–62

[†] Spinal pain last week/nocturnal spinal pain last week. I: indication of clinimetric properties; r: Pearson correlation coefficient; ICC: intraclass correlation coefficient; SRM: standardized response mean; ES: effect size; MIC: minimally important change; MCID: minimal clinically important difference; CIC: clinically important change; MID: minimally important difference; SDD: smallest detectable difference; VAS: visual analog scale; NRS: numerical rating scale; VRS: verbal rating scale; MPQ: McGill Pain Questionnaire; SF-MPQ: Short-Form McGill Pain Questionnaire; IBD: Index bidimensionnel de la douleur; SF-36: Short Form 36; AIMS: Arthritis Impact Measurement Scales; SF-AIMS: Short-Form AIMS: DAREA: Disease Activity Index for Reactive Arthritis; ADL: Activities of Daily Living; NHP: Nottingham Health Profile; RAPS: Rheumatoid Arthritis Pain Scale; RADAI: Rheumatoid Arthritis Disease Activity Index; FFI: Foot Function Index; MHQ: Michigan Hand Outcomes Questionnaire; RAOS: Rheumatoid Arthritis Outcome Score; BASDAI: Bath Ankylosing Spondylitis Disease Activity Index.

which is regarded as being satisfactorily consistent[10]. Due to the approach of calculating internal consistency, which requires at least 2 items, there were no coefficients for single-item pain measures such as the VAS, the verbal rating scale (VRS), or the numerical rating scale (NRS). The results for retest reliability were considerably heterogeneous depending on the coefficient and the underlying study design that was applied. While ICC showed satisfactory results of ≥ 0.7 according to suggestions in the literature[11], the product-moment correlation of 2 subsequent timepoints of pain measurement ranged from r = 0.15 (short form of the AIMS; SF-AIMS) to r = 0.96 (NRS). Unfortunately, for these considerably differing coefficients, there were no comparable ICC reported. However, a closer look at the study design revealed that the poor product-moment correlations of the AIMS and the SF-AIMS were associated with larger periods until followup assessment (i.e., ≥ 6 months), whereas the 2 assessments of the NRS had been done before and immediately after a regular medical consultation. A similar pattern of heterogeneity could be seen from the data on responsiveness that were available. In this context, measures of ES covered the whole spectrum of values ranging from negligible effects of < 0.20 to large effects of ≥ 0.80 for both ES and SRM. Results showing consistently large effects were reported for VAS (SRM), VRS (SRM), NRS (ES), AIMS (SRM and ES), AIMS2 (SRM), DAREA (ES), and the neck, back and hip pain item of the BASDAI (SRM). Additional data on SRM and ES varied depending on the corresponding followup period and the pain scale that was applied. Table 2 shows corresponding worst- and best-case scenarios for information on ES obtained from the literature. Definitions of relevant changes in pain perception were obtained only for VAS (clinically important change in RA, minimally important difference in SpA, and smallest detectable difference in AS), SF-36 BP (minimally important change in RA), and MPQ (minimal clinically important difference in RA).

DISCUSSION

This SLR summarizes and evaluates evidence from the literature on measuring pain and efficacy of pain treatment in IA. Combined with expert opinion from a broad panel of rheumatologists in the 3e Initiative, our results served as an evidence base for generating one of the 10 clinical recommendations on how to approach pain in IA[2].

From the available data, we were able to demonstrate that the VAS for overall pain intensity is currently the best-evaluated pain measure in RA. Moreover, the NRS and the VRS as alternative single-item pain measures are also a sensitive alternative with comparably good clinimetric profiles. Multidimensional tools including a well-evaluated pain subscale such as the different versions of the AIMS or the SF-36 can be a good approach to integrate the assessment of pain into a larger context; however, anatomically-specific pain measures are indicated for more sophisticated tasks in clinical research. Some readers might have expected the coefficients of convergent validity to show higher correlations in general. But with inflammatory pain having many facets (such as intensity, frequency, duration, location) and with pain measures applying various scaling techniques, the correlation coefficients obtained still appear to be of satisfactory value in accordance with the requirements mentioned in the literature[9]. However, apart from the field of RA, data on other disease-specific pain measures could be identified only for reactive arthritis and AS. The results showed that although a lot of information on several patient-derived outcomes measuring pain in IA was available, drawing a complete clinimetric profile of each instrument remains a challenge. Similarly to data on responsiveness or stability of pain measurement scores, results on the establishment of meaningful differences were available for only a limited number of instruments. Additionally, information for completing a clinimetric profile of a pain measurement tool often had to be taken from separate studies with different underlying designs and varying periods between followup assessments. As a consequence, data on some of the properties related to retest reliability or responsiveness show considerable heterogeneity, which must not be mistaken for poor outcomes due to an invalid tool, but needs to be considered in the light of different individual study designs. Thus, in view of comparable results for retest reliability and responsiveness, it seems to be more a question of when and how to measure pain than one of differences due to the instrument used. With some profiles of pain measures remaining poorly described and a standard terminology of meaningful differences still lacking, several research questions will have to be addressed in the future. Besides the task of completing clinimetric profiles of various pain measurement tools, the question of whether a meaningful difference remains stable across the whole scale of a tool will be of special interest in this context. For this SLR we decided to strictly limit the term "pain measures" to patient-derived outcomes that were obtained without any influence of a third party. Hence, information on clinimetric properties of tools containing a pain subscale was taken into further account only if distinct data on the pain subscales were presented. We are aware that pain is also closely related to other domains of daily living such as physical functioning or physical and mental well-being and may interact with these variables. Moreover, information on painful joints, for example, can be obtained by physical examination. However, the purpose of this SLR was to deduce information on pain measures in order to make a recommendation that is both a precise response regarding patient-reported outcome measures as well as useful guidance for clinical practice. Although clinimetric data on some tools are still incomplete, we were able to draw a conclusion from the information that was available.

In summary, taking the heterogeneous and partly-lacking data into account, it can be concluded that pain in IA should be measured routinely with validated scales such as the VAS,

NRS, or VRS. Multidimensional tools or anatomically-specific measures might be additionally considered where appropriate with respect to the individual research agenda. This conclusion was incorporated as one of the recommendations of the 3e Initiative on the measurement and treatment of pain in IA[2].

ACKNOWLEDGMENT

REFERENCES

1. Gossec L, Dougados M, Rincheval N, Balanescu A, Boumpas DT, Canadelo S, et al. Elaboration of the preliminary Rheumatoid Arthritis Impact of Disease (RAID) score: A EULAR initiative. Ann Rheum Dis 2009;68:1680-5.
2. Whittle SL, Colebatch AN, Buchbinder R, Edwards CJ, Adams K, Englbrecht M, et al. Multinational evidence-based recommendations for pain management by pharmacotherapy in inflammatory arthritis: Integrating systematic literature research and expert opinion of a broad panel of rheumatologists in the 3e Initiative. Rheumatology 2012; [in press].
3. Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions, Version 5.1.0 [updated March 2011; accessed April 18, 2012]. The Cochrane Collaboration, 2011. Available from: www.cochrane-handbook.org
4. Feinstein AR. An additional basic science for clinical medicine: IV. The development of clinimetrics. Ann Intern Med 1983;99:843-8.
5. Bravais A. Analyse mathematique sur les probabilités des erreurs de situation de point. Memoires presentes par divers savants a l'Academie des Sciences de l'institut de France. 1846;9:255-332.
6. Pearson K. On further methods of determining correlation. London: Dulau and Co., 1907. Cambridge: Draper's Company Research. Memoirs (Biometric Series IV) 1907: 39 pp.
7. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16:297-334.
8. Nunnally JC. Assessment of reliability. In: Nunnally JC, Bernstein IH, editors. Psychometric theory. New York: McGraw-Hill; 1978:225-55.
9. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
10. DeVon HA, Block ME, Moyle-Wright P, Ernst DM, Hayden SJ, Lazzara DJ, et al. A psychometric toolbox for testing validity and reliability. J Nurs Scholarsh 2007;39:155-64.
11. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res 2005;19:231-40.
12. Arkela-Kautiainen M, Kauppi M, Heikkila S, Kautiainen H, Malkia E, Leirisalo-Repo M. Evaluation of the Arthritis Impact Measurement Scales (AIMS2) in Finnish patients with rheumatoid arthritis. Scand J Rheumatol 2003;32:300-5.
13. Askary-Ashtiani AR, Mousavi SJ, Parnianpour M, Montazeri A. Translation and validation of the Persian version of the Arthritis Impact Measurement Scales 2-Short Form (AIMS2-SF) in patients with rheumatoid arthritis. Clin Rheumatol 2009;28:521-7.
14. Auleley GR, Benbouazza K, Spoorenberg A, Collantes E, Hajjaj-Hassouni N, van der Heijde DM, et al. Evaluation of the smallest detectable difference in outcome or process variables in ankylosing spondylitis. Arthritis Care Res 2002;47:582-7.
15. Bellamy N, Campbell J, Syrotuik J. Comparative study of self-rating pain scales in rheumatoid arthritis patients. Curr Med Res Opin 1999;15:121-7.
16. Bombardier C, Raboud J. A comparison of health-related quality-of-life measures for rheumatoid arthritis research. The Auranofin Cooperating Group. Control Clin Trials 1991;12 Suppl:243S-56S.
17. Buchbinder R, Bombardier C, Yeung M, Tugwell P. Which outcome measures should be used in rheumatoid arthritis clinical trials? Clinical and quality-of-life measures' responsiveness to treatment in a randomized controlled trial. Arthritis Rheum 1995;38:1568-80.
18. Callahan LF, Brooks RH, Summey JA, Pincus T. Quantitative pain assessment for routine care of rheumatoid arthritis patients, using a pain scale based on activities of daily living and a visual analog pain scale. Arthritis Rheum 1987;30:630-6.
19. Coates LC, Fransen J, Helliwell PS. Defining minimal disease activity in psoriatic arthritis: A proposed objective target for treatment. Ann Rheum Dis 2009;69:48-53.
20. Eberl G, Studnicka-Benke A, Hitzelhammer H, Gschnait F, Smolen JS. Development of a disease activity index for the assessment of reactive arthritis (DAREA). Rheumatology 2000;39:148-55.
21. Ferraz MB, Quaresma MR, Aquino LR, Atra E, Tugwell P, Goldsmith CH. Reliability of pain scales in the assessment of literate and illiterate patients with rheumatoid arthritis. J Rheumatol 1990;17:1022-4.
22. Goldsmith CH, Boers M, Bombardier C, Tugwell P. Criteria for clinically important changes in outcomes: Development, scoring and evaluation of rheumatoid arthritis patient and trial profiles. OMERACT Committee. J Rheumatol 1993;20:561-5.
23. Houssien DA, McKenna SP, Scott DL. The Nottingham Health Profile as a measure of disease activity and outcome in rheumatoid arthritis. Br J Rheumatol 1997;36:69-73.
24. Linde L, Sorensen J, Ostergaard M, Horslev-Petersen K, Hetland ML. Health-related quality of life: Validity, reliability, and responsiveness of SF-36, 15D, EQ-5D [corrected] RAQoL, and HAQ in patients with rheumatoid arthritis. J Rheumatol 2008;35:1528-37.
25. Pincus T, Bergman M, Sokka T, Roth J, Swearingen C, Yazici Y. Visual analog scales in formats other than a 10 centimeter horizontal line to assess pain and other clinical data. J Rheumatol 2008;35:1550-8.
26. Pope JE, Khanna D, Norrie D, Ouimet JM. The minimally important difference for the Health Assessment Questionnaire in rheumatoid arthritis clinical practice is smaller than in randomized controlled trials. J Rheumatol 2009;36:254-9.
27. Taal E, Rasker JJ, Riemsma RP. Psychometric properties of a Dutch short form of the Arthritis Impact Measurement Scales 2 (Dutch-AIMS2-SF). Rheumatology 2003;42:427-34.
28. ten Klooster PM, Vlaar AP, Taal E, Gheith RE, Rasker JJ, El-Garf AK, et al. The validity and reliability of the graphic rating scale and verbal rating scale for measuring pain across cultures: A study in Egyptian and Dutch women with rheumatoid arthritis. Clin J Pain 2006;22:827-30.
29. van der Heijde DM, van 't Hof M, van Riel PL, van de Putte LB. Validity of single variables and indices to measure disease activity in rheumatoid arthritis. J Rheumatol 1993;20:538-41.
30. van Lankveld W, van 't Pad Bosch P, van de Putte L, van der Staak C, Naring G. Pain in rheumatoid arthritis measured with the visual analogue scale and the Dutch version of the McGill Pain Questionnaire [Dutch]. Ned Tijdschr Geneeskd 1992;136:1166-70.
31. Verhoeven AC, Boers M, van der Linden S. Responsiveness of the core set, response criteria, and utilities in early rheumatoid arthritis. Ann Rheum Dis 2000;59:966-74.
32. Ward MM. Clinical measures in rheumatoid arthritis: Which are most useful in assessing patients? J Rheumatol 1994;21:17-27.
33. Ward MM, Guthrie LC. Patients' ratings of clinically important

changes in pain and physical function in rheumatoid arthritis [abstract]. Arthritis Rheum 2008;58 Suppl:S205.

34. Wells GA, Tugwell P, Kraag GR, Baker PR, Groh J, Redelmeier DA. Minimum important difference between patients with rheumatoid arthritis: the patient's perspective. J Rheumatol 1993;20:557-60.

35. Wheaton L, Pope J. The minimally important difference for patient-reported outcomes in spondyloarthropathies including pain, fatigue, sleep, and Health Assessment Questionnaire. J Rheumatol 2010;37:816-22.

36. Veehof MM, ten Klooster PM, Taal E, van Riel PL, van de Laar MA. Comparison of internal and external responsiveness of the generic Medical Outcome Study Short Form-36 (SF-36) with disease-specific measures in rheumatoid arthritis. J Rheumatol 2008;35:610-7.

37. Burckhardt CS, Bjelle A. A Swedish version of the short-form McGill Pain Questionnaire. Scand J Rheumatol 1994;23:77-81.

38. Yakut Y, Yakut E, Bayar K, Uygur F. Reliability and validity of the Turkish version short-form McGill Pain Questionnaire in patients with rheumatoid arthritis. Clin Rheumatol 2007;26:1083-7.

39. Harrewyn JM. Clinical evaluation of pain. Value of a 2-dimensional index of pain [French]. Rev Rhum Mal Osteoartic 1989;56:75-8.

40. Kosinski M, Zhao SZ, Dedhiya S, Osterhaus JT, Ware JE. Determining minimally important changes in generic and disease-specific health-related quality of life questionnaires in clinical trials of rheumatoid arthritis. Arthritis Rheum 2000;43:1478-87.

41. Ruta DA, Hurst NP, Kind P, Hunter M, Stubbings A. Measuring health status in British patients with rheumatoid arthritis: Reliability, validity and responsiveness of the short form 36-item health survey (SF-36). Br J Rheumatol 1998;37:425-36.

42. Salaffi F, Stancati A, Carotti M. Responsiveness of health status measures and utility-based methods in patients with rheumatoid arthritis. Clin Rheumatol 2002;21:478-87.

43. Fries JF, Spitz PW, Young DY. The dimensions of health outcomes: The Health Assessment Questionnaire, disability and pain scales. J Rheumatol 1982;9:789-93.

44. Hill J, Bird HA, Lawton CW, Wright V. The Arthritis Impact Measurement Scales: An anglicized version to assess the outcome of British patients with rheumatoid arthritis. Br J Rheumatol 1990;29:193-6.

45. Lorish CD, Abraham N, Austin JS, Bradley LA, Alarcon GS. A comparison of the full and short versions of the Arthritis Impact Measurement Scales. Arthritis Care Res 1991;4:168-73.

46. Meenan RF, Anderson JJ, Kazis LE, Egger MJ, Altz-Smith M, Samuelson CO Jr, et al. Outcome assessment in clinical trials. Evidence for the sensitivity of a health status measure. Arthritis Rheum 1984;27:1344-52.

47. Potts MK, Brandt KD. Evidence of the validity of the Arthritis Impact Measurement Scales. Arthritis Rheum 1987;30:93-6.

48. Wallston KA, Brown GK, Stein MJ, Dobbins CJ. Comparing the short and long versions of the Arthritis Impact Measurement Scales. J Rheumatol 1989;16:1105-9.

49. Husted J, Gladman DD, Farewell VT, Long JA. Validation of the revised and expanded version of the Arthritis Impact Measurement Scales for patients with psoriatic arthritis. J Rheumatol 1996;23:1015-9.

50. Riemsma RP, Taal E, Rasker JJ, Houtman PM, Van Paassen HC, Wiegman O. Evaluation of a Dutch version of the AIMS2 for patients with rheumatoid arthritis. Br J Rheumatol 1996;35:755-60.

51. Uutela T, Hakala M, Kautiainen H. Validity of the Nottingham Health Profile in a Finnish out-patient population with rheumatoid arthritis. Rheumatology 2003;42:841-5.

52. Anderson DL. Development of an instrument to measure pain in rheumatoid arthritis: Rheumatoid Arthritis Pain Scale (RAPS). Arthritis Care Res 2001;45:317-23.

53. Fransen J, Langenegger T, Michel BA, Stucki G. Feasibility and validity of the RADAI, a self-administered Rheumatoid Arthritis Disease Activity Index. Rheumatology 2000;39:321-7.

54. Budiman Mak E, Conrad KJ, Roach KE. The Foot Function Index: A measure of foot pain and disability. J Clin Epidemiol 1991;44:561-70.

55. Saag KG, Saltzman CL, Brown CK, Budiman-Mak E. The foot function index for measuring rheumatoid arthritis pain: Evaluating side-to-side reliability. Foot Ankle Int 1996;17:506-10.

56. Shauver MJ, Chung KC. The minimal clinically important difference of the Michigan Hand Outcomes Questionnaire. J Hand Surg 2009;34:509-14.

57. van der Giesen FJ, Nelissen RG, Arendzen JH, de Jong Z, Wolterbeek R, Vliet Vlieland TP. Responsiveness of the Michigan Hand Outcomes Questionnaire — Dutch language version in patients with rheumatoid arthritis. Arch Phys Med Rehabil 2008;89:1121-6.

58. Karatepe AG, Gunaydin R, Turkmen G, Kaya T, Ozbek G. The validity and reliability of the Turkish version of the Rheumatoid Arthritis Outcome Score (RAOS) in patients with rheumatoid arthritis. Clin Rheumatol 2009;28:47-51.

59. Calin A, Nakache JP, Gueguen A, Zeidler H, Mielants H, Dougados M. Defining disease activity in ankylosing spondylitis: Is a combination of variables (Bath Ankylosing Spondylitis Disease Activity Index) an appropriate instrument? Rheumatology 1999;38:878-82.

60. Fernandez-Sueiro JL, Willisch A, Pertega-Diaz S, Tasende JA, Fernandez-Lopez JC, Villar NO, et al. Validity of the Bath Ankylosing Spondylitis Disease Activity Index for the evaluation of disease activity in axial psoriatic arthritis. Arthritis Care Res 2010;62:78-85.

61. van der Heijde D, Lie E, Kvien TK, Sieper J, van den Bosch F, Listing J, et al. ASDAS, a highly discriminatory ASAS-endorsed disease activity score in patients with ankylosing spondylitis. Ann Rheum Dis 2009;68:1811-8.

62. Waldner A, Cronstedt H, Stenstrom CH. The Swedish version of the Bath Ankylosing Spondylitis Disease Activity Index. Reliability and validity. Scand J Rheumatol Suppl 1999;111:10-6.