Interpretation of the Juvenile Arthritis Disease Activity Score: responsiveness, clinically important differences and levels of disease activity in prospective cohorts of patients with juvenile idiopathic arthritis

Maja Bulatović Ćalasan1, Lara D. de Vries1, Sebastiaan J. Vastert1, Marloes W. Heijstek1 and Nico M. Wulffraat1

Abstract

Objectives. The objectives of this study were to assess 27-joint Juvenile Arthritis Disease Activity Score (JADAS-27) responsiveness, JADAS-27 changes corresponding to clinically important differences and cut-off scores for low and high disease activity in a large prospective JIA cohort.

Methods. JADAS-27 responsiveness, using effect size and standardized response mean (SRM), and changes in the JADAS-27 corresponding to clinically important differences were determined for clinical improvement (ACRpedi30) and worsening (flare). To assess whether various degrees of change in the JADAS-27 could be used to demonstrate improvement or worsening in individual patients, diagnostic parameters were computed for cut-off score changes. Finally, cut-off scores for low and high disease activity and their diagnostic parameters were determined.

Results. In 228 patients with 529 consecutive visits, ACRpedi30 was detected in 109 and flare in 111 visits. Regarding responsiveness, the effect size was 0.93 and SRM was 1.26 for clinical improvement, while for clinical worsening the effect size was 0.65 and SRM was 0.60. Changes in the JADAS-27 corresponding to clinically important difference were −5.5 for improvement and +1.7 for worsening. Cut-off score changes in the JADAS-27 had 65–90% sensitivity and 67–86% specificity for improvement, and 31–64% sensitivity and 89–97% specificity for worsening. The JADAS-27 cut-off score for low disease activity was ≤2.7 with 76% sensitivity and 62% specificity, and the cut-off score for high disease activity was ≥6 with 77% sensitivity and 77% specificity.

Conclusion. The JADAS-27 had moderate to good responsiveness and was changed by clinically important differences. The JADAS-27 cut-off scores differentiated between low and high disease activity. These JADAS-27 interpretations could be potentially applicable in clinical care and trials.

Key words: juvenile idiopathic arthritis, responsiveness, clinically important difference, disease activity.

Introduction

JIA is the most common chronic rheumatic disease in childhood [1]. JIA disease activity is variable and appropriate disease control is crucial to prevent irreversible joint destruction and long-term disabilities [2]. To assess and monitor the extent of disease control in JIA, a new composite disease activity score, the 27-joint Juvenile Arthritis Disease Activity Score (JADAS-27), was recently developed and validated [3]. In contrast to relative measures such as the ACR paediatric (ACRpedi) criteria, the JADAS-27 is an absolute disease activity measure that can be used to determine and evaluate disease activity status and course in individual patients. The ultimate goal is to use this composite score in daily clinical practice and clinical trials, as is done for the DAS...
in adult RA [4, 5]. In order to use the JADAS-27 in these clinical settings, it is crucial to determine its responsiveness, changes in JADAS-27 that correspond to clinically important differences and the cut-off scores that categorize patients into low and high disease activity. Although these JADAS-27 interpretations are essential for monitoring disease activity in individual patients over time and for comparison of disease activity status between patients, they have not been determined to date. The objectives of this study were to determine the JADAS-27 responsiveness, changes in score corresponding to clinically important difference and cut-off values for low and high disease activity in a large prospective JIA cohort.

Patients and methods

Patients

Clinical data on disease characteristics, disease activity and medication use of JIA patients was prospectively gathered every 3 months for 1 year between August 2007 and April 2011 for three investigator-initiated clinical trials concerning the safety and efficacy of vaccinations (NCT00731965, NCT00815282) and the occurrence of MTX intolerance (ISRCTN13524271). Each patient had up to five outpatient ward visits. The original clinical trials were performed at the University Medical Centre Utrecht (UMCU) and were approved by the Ethics Committee of the UMCU and the Central Committee on Research involving Human Subjects. JIA patients of six subtypes (persistent oligoarticular, extended oligoarticular, polyarticular, psoriatic, enthesitis-related and systemic JIA) with a confirmed diagnosis according to the ILAR criteria were included [6]. Patients having uveitis without joint involvement were excluded. Full ethics approval of the data analysis described in this article was retrospectively obtained from the Central Committee on Research involving Human Subjects.

JADAS-27 computation

The JADAS-27 (range 0–57) was computed by summing the scores of four core-set criteria [3]: physician’s global assessment of disease activity (PGA) on a 10 cm visual analogue scale (VAS); parent/patient global assessment of well-being on a 10 cm VAS [7]; active arthritis, defined as joint swelling or limitation of movement accompanied by pain and tenderness, assessed in 27 joints; and ESR (mm/h) normalized to a 0 by pain and tenderness, assessed in 27 joints; and ESR as joint swelling or limitation of movement accompanied of well-being on a 10 cm VAS [7]. The JADAS-27 computation was retrospectively obtained from the Central Committee of the UMCU and the Central Committee on Research involving Human Subjects. JIA patients of six subtypes (persistent oligoarticular, extended oligoarticular, polyarticular, psoriatic, enthesitis-related and systemic JIA) with a confirmed diagnosis according to the ILAR criteria were included [6]. Patients having uveitis without joint involvement were excluded. Full ethics approval of the data analysis described in this article was retrospectively obtained from the Central Committee on Research involving Human Subjects.

Responsiveness

Responsiveness represents the instrument’s capacity to detect a change in health status [8]. More specifically, responsiveness represents the capacity of the JADAS-27 to detect a change in disease status. The changes in disease status were classified into disease improvement, defined as an ACRp30 response or flare, for disease worsening, defined as a flare [10]. A flare is defined as worsening of \( >40\% \) in two or more core-set criteria (PGA, number of active joints, number of joints with limitation of movement, physical functional ability [measured with the Childhood Health Assessment Questionnaire (CHAQ)], parent/patient assessment of patient’s well-being and ESR) without an improvement of \( >30\% \) in two or more of the remaining core-set criteria. To determine the changes in the JADAS-27 and subsequently the changes in disease status, two consecutive visits during the follow-up were compared to each other (3-month visit was compared with the baseline, 6-month compared with 3-month, 9-month compared with 6-month and 12-month compared with 9-month).

To assess the JADAS-27 responsiveness, we computed the effect size and the standardized response mean (SRM) for disease improvement and disease worsening [11]. The effect size was calculated by dividing the mean change in the JADAS-27 between consecutive visits by the s.d. of JADAS-27 scores at baseline (baseline is defined as the visit which the following visit is compared with). The SRM was calculated by dividing the mean change in the JADAS-27 between consecutive visits with the s.d. of that change. The threshold levels for both the effect size and SRM are \( >0.20=\text{small}, >0.50=\text{moderate} \text{and } >0.80=\text{good} [11] \).

Clinical importance difference

A clinically important difference is a change in score of a construct that would be considered important (meaningful) from the perspective of a patient or a clinician [8, 12]. More specifically, a clinically important difference is a change in the JADAS-27 considered important from the perspective of a clinician. To determine clinically important differences, an anchor-based approach was used [8, 12]. Anchor-based methods require external patient-based or clinical criteria, so-called anchors, to inform whether changes in a construct are important (meaningful) [8]. The external criteria (anchors) used here were ACRp30, for disease improvement, and flare, for disease worsening. To determine whether disease improvement or worsening had occurred, two consecutive visits during the follow-up were compared.

First, clinically important differences were calculated as the median changes of the score for consecutive visits in which patients had satisfied the external criterion—ACRp30 or flare [8, 13]. We also tested whether the median changes in the JADAS-27, corresponding to clinically important difference, differed between JIA subtypes, using one-way analysis of variance with Bonferroni adjustment for multiple comparisons. Second, to determine how well calculated changes in the JADAS-27 discriminated between visits in which patients had an ACRp30 response or flare compared with visits in which they did not, receiver operating characteristic (ROC) curves and the corresponding areas under the ROC curves (AUC) were computed. Third, to examine
whether various degrees of change in the JADAS-27 could be used to demonstrate improvement or worsening in individual patients in daily clinical practice, we computed the sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) for several integer cut-off score changes. These diagnostic parameters provide answers to the following questions: How likely is it that the change in score (i.e. $\leq -2$) will detect improvement, which has clinically occurred? (sensitivity); How likely is it that the change in score (i.e. $> -2$) will detect no improvement, which has also occurred clinically? (specificity); Given that a patient has a change in score of e.g. $\leq -2$, indicating improvement, how likely is this to be clinically correct? (PPV); and Given that a patient has a change in score of e.g. $> -2$, indicating no improvement, how likely is this to be clinically correct? (NPV).

**Cut-off values for low and high disease activity**

The JADAS-27 was computed for visits classified into low and high disease activity. Low disease activity was defined as stopping MTX or biologics or having NSAID monotherapy or no medication. Visits in which MTX or biologics were stopped due to adverse effects were excluded. High disease activity was defined as starting MTX, biologics or oral steroids. The JADAS-27 cut-off values were determined using a method previously employed to determine cut-off values for the DAS28 in RA [14]. The cut-off for low disease activity was set at the 75th percentile of the JADAS-27 of the low disease activity group, whereas the cut-off for high disease activity was set at the 25th percentile of the JADAS-27 of the high disease activity group. Determining the cut-off values based on the above-mentioned percentiles pre-defines sensitivity at $\sim 75\%$. Subsequently the cut-off values’ specificity, PPV and NPV were determined. All statistical analyses were performed using IBM SPSS Statistics, version 20.

**Results**

Of 273 eligible patients, 1 patient with uveitis without joint involvement and 2 patients who failed to visit the outpatient ward were excluded. Two hundred and seventy patients with 1035 visits were included (1 visit, $n=16$ patients; 2 visits, $n=22$ patients; 3 visits, $n=39$ patients; 4 visits, $n=107$ patients; 5 visits, $n=86$ patients). The JADAS-27 could be calculated in 789 (76.2%) visits (Table 1). The baseline JADAS-27 distribution (median 4.0; range 0–40.5) was skewed to the left (skewness 1.8, kurtosis 4.0). The missing JADAS-27 was due to the absence of the following core-set criteria: active joint count ($n=17$), ESR ($n=93$), parent/patient global assessment of well-being ($n=53$) or a combination of these ($n=83$). There were no differences in the remaining core-set criteria between visits with a computable JADAS-27 and those with a missing JADAS-27.

**Responsiveness**

To determine responsiveness, the change in the JADAS-27 was calculated for 529 consecutive visits of 228 patients. Of 529 visits, an ACRpedi30 response was detected in 109 (21%) and flare in 111 (21%). For visits in which patients had an ACRpedi30 response, the effect size was 0.93 and SRM was 1.26, whereas for visits in which patients had a flare, the effect size was 0.65 and SRM was 0.60. Therefore the JADAS-27 revealed good responsiveness to change for clinical improvement and moderate responsiveness to change for clinical worsening.

**Clinically important difference**

The clinically important difference for disease improvement (ACRpedi30) was a median change in the JADAS-27 of $-5.5$ [interquartile range (IQR) of $-9.5$ to $-2.7$]. The clinically important difference for disease worsening (flare) was a median change in the JADAS-27 of $+1.7$ (IQR $+0.3$ to $+5.0$). For disease improvement, the change (decrease) in the JADAS-27 in polyarticular JIA patients was higher than in oligoarticular JIA patients by 2.5 (95% CI 1.1, 3.9; $P=0.001$). For disease worsening, the change (increase) in the JADAS-27 was higher in polyarticular JIA patients than in oligoarticular JIA patients by 2.5 (95% CI 1.1, 3.9; $P < 0.001$).

The AUC under the ROC curve of the above-mentioned changes in the JADAS-27 for clinical improvement was 0.86 (95% CI 0.83, 0.90), indicating that 86% of visits in which patients had an ACRpedi30 response were

**Table 1** Baseline characteristics of 270 JIA patients

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female, $n$ (%)</td>
<td>183 (67.8)</td>
</tr>
<tr>
<td>Age at inclusion, mean (S.D.), years</td>
<td>9.9 (4.2)</td>
</tr>
<tr>
<td>Age at onset, mean (S.D.), years</td>
<td>6.1 (4.4)</td>
</tr>
<tr>
<td>Subtype JIA, $n$ (%)</td>
<td></td>
</tr>
<tr>
<td>Persistent oligoarticular JIA</td>
<td>102 (37.8)</td>
</tr>
<tr>
<td>Extended oligoarticular JIA</td>
<td>30 (11.1)</td>
</tr>
<tr>
<td>Polyarticular JIA</td>
<td>92 (34.1)</td>
</tr>
<tr>
<td>Psoriatic arthritis</td>
<td>14 (5.2)</td>
</tr>
<tr>
<td>Enthesitis-related arthritis</td>
<td>8 (3.0)</td>
</tr>
<tr>
<td>Systemic onset JIA</td>
<td>24 (8.9)</td>
</tr>
<tr>
<td>Core-set criteria, median (range)</td>
<td></td>
</tr>
<tr>
<td>PGA (0–10)</td>
<td>1.4 (0–8.5)</td>
</tr>
<tr>
<td>Number of joints with limited range of motion</td>
<td>1 (0–25)</td>
</tr>
<tr>
<td>Number of joints with active arthritis</td>
<td>1 (0–27)</td>
</tr>
<tr>
<td>CHAQ disability (0–3)</td>
<td>0.5 (0–2.9)</td>
</tr>
<tr>
<td>Parent/patient global assessment of well-being (0–10)</td>
<td>2.0 (0–9.9)</td>
</tr>
<tr>
<td>ESR, mm/h</td>
<td>10 (2–140)</td>
</tr>
<tr>
<td>Medication, $n$ (%)</td>
<td></td>
</tr>
<tr>
<td>NSAIDs</td>
<td>165 (60.7)</td>
</tr>
<tr>
<td>MTX</td>
<td>164 (60.7)</td>
</tr>
<tr>
<td>Dosage, median (range), mg/m²/week</td>
<td>10.2 (5.1–28.2)</td>
</tr>
<tr>
<td>Biologics</td>
<td>28 (10.4)</td>
</tr>
<tr>
<td>Steroids oral</td>
<td>9 (3.3)</td>
</tr>
<tr>
<td>Dosage, median (range), mg/kg/day</td>
<td>3.8 (0.1–15.0)</td>
</tr>
<tr>
<td>Steroids local</td>
<td>14 (5.2)</td>
</tr>
<tr>
<td>JADAS-27, median (range)</td>
<td>4.0 (0–40.5)</td>
</tr>
</tbody>
</table>

PGA: physician’s global assessment of disease activity; CHAQ: Childhood Health Assessment Questionnaire; JADAS-27: Juvenile Arthritis Disease Activity Score involving 27 joints.

aRF positive, $n=14$ (5.2%); RF negative, $n=70$ (25.9%); RF unknown, $n=8$ (3.0%).
bEtanercept, $n=20$ (7.4%); anakinra, $n=7$ (2.6%); adalimumab, $n=1$ (0.4%).
classifying correctly. The AUC under the ROC curve of the above-mentioned changes in the JADAS-27 for clinical worsening was 0.84 (95% CI 0.80, 0.88), indicating that 84% of visits in which patients had a flare were classified correctly.

Table 2 presents the diagnostic parameters of various degrees of change in the JADAS-27 that could be used to monitor and follow individual patients in daily clinical practice. For clinical improvement, cut-off scores showed moderate to (very) good sensitivity of 65-90% (the cut-off score and score changes below it are likely to detect improvement), which has also occurred clinically) and moderate to (very) good specificity of 67-86% (score changes above the cut-off score are likely to detect no improvement, which has also occurred clinically). The best balance between sensitivity (80%) and specificity (78%) was reached at \( \leq -2 \). In case of clinical worsening, while cut-off scores had good to very good specificity of 89–97%, they showed relatively poor sensitivity of 31–64%. The best balance between sensitivity (64%) and specificity (89%) was reached at \( \geq 1 \). PPV, which reveals how likely it is that a patient is clinically improved or worsened if he or she has a certain change in the JADAS-27, is an important parameter since it could be used by clinicians to interpret the score changes in individual patients and to follow them over time. While for clinical worsening the PPV was moderate (62–71%), for clinical improvement the PPV was relatively low (42–54%). On the other hand, the NPV was high for both clinical improvement and worsening (84–96%).

### Cut-off values for low and high disease activity

In 316 (30.5%) visits, patients had low disease activity with a median JADAS-27 of 0.5 (IQR 0.0–2.7), whereas in 190 (18.4%) visits, patients had high disease activity with a median JADAS-27 of 11.2 (IQR 6.0–17.6) \( (P < 0.001) \) (Fig 1).

Two cut-off values were selected: at the 75th percentile for low disease activity (JADAS-27 \( \leq 2.7 \)) and at the 25th percentile for high disease activity (JADAS-27 \( \geq 6.0 \)). At the cut-off \( \leq 2.7 \), sensitivity was 76% and specificity was 62%; in 42% of visits with a JADAS-27 \( \leq 2.7 \) (PPV), patients were identified correctly as having low disease activity and in 88% of visits with a JADAS-27 \( > 2.7 \) (NPV), patients were identified correctly as having non-low disease activity. At the cut-off \( \geq 6 \), sensitivity and specificity were 77%, with a PPV of 41% and an NPV of 94%.

### Discussion

We showed that the JADAS-27 has moderate to good responsiveness to changes in disease activity status. Changes in the JADAS-27 score corresponding to clinically important differences are \( \pm 5.5 \) for disease improvement and \( \pm 1.7 \) for disease worsening. Moreover, the JADAS-27 cut-off score is \( \leq 2.7 \) for low disease activity and \( \geq 6 \) for high disease activity.

The JADAS-27 responsiveness, determined using two measures, i.e. effect size and SRM, was moderate for disease worsening (\( \pm 0.50 \)) and good for disease improvement (\( \pm 0.80 \)). Similarly, others have demonstrated good responsiveness of the JADAS-27 in two longitudinal cohorts, with SRM values \( \geq 0.8 \) (1.27 and 0.98) [3]. Conversely, a clinically important difference for the
JADAS-27 has not been previously demonstrated. In our cohort, clinically important differences were a median decrease in score of 5.5 for disease improvement and a median increase in score of 1.7 for disease worsening. Moreover, changes in scores for improvement and worsening were able to discriminate well between visits in which patients had a change in disease activity and visits in which they did not have a change in disease activity (AUCs 0.86 and 0.84, respectively).

In order to use the changes in score in daily clinical practice, clinicians should know how likely it is that patients improved or worsened if they had a certain change in the JADAS-27. Thus diagnostic parameters for various cut-off score changes were computed for disease improvement and worsening (Table 2). While all cut-off scores showed moderate to good sensitivity and specificity for disease improvement, they had relatively low PPVs (54%), since 54% of visits in which patients had a given JADAS-27 change were not accompanied by ACRpedi30 improvement. Low PPVs could impede the use of cut-off scores for disease improvement in daily clinical practice, as clinicians would not be able to establish with great certainty that a patient with a particular change in score is indeed clinically improved (has reached ACRpedi30) or worsened (had a flare). On the other hand, the cut-off scores for clinical worsening had moderately good PPVs, but relatively low sensitivities, which could be due to an insufficient ability of the JADAS-27 to detect disease worsening or to the definition for disease activity worsening (flare). A flare occurred in an unexpectedly large proportion of visits (21%), suggestive of a lenient definition of flare. Indeed, if a stricter definition of worsening was used, namely a flare with an increase of at least 20% in PGA and 15% in ESR, only 3% of visits fulfilled the requirements for disease activity worsening. Contrary to PPVs, the NPVs for both clinical improvement and worsening were high, indicating that clinicians would be able to establish with great certainty that patients without a particular change in score are indeed clinically not improved or not worsened.

It is noteworthy that clinically important difference is not termed minimal clinically important difference for the following reason. To inform whether changes in the JADAS-27 were clinically important, external clinical criteria (anchors) of ACRpedi30 and flare were used. Whether the change in the JADAS-27 is minimal depends on the anchor used. Although ACRpedi30 and flare are important from the prospective of a clinician, as they could provoke changes in the therapeutic approach, it is nevertheless possible that ACRpedi30 and flare are more than minimally important from the clinician’s prospective. In order to establish minimal clinically important difference, multiple anchors should be used, e.g. patient’s or parent’s opinion on the extent of change in disease activity (small vs moderate vs large). Furthermore, as we were unable to calculate the S.E.M. [12] due to a lack of more frequent (i.e. monthly) visits to measure the JADAS-27, we cannot exclude that JADAS-27 changes corresponding to the clinically important difference could be the result of a measurement error rather than true observed changes. However, keeping in mind that changes in the JADAS-27 were able to discriminate well between patients with and without disease improvement and worsening, it is unlikely that the changes in score are the result of a measurement error.

The JADAS-27 is able to discriminate between patients with low and high disease activity. The devised cut-off scores for low and high disease activity can be used to interpret disease activity status and to compare disease activity status between individual patients and patient groups. Recently Consolaro et al. [15] determined cut-off values for minimal, acceptable and inactive disease. Their cut-off for minimal disease activity (2.0–3.8) and inactive disease (1) corresponds to our cut-off for low disease activity of ≤2.7 [15]. In addition to the above-mentioned study, we also computed the JADAS-27 cut-off for high disease activity. Although the sensitivity and specificity of the proposed cut-off values for disease activity were satisfactory, their PPVs were relatively low: the PPV of 42% for low disease activity indicates that in 58% of visits in which patients had a JADAS-27 ≤2.7, disease activity was non-low. Similarly, in 59% of visits with a JADAS-27 ≥6, disease activity was non-high. On the contrary, the NPVs for low and high disease activity reached 88% and 94%, respectively. In order to use the cut-off values in daily clinical practice, optimization of diagnostic parameters, and PPV in particular, is warranted. This could be achieved by determining cut-off values for low and high disease activity, defined using a different external criterion, such as parent or patient assessment of disease status, as has been previously done for cut-off values with minimal, acceptable and inactive disease [15].

In conclusion, we determined the responsiveness of the JADAS-27, changes in the JADAS-27 corresponding to a clinically important difference for disease improvement and worsening and cut-off scores for low and high disease activity in a large prospective JIA cohort. If these results are refined and confirmed using different external criteria to define disease activity changes and states as well as validated in an independent JIA cohort, the above mentioned JADAS-27 interpretations could be potentially applicable in clinical practice and trials for monitoring and comparison of disease activity (changes) in and between individual patients.

Rheumatology key messages

- The JADAS-27 is responsive to change and can be changed by clinically important differences.
- The JADAS-27 differentiates between JIA patients with low and high disease activity.
- These JADAS-27 interpretations could be potentially applicable in clinical practice and trials to assess and monitor (changes in) disease activity.

Disclosure statement: The authors have declared no conflicts of interest.
References

8 Revicki DA, Cella D, Hays RD et al. Responsiveness and minimal important differences for patient reported outcomes. Health Qual Life Outcomes 2006;4:70.