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

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 improvement1. 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 Arthritis2; 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
METHODS
This SLR was performed according to the guidelines of the Cochrane Handbook for Systematic Literature Reviews\(^5\) 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, spondyloarthritides (SpA), inflammatory bowel disease, or reactive arthritides], 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\(^6\).

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\(^7\) also known as product-moment correlation. Cronbach \(\alpha\) 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 fulfill 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-
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 $r \geq 0.50$ 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 $\geq 0.70$,
<table>
<thead>
<tr>
<th>Tool</th>
<th>Description</th>
<th>Score</th>
<th>F</th>
<th>S</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overall Pain Measures</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VAS</td>
<td>Line of 10 cm/100 mm</td>
<td>0–10 or 0–100, no pain to worst pain</td>
<td>3</td>
<td>3</td>
<td>RA, SpA, AS</td>
</tr>
<tr>
<td>NRS</td>
<td>11-point numerical scale</td>
<td>0–10</td>
<td>3</td>
<td>3</td>
<td>RA</td>
</tr>
<tr>
<td>VRS</td>
<td>Mostly 5-point Likert scale</td>
<td>0–4 or 1–5, no pain to worst pain</td>
<td>3</td>
<td>3</td>
<td>RA</td>
</tr>
<tr>
<td>MPQ</td>
<td>3 domains; 107 items: description of pain feeling; change in pain over time;</td>
<td>Total score: 0–78, no pain to maximum pain</td>
<td>2</td>
<td>2</td>
<td>RA</td>
</tr>
<tr>
<td>SF-MPQ</td>
<td>1 domain (0–3; 15 items): description of pain feeling</td>
<td>Total score: 0–45, no pain to maximum pain</td>
<td>3</td>
<td>2</td>
<td>RA</td>
</tr>
<tr>
<td>IBD</td>
<td>2 domains: pain intensity (0–4); pain frequency (0–5)</td>
<td>Total score: 0–20, best to worst</td>
<td>2</td>
<td>2</td>
<td>RA</td>
</tr>
<tr>
<td>SF-36, pain subscale</td>
<td>2 items: pain intensity (1–6); impact on daily living (1–5)</td>
<td>0–100, worst to best</td>
<td>2</td>
<td>2</td>
<td>RA</td>
</tr>
<tr>
<td>AIMS, pain subscale</td>
<td>4 items (1–6): description of arthritis pain; duration of morning stiffness;</td>
<td>0–10 (standardized), best to worst</td>
<td>3</td>
<td>3</td>
<td>RA</td>
</tr>
<tr>
<td>AIMS2, pain subscale</td>
<td>4 items (1–5): frequency of severe arthritis pain; frequency of morning</td>
<td>0–10 (standardized), best to worst</td>
<td>3</td>
<td>3</td>
<td>RA</td>
</tr>
<tr>
<td>SF-AIMS, pain subscale</td>
<td>First 2 items of AIMS</td>
<td>0–10 (standardized), best to worst</td>
<td>3</td>
<td>3</td>
<td>RA</td>
</tr>
<tr>
<td>SF-AIMS2, pain subscale</td>
<td>First 3 items of AIMS2</td>
<td>0–10 (standardized), best to worst</td>
<td>3</td>
<td>3</td>
<td>RA</td>
</tr>
<tr>
<td>DAREA, pain item</td>
<td>Single item (0 = none to 2 = severe): patient’s pain assessment</td>
<td>—</td>
<td>3</td>
<td>3</td>
<td>ReA</td>
</tr>
<tr>
<td>ADL, pain subscale</td>
<td>8 items (1–4): frequency of perceived pain during various activities</td>
<td>1–4, best to worst</td>
<td>3</td>
<td>3</td>
<td>RA</td>
</tr>
<tr>
<td>NHP, pain subscale</td>
<td>8 items (0 = yes or 1 = no): pain at night; unbearable pain; presence of</td>
<td>0–8, worst to best</td>
<td>3</td>
<td>2</td>
<td>RA</td>
</tr>
<tr>
<td><strong>Combination Measures (overall pain and anatomically-specific pain)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAPS</td>
<td>4 domains (0–6; 24 items): physiologic component; affective component;</td>
<td>Not reported</td>
<td>2</td>
<td>1</td>
<td>RA</td>
</tr>
<tr>
<td>RAOS, pain items</td>
<td>sensory component; cognitive component</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Anatomically-specific Pain Measures</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FFI, pain subscale</td>
<td>9 items (0–9): severity of foot pain during various activities</td>
<td>0–9, best to worst</td>
<td>3</td>
<td>2</td>
<td>RA</td>
</tr>
<tr>
<td>RAOS, pain subscale</td>
<td>5 items (1–5): frequency of hand pain during various activities</td>
<td>5–25, (unweighted), best to worst</td>
<td>3</td>
<td>2</td>
<td>RA</td>
</tr>
<tr>
<td>RAOS, pain subscale</td>
<td>9 items (0–4): frequency of hip, knee, or foot pain; intensity of hip,</td>
<td>0–100, worst to best</td>
<td>3</td>
<td>2</td>
<td>RA</td>
</tr>
<tr>
<td>BASDAI, pain items</td>
<td>2 items (0–10): neck, back, and hip pain; pain or swelling in other joints</td>
<td>0–10, worst to best</td>
<td>3</td>
<td>2</td>
<td>AS</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Tool</th>
<th>I</th>
<th>Convergent Validity</th>
<th>Internal Consistency (α)</th>
<th>Retest Reliability</th>
<th>Sensitivity to Change</th>
<th>Defined Differences</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Overall Pain Measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VAS</td>
<td>RA</td>
<td>$r_{ADL} = 0.54$</td>
<td>—</td>
<td>$r = 0.59–0.94$</td>
<td>SRM 0.7–1.05</td>
<td>CIC</td>
<td>12–35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$r_{IBD} = 0.56$</td>
<td>$r_{MPQ} = 0.54$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$r_{RAPS} = 0.68$</td>
<td>$r_{SF-MPQ} = 0.53$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$r_{NHP(pain)} = 0.62$</td>
<td>$r_{5SF-36(BP)} = -0.80$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$r_{AIMS(pain)} = 0.75$</td>
<td>$r_{AIMS2(pain)} = 0.66$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$r_{SF-AIMS2(pain)} = 0.76$</td>
<td>$r_{SF-AIMS2(pain)} = 0.59–0.60$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SpA</td>
<td>—</td>
<td>—</td>
<td></td>
<td>MID</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AS</td>
<td></td>
<td>0.78–0.93/0.75–0.92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NRS</td>
<td>RA</td>
<td>$r_{SF-MPQ} = 0.64–0.70$</td>
<td>$r = 0.95–0.96$</td>
<td>SRM 0.75–1.08</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td>VRS</td>
<td>RA</td>
<td>—</td>
<td>$r = 0.82–0.90$</td>
<td>SRM 0.97</td>
<td>ES –0.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MPQ</td>
<td>RA</td>
<td>$r_{VAS} = 0.54$</td>
<td>0.87</td>
<td>$r = 0.70$</td>
<td>SRM 0.68</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SF-MPQ</td>
<td>RA</td>
<td>$r_{VAS} = 0.53$</td>
<td>—</td>
<td>ICC 0.89</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IBD</td>
<td>RA</td>
<td>$r_{VAS} = 0.56$</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SF-36, bodily pain subscale</td>
<td>RA</td>
<td>$r_{SF-AIMS2} = 0.69$</td>
<td>(95% CI 0.84–0.93)</td>
<td>$r = 0.36–0.51$</td>
<td>SRM 0.44–1.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AIMS, pain subscale</td>
<td>RA</td>
<td>$r_{VAS} = -0.80$</td>
<td>0.82–0.90</td>
<td>$r = 0.36–0.51$</td>
<td>SRM 1.2–1.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AIMS2, pain subscale</td>
<td>RA</td>
<td>$r_{VAS} = 0.66$</td>
<td>—</td>
<td>ICC 0.75–0.88</td>
<td>SRM 0.82–2.67</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SF-AIMS, pain subscale</td>
<td>RA</td>
<td>$r_{VAS} = 0.76$</td>
<td>0.84–0.91</td>
<td>$r = 0.19–0.38$</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SF-AIMS2, pain subscale</td>
<td>RA</td>
<td>$r_{VAS} = 0.59–0.60$</td>
<td>0.75–0.84</td>
<td>ICC 0.71–0.83</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DAREA, pain item</td>
<td>ReA</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>ES 2.72</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ADL, pain subscale</td>
<td>RA</td>
<td>$r_{VAS} = 0.54$</td>
<td>—</td>
<td>$r = 0.81$</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NHP, pain subscale</td>
<td>RA</td>
<td>$r_{VAS} = 0.62$</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Combination Measures (overall pain and anatomically-specific pain)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAPS</td>
<td>RA</td>
<td>$r_{VAS} = 0.68$</td>
<td>0.64–0.86</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>52</td>
</tr>
<tr>
<td>RADAI, pain items</td>
<td>RA</td>
<td>$r_{items} = 0.67$</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Anatomically-specific Pain Measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FFI, pain subscale</td>
<td>RA</td>
<td>—</td>
<td>0.95</td>
<td>ICC 0.7–0.89</td>
<td>—</td>
<td>—</td>
<td>54,55</td>
</tr>
<tr>
<td>MHQ, pain subscale</td>
<td>RA</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>SRM 0.15 ES 0.19</td>
<td>MCID</td>
<td>56,57</td>
</tr>
<tr>
<td>RAOS, pain subscale</td>
<td>RA</td>
<td>$r_{AIMS2(pain)} = 0.36$</td>
<td>0.90</td>
<td>ICC 0.88</td>
<td>—</td>
<td>ES –0.28</td>
<td>58</td>
</tr>
<tr>
<td>BASDAl, pain items</td>
<td>AS</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>SRM $\text{mean} = 0.94–1.24$</td>
<td>—</td>
<td>59–62</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SRM $\text{henn} = 0.55–1.18$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

† 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; BASDAl: Bath Ankylosing Spondylitis Disease Activity Index.
which is regarded as being satisfactorily consistent\textsuperscript{10}. Due to
the approach of calculating internal consistency, which
requires at least 2 items, there were no coefficients for sin-
gle-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 depend-
ing on the coefficient and the underlying study design that was
applied. While ICC showed satisfactory results of $\geq 0.7$
according to suggestions in the literature\textsuperscript{11}, the product-
moment correlation of 2 subsequent timepoints of pain meas-
urement ranged from $r = 0.15$ (short form of the AIMS;
SF-AIMS) to $r = 0.96$ (NRS). Unfortunately, for these consider-
erably 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 follow-
up assessment (i.e., $\geq 6$ months), whereas the 2 assess-
ments of the NRS had been done before and immediately after
a regular medical consultation. A similar pattern of hetero-
genility 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 $\geq 0.80$ 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 peri-
od and the pain scale that was applied. Table 2 shows corre-
sponding 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 (cli-
cially important change in RA, minimally important differ-
ence 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 litera-
ture on measuring pain and efficacy of pain treatment in IA.
Combined with expert opinion from a broad panel of rheuma-
tologists 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\textsuperscript{2}.

From the available data, we were able to demonstrate that
the VAS for overall pain intensity is currently the best-evalu-
ated 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 sub-
scale 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 meas-
ures applying various scaling techniques, the correlation co-
efficients obtained still appear to be of satisfactory value in
accordance with the requirements mentioned in the literature\textsuperscript{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 pro-
file 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. Addition-
ally, 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 respon-
siveness 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 relia-
bility 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 ques-
tions will have to be addressed in the future. Besides the task
of completing clinimetric profiles of various pain measure-
ment tools, the question of whether a meaningful difference
remains stable across the whole scale of a tool will be of spe-
cial interest in this context. For this SLR we decided to strict-
ly 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 contain-
ing a pain subscale was taken into further account only if dis-
tinct data on the pain subscales were presented. We are aware
that pain is also closely related to other domains of daily liv-
ing such as physical functioning or physical and mental well-
being and may interact with these variables. Moreover, infor-
mation 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 IA².

ACKNOWLEDGMENT

The authors would especially like to thank the following persons for their support and dedication in developing the search algorithm for this SLR: Dr. Louise Falzon, The Cochrane Collaboration, Australia, and Dr. Volker Müller and Dr. Sandra Heuser, University of Erlangen-Nuremberg, Erlangen, Germany.

REFERENCES


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


