



HHS Public Access

Author manuscript

Arthritis Care Res (Hoboken). Author manuscript; available in PMC 2024 February 01.

Published in final edited form as:

Arthritis Care Res (Hoboken). 2023 February ; 75(2): 373–380. doi:10.1002/acr.24763.

Development and validation of the juvenile spondyloarthritis disease flare (JSpAflare) measure: Ascertaining flare in patients with inactive disease

PF Weiss, MD, MSCE,

Departments of Pediatrics and Epidemiology, Perelman School of Medicine; Division of Rheumatology, Children's Hospital of Philadelphia; Center for Pediatric Clinical Effectiveness, Children's Hospital of Philadelphia, Philadelphia, PA

TG Brandon, MPH,

Department of Pediatrics, Division of Rheumatology, Children's Hospital of Philadelphia; Center for Pediatric Clinical Effectiveness, Children's Hospital of Philadelphia, Philadelphia, PA

ME Ryan, DO,

Division of Rheumatology, University of Minnesota Masonic Children's Hospital, Minneapolis, MN

EB Treemarcki, DO,

Department of Pediatrics, Division of Rheumatology, University of Utah and Primary Children's Hospital, Salt Lake City, UT

S Armendariz, APRN, FNP-C,

Division of Rheumatology, Scottish Rite Hospital for Children, Dallas, TX

TB Wright, MD,

Department of Pediatrics Division of Rheumatology, UT Southwestern Medical Center; Scottish Rite Hospital for Children, Dallas, TX

C Godiwala, MS,

Division of Rheumatology, Scottish Rite Hospital for Children, Dallas, TX

ML Stoll, MD, PhD, MSCS,

Department of Pediatrics, Division of Pediatric Rheumatology, University of Alabama at Birmingham, Birmingham, Alabama

R Xiao, PhD,

Department of Biostatistics, Epidemiology and informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

D Lovell, MD

Department of Pediatrics, Division of Pediatric Rheumatology, Cincinnati Children's Hospital Medical Center

Corresponding Author: Pamela F. Weiss, MD, MSCE, The Children's Hospital of Philadelphia, Roberts Center for Pediatric Research, 2716 South Street, Room 11121, Philadelphia, PA 19146. weisspa@chop.edu, tel: 215-590-2547, fax: 267-426-0380.

Disclosures: Dr. Weiss has served as a consultant for Lilly and Pfizer.

Abstract

Objective: Our objective was to develop and validate a composite disease flare definition for juvenile spondyloarthritis that would closely approximate the clinical decision made to reinitiate/not reinitiate systemic therapy after therapy de-escalation.

Methods: Retrospective chart reviews of children with spondyloarthritis who underwent systemic therapy de-escalation of biologic or conventional disease-modifying antirheumatic drugs (bDMARDs; cDMARDs) were used to develop and validate the flare outcome. Independent cohorts for development (1 center) and validation (4 centers) were collected from large tertiary healthcare systems. Core measure thresholds and candidate disease flare outcomes were assessed using sensitivity, specificity, positive (PPV) and negative predictive values (NPV), and area under the receiver operating characteristic (AUROC) curve with physician assessment of “active disease” *plus* re-initiation of standard dose of systemic therapy as the reference standard.

Results: Of the candidate definitions, clinically meaningful worsening in 3 of the following five core measures performed best: caregiver/patient assessment of well-being, physician assessment of disease activity, caregiver/patient assessment of pain, physical function, and active joint count. AUROC was 0.91, PPV 87.5%, NPV 98.1%, sensitivity 82.4%, and specificity 98.7%. Cronbach’s α was 0.81, signifying internal consistency and factor analysis demonstrated the outcome measured one construct. “JSpAflare” had face validity according to 21 surveyed pediatric rheumatologists. JSpAflare had AUROC 0.85, PPV 92.3%, and NPV 96.8% in the validation cohort.

Conclusions: There is initial support for the validity of JSpAflare as a tool to identify disease flare in juvenile spondyloarthritis patients de-escalating therapy and is potentially applicable in clinical practice, observational studies, and therapeutic trials.

Clinical trials in juvenile spondyloarthritis are greatly needed both for efficacy of novel targeted agents as well as for therapy de-escalation strategies in those who have achieved sustained remission. Since the introduction of biologic disease modifying anti-rheumatic drugs (bDMARDs) such as tumor necrosis factor inhibitors (TNFi), inactive disease is a realistic goal. In 2018, an international task force of pediatric rheumatologists developed recommendations for treating juvenile arthritis, including spondyloarthritis, to target.(1) The primary treatment target for juvenile arthritis was inactive disease, defined as the absence of all clinical signs and patient-experienced symptoms of inflammatory disease activity. Current treatment approaches for children with spondyloarthritis have resulted in up to 60% attaining inactive disease while on therapy.(2–4)

The international pediatric task force specified several overarching principles for the management of juvenile arthritis, which included not only controlling signs and symptoms of disease but also avoidance of drug toxicities and optimization of personal well-being. (1) The use of injectable biologics can affect quality of life, cause anxiety, and create a sense of “being different” in children.(5) Recently, the COVID-19 pandemic has prompted numerous inquiries from families and patients regarding whether being on a conventional or biologic disease modifying antirheumatic agent (cDMARD; bDMARD) increased the risk of infection with SARS-CoV-2, and, if infected, whether being on a cDMARD or bDMARD increased the risk of having more severe disease. In the absence of definitive answers, many

families' next question was, "Can we stop the medication?" There is limited to no data to guide therapy de-escalation in juvenile SpA. In order to study the risk of flare after therapy de-escalation, a validated composite measure of flare that closely approximates the clinical decision made in routine care to reinitiate or not reinitiate systemic therapy after therapy de-escalation is greatly needed.

There are no validated definitions of flare in juvenile spondyloarthritis though juvenile idiopathic arthritis (JIA) flare measures have been developed and used in randomized withdrawal trials and biologic de-escalation trials in polyarticular JIA.(6–8) The validity of those metrics has not been examined in juvenile spondyloarthritis, which encompasses different disease manifestations that are important to consider. The existing flare definitions, (6–8) based on the American College of Rheumatology (ACR) pediatric six core response variables,(9) are heavily dependent on the presence of peripheral joint disease with separate measures for the active joint and limited range of motion counts. A recent TNFi withdrawal trial of polyarticular patients in remission used a modified ACR core set definition of flare. In addition to meeting the 30% worsening from baseline in three cores, the change from baseline had to exceed the clinically important change in each core (e.g. an increase in the active or limited range of motion joint count by at least 2). (7) The potential application of this flare definition to the spondyloarthritis population is problematic as only one-third of children have polyarticular disease, and early disease flare in spondyloarthritis can also manifest as recurrence of enthesitis and/or axial symptoms, neither of which are encompassed with the juvenile arthritis flare criteria. In 2014, a juvenile spondyloarthritis disease activity (JSpADA) index was developed using modified Delphi consensus techniques with input from 106 physicians from 14 countries.(10) The JSpADA index contains eight items all scored 0-1 for a total score of 8. JSpADA items include active joint and enthesitis counts, patient pain scores, inflammatory markers (erythrocyte sedimentation rate or C-reactive protein), morning stiffness, clinical sacroiliitis, uveitis, and back mobility. A 7-item version of the JSpADA that excluded the measure of back mobility was also prospectively validated.(11)

We aimed to leverage the work done to develop the JIA flare criteria used in polyarticular JIA (7, 9) and the JSpADA index (10) to develop and validate a composite flare outcome for patients with juvenile spondyloarthritis called the juvenile spondyloarthritis flare (JSpAflare). JSpAflare was designed to closely mirror the disease activity threshold that triggers reinitiation of systemic therapy after therapy de-escalation in children with spondyloarthritis in routine care.

PATIENTS AND METHODS

Human Subjects protections.

The protocol for this retrospective study was reviewed and approved by the Institutional Review Boards at Children's Hospital of Philadelphia, the University of Utah and Primary Children's Hospital, University of Texas Southwestern, University of Alabama at Birmingham, and University of Minnesota.

Patients.

The source population for the development cohort was a retrospective longitudinal data set of children with spondyloarthritis who were evaluated at a large tertiary care center in a rheumatology clinic. Patients who met the following criteria were eligible for inclusion: fulfilled International League of Associations for Rheumatology enthesitis related arthritis criteria, treatment de-escalation of a cDMARD or bDMARD was initiated secondary to inactive disease and outcome data available for the reference visit and at least 1 follow-up visit. Children with a history of inflammatory bowel disease, psoriasis, uveitis, or amplified pain were excluded as these comorbidities may confound treatment recommendations. The source population for the validation cohort was an independent data set from four large tertiary care centers and the inclusion and exclusion criteria were the same as for the development cohort.

Data.

Data was abstracted from the electronic medical record or paper charts. Patient reported outcomes at 1 site (Utah) were obtained from the Childhood and Arthritis Research Alliance (CARRA) Registry. The following data elements were abstracted from each eligible clinic: demographics, clinical exam features (active joint count, tender enthesitis count), caregiver/patient assessment of well-being, physician assessment of disease activity, caregiver/patient assessment of pain, physical function (Child Health Assessment Questionnaire [CHAQ](12), NIH Patient Reported Outcomes Measurement Information System [PROMIS] mobility(13) or upper extremity function(13) short forms, or 4-question function form modeled to be a streamlined version of the CHAQ), medication use, and physician's impression of disease activity (inactive, active, or uncertain).

Active joint count was defined as the number of joints with swelling or, in the absence of swelling, limitation of motion accompanied by pain or warmth. The caregiver/patient assessment of well-being was scored 0-10 with anchors "Very well" and "Very poor" with higher scores indicating poorer well-being. The physician assessment of disease activity was scored 0-10 with anchors "Not active" and "Very active" with higher scores indicating higher magnitude of disease activity. The caregiver/patient assessment of pain was scored 0-10 with anchors "No pain" and "Very severe pain" with higher scores indicating higher magnitude of pain. All visual analogue scale measures were scored on a traditional 10-centimeter line, an integer-based scale, or an electronic sliding scale depending on the way these measures were collected as part of routine care in the clinic. Function was assessed by the NIH PROMIS mobility or upper extremity function short forms, the CHAQ, or the 4-question function-related portion of a clinic intake form. The PROMIS short forms have been validated in children with JIA(13) and each include 8 questions. A T-score of '50' represents the healthy population mean score with standard deviation equal to 10. The minimal clinically important difference (MCID) for the pediatric PROMIS measures is approximately 3.(14) The CHAQ is a validated pediatric measure and includes 30 questions covering 8 functional ability domains with scores ranging from 0 to 3 with higher scores indicating more functional impairment(15). The MCID for the CHAQ is 0.125.(16) The clinic function form at 1 site was a streamlined functional questionnaire consisting of 4 questions – 1) limitations in rigorous/athletic activities due to arthritis 2) limitations in

normal daily activities due to arthritis, 3) required assistance for others for normal activities, 4) use of aids or devices for normal activities. A positive response on any of the 4 questions was determined to be equivalent to an increase of at least the MCID seen in the CHAQ (0.125).

The reference visit was defined as the visit at which inactive disease was identified by the treating provider (or there was a physician global assessment of disease activity of 0 if disease activity was not explicitly assessed as “active” or “inactive”) *and* a bDMARD or cDMARD medication was withdrawn or tapered (either dosing interval increased or dose decreased). In the event a patient self-discontinued bDMARD or cDMARD due to inactive disease between clinical assessments, the next visit was used as the index visit if the disease was assessed as “inactive” or the physical global assessment was 0. In the event core measures were missing at the first inactive disease visit, the second visit was used as the reference visit (N=4, validation cohort) if the disease was assessed as “inactive” or the physical global assessment was 0. Data was collected for all visits until bDMARD or cDMARD was restarted, standard dose or dosing interval was restarted or the end of follow-up. Only visits with complete core flare variable data or for which flare or no flare could be definitively concluded were included. If subjects underwent more than 1 episode of therapy de-escalation for inactive disease, all episodes were included.

Developing Candidate Flare Definitions.

The preliminary core set of measures evaluated for JSpAflare included six core measures: caregiver/patient assessment of well-being, physician assessment of disease activity, caregiver/patient assessment of pain, physical function, active joint count, and tender entheses count.

Initial candidate definitions for disease flare were based upon published values of minimally important change in disease measures.^(9, 14, 20) In order to verify the optimal cut-off for change in each of the 6 core measures in the development cohort, we calculated the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and area under the receiver operating characteristic (AUROC) curve for each of the core variables. Various thresholds for absolute change from the reference time point (date of inactive disease and clinical decision to start therapy de-escalation) were tested using physician assessment of active disease *plus* re-initiation of systemic therapy as the reference standard of flare. Maximization of the AUROC was prioritized among the metrics for the choice of optimal threshold.

Next, using physician assessment of “Active disease” *plus* re-initiation of systemic therapy as the reference standard of flare, we assessed the measurement characteristics of the composite candidate flare definitions by calculating the sensitivity, specificity, PPV, and NPV to detect flare. We also plotted receiver operator characteristic curves and calculated AUROC. An AUROC >0.5 is typically considered responsive to change.⁽⁹⁾ From the pool of generated candidate flare definitions, the final definition was selected according to the AUROC, PPV, and NPV, in order of importance.

Internal consistency of the measure was assessed using Cronbach's α coefficient. Estimates less than 0.6 were considered as poor, 0.6-0.64 for slight, 0.65-0.69 for fair, 0.7-0.79 for moderate, 0.8-0.89 for substantial, and greater than 0.9 for almost perfect consistency.(21) Principal component analysis (PCA) was done to confirm the dimensional structure of the candidate items measured a single construct. This analysis generates loadings that indicate the strength of association of the core measures with its latent factor(s).

Validation.

The flare definitions derived from the development cohort were validated using an independent dataset of children with spondyloarthritis who underwent cDMARD or bDMARD therapy de-escalation at four large tertiary care centers. We compared the development and validation cohorts using two-sample t-test for continuous variables and Chi-square test for categorical variables. Face validity of the final flare definition was assessed through a REDCap survey to rheumatologists at 21 centers across the US. These rheumatologists were chosen because they all have established expertise in JIA clinical research and their centers expressed interest in participation in a therapy de-escalation trial for spondyloarthritis. Each rheumatologist was provided with the list of core measures, range of possible scores and clinically meaningful change for each. Respondents were asked to indicate "do you think this outcome for juvenile spondyloarthritis has face validity to capture flare". If the respondent answered "no", then s/he was subsequently asked to share why not.

We assessed discrimination, the ability to correctly classify patients with or without flare, in the validation cohort using the model derived from the development cohort. Specifically, we calculated AUROC, specificity, sensitivity, PPV, and NPV using physician assessment of "active disease" *plus* re-initiation of systemic therapy as the reference standard.

RESULTS

Patients.

The demographics and core measures at the reference time point for both the development and validation cohorts are shown in Table 1. Thirty-nine patients, 45 systemic therapy de-escalation episodes and 180 visits met inclusion criteria for the flare development cohort. Sixty-four percent were male, and median age at therapy de-escalation was 14.1 years (IQR 11.9-17.3). In the validation cohort, there were 36 patients, 39 systemic therapy de-escalation episodes, and 164 visits. Sixty-four percent of the validation cohort was male and the median age at the index visit was 14.3 years (IQR 12-15.8). The median patient follow-up time was 18.7 months (IQR 8.5-30.4) in the development cohort and 14.2 months (IQR 9.6-23.4) in the validation cohort. As defined by the reference standard, there were 17 episodes of flare in the development cohort and 10 in the validation cohort. As expected, the mean and standard deviation (SD) for the core measures indicated minimal clinical signs and patient-experienced symptoms of disease activity in both the development and validation cohorts. Missingness of core measures in the development cohort ranged from 0 to 7.7% and from 0-24.4% in the validation cohort. Patient-reported physical function outcomes had the highest percentage of missingness in both cohorts.

Absolute change thresholds for core measure.

Absolute change was tested for each core measure and not percentage change, given the minimal values at the reference visit. The data, as measured by AUROC, PPV, NPV, sensitivity, and specificity, supported using a meaningful absolute change of 2 for all visual analogue scale measures (caregiver/patient assessment of well-being, physician assessment of disease activity, and caregiver/patient assessment of pain). Physical function outcome data supported setting a cutoff of 3 unit change in the PROMIS T-scores, 0.125 unit change in the CHAQ, and 1 item in the clinic intake questionnaire, 1 unit change in active joint count, and 2 unit change in the tender enthesis count (see supplemental table).

Flare definition based on absolute change of core measures.

The test properties of the candidate flare definitions are shown in Table 2. Flare definitions based on meaningful change in the physician disease activity assessment plus meaningful change in 2 or 3 of the other core measures did not perform better than those including all the core measures. Flare definitions inclusive of meaningful change in tender enthesis count (6 core measures) did not perform better than those excluding the tender enthesis count (5 core measures). Based on high AUROC, PPV, and NPV, the best candidate definition for “JSpAflare” included five core measures and defined flare as meaningful change in 3 or more core measures (Table 2; Figure 1). AUROC was 0.91, PPV was 87.5, and NPV was 98.1. The frequency of meaningful change in each of the core measures in children who met the final JSpAflare criteria were as follows: caregiver/patient global assessment of well-being (75%), physician assessment of disease activity (81%), caregiver/patient assessment of pain (88%), physical function (80%), and active joint count (75%)(Table 3). Of note, only 13% of children had meaningful change in the tender enthesis count.

The Cronbach’s α coefficient in the development cohort was 0.81, demonstrating substantial internal consistency. Factor analysis demonstrated that the JSpAflare measured only one construct. Correlation between individual items and the latent factor was greatest for caregiver/patient assessment of pain (0.43), caregiver/patient global assessment of well-being (0.43), physician disease activity assessment (0.51), active joint count (0.48), and patient-reported function (0.36). This one key factor explained 53% of the variance.

Validation.

Validation of JSpAflare was performed on an independent data set of children with spondyloarthritis who underwent cDMARD or bDMARD therapy de-escalation at four additional large tertiary care centers. Face validity of the final flare definition was verified through a REDCap survey of pediatric rheumatologists at 21 centers across the US. Twenty-one of 21 rheumatologists (100%) agreed that the 5 core measures and thresholds for clinically meaningful change had face validity to capture flare in juvenile spondyloarthritis.

We assessed the ability of the top performing JSpAflare model from the development cohort to discriminate between patients with or without flare in the validation cohort using the same test statistics and reference standards (Table 2). The AUROC (0.85, 95% CI: 0.74-0.96), PPV (92.3, 95% CI: 64-99.8), NPV (96.8, 95% CI: 92.8-99), sensitivity (70.6, 95% CI: 44-89.7), and specificity (99.4, 95% CI: 96.4-100) were all high.

DISCUSSION

We describe the development and validation of a composite measure of spondyloarthritis disease flare, the JSpAflare, which includes 5 core measures evaluating overall disease activity, well-being, pain, function, and peripheral disease activity. A flare definition based on meaningful change in the physician disease activity assessment plus meaningful change in 2 or 3 of the other core measures did not perform better than those including all the core measures. Flare definitions inclusive of meaningful change in tender entheses count did not perform better than those excluding the tender entheses count; further, there was a low frequency of meaningful change in the tender entheses count in the development cohort. The final JSpAflare definition harmonizes 4 measures that are part of the pediatric ACR core response(9) and 2 measures included in the JSpADA index.(10) The JSpAflare is based upon absolute change in 3 or more core measures from the reference visit and is straightforward to determine. The AUROC and PPV were very high. The JSpAflare demonstrated substantial internal consistency and factor analysis revealed the core variable measured one construct. Validation was performed on an independent cohort from 4 large tertiary care referral centers, which likely represent the spectrum of children with spondyloarthritis seen by pediatric rheumatologists across the United States.

Our aim was to develop an outcome that closely approximates the point-of-care decision to re-initiate systemic therapy after cDMARD or bDMARD de-escalation that can be used in pragmatic trials of therapy de-escalation. While physician global assessment is often considered a gold standard measure of disease activity, it is a metric that is subjective and varies considerably amongst providers and institutions.(22) Therefore, while we felt it was important to include, we did not want the outcome of disease flare to rely solely upon this metric with poor interrater reliability.(22) The JSpAflare core measures are a mix of physician and patient-reported outcomes, which is critical given that only patients can report on domains like well-being, pain, and function, and physician and patient-reported disease assessments are often not correlated.(23) Additionally, JSpAflare includes an enhanced functional assessment over prior metrics. The CHAQ is the functional assessment traditionally included in tools such as the juvenile arthritis flare criteria. Prior work has demonstrated the MCID for worsening of the CHAQ approximates the smallest possible change in the CHAQ and that measure is relatively non-responsive to both short-term and clinically meaningful change in juvenile arthritis activity.(16) Conversely, the PROMIS functional measures of mobility and upper extremity function, which are included along with the CHAQ in the JSpAflare, do discriminate between meaningfully different disease states in juvenile arthritis.(24, 25)

There are several caveats of the study that should be kept in mind while interpreting the results. First, this was a retrospective study and, as expected, there was missing data. However, the missing data was minimal in the final analysis cohorts. Second, the flare definition was developed and validated in children with inactive disease undergoing open-label use and de-escalation of cDMARDs and bDMARDs. Therefore, the flare definition may not be valid for a randomized withdrawal trial where at the time of randomization children are typically not in inactive disease. Future work should include a post-hoc analysis of such a trial. Next, since the data collection and exam were not protocolized, there

may have been differences in the number of joints and entheses examined. However, this variability as part of routine care is also something we aimed to capture with this measure, as the hope is to use this measure in pragmatic trials that may not specify which peripheral joints to examine. Additionally, the peripheral joint exam is not protocolized in other well-accepted tools such as the clinical juvenile disease activity score (cJADAS) (26) and the JSpADA (10). Lastly, our sample size for the development and validation cohorts is small, but this is offset to some extent by having a robust number of clinical visits in each cohort for use in the analysis. However, despite our sample size we were able to demonstrate excellent performance of the JSpAflare across a cohort from 4 institutions. Further, the size of the development cohort is comparable to that used to develop the juvenile arthritis flare criteria.(9)

In summary, we have developed and validated a new composite flare outcome, JSpAflare, for spondyloarthritis that comprises 5 clinically relevant core measures for this condition. This tool is feasible and easily applicable at point-of-care for children who have inactive disease and have started to de-escalate therapy. In validation analyses, the JSpAflare had excellent measurement properties, indicating that it is potentially applicable in clinical practice, observational studies, and therapeutic trials. Future studies should test the validity of the JSpAflare in a prospective cohort of patients and an ad-hoc analysis of a randomized withdrawal phase 3 clinical trial including spondyloarthritis patients. In accordance with recent recommendations to not only control signs and symptoms of disease but also to avoid drug toxicities and optimize personal well-being,(1) the development of the JSpAflare makes evaluation of cDMARD and bDMARD therapy de-escalation feasible for children in sustained remission.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENT

We would like to thank the following individuals who helped with various stages of identification, abstraction, and transfer of subject data: Colleen Correll (University of Minnesota), Livie Huie (University of Alabama at Birmingham), and Suzy Richins (University of Utah). This work could not have been accomplished without the aid of the following organizations: The NIH's National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) & the Arthritis Foundation. We would also like to thank all participants and hospital sites that recruited patients for the CARRA Registry. The authors thank the following CARRA Registry site principal investigators, sub-investigators and research coordinators: John Bohnsack, Cassandra Davis, Deborah Durkee, Sylvie Fadrhonic, Aimee Hersh, Karen James, Rebecca Overbury, Clare Peckenpaugh, and Sara Stern.

Funding:

Dr. Weiss is funded by NIH R01 AR074098

REFERENCES

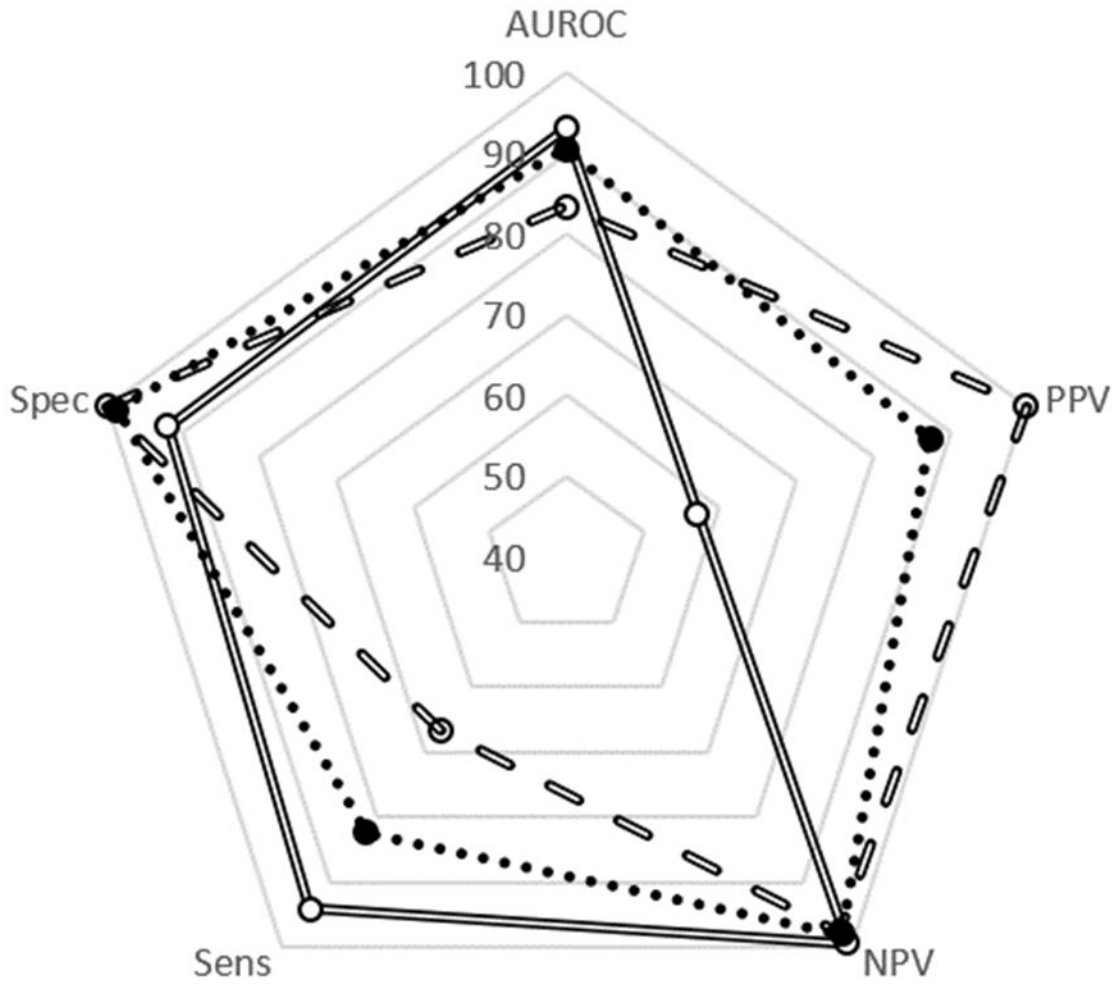
1. Ravelli A, Consolaro A, Horneff G, Laxer RM, Lovell DJ, Wulffraat NM, et al. Treating juvenile idiopathic arthritis to target: recommendations of an international task force. *Ann Rheum Dis*. 2018;77(6):819–28. [PubMed: 29643108]

2. Constantin T, Foeldvari I, Vojinovic J, Horneff G, Burgos-Vargas R, Nikishina I, et al. Two-year Efficacy and Safety of Etanercept in Pediatric Patients with Extended Oligoarthritis, Enthesitis-related Arthritis, or Psoriatic Arthritis. *J Rheumatol*. 2016;43(4):816–24. [PubMed: 26932344]
3. Burgos-Vargas R, Tse SM, Horneff G, Pangan AL, Kalabic J, Goss S, et al. A Randomized, Double-Blind, Placebo-Controlled Multicenter Study of Adalimumab in Pediatric Patients With Enthesitis-Related Arthritis. *Arthritis Care Res (Hoboken)*. 2015;67(11):1503–12. [PubMed: 26223543]
4. Guzman J, Henrey A, Loughin T, Berard RA, Schiff NJ, Jurecak R, et al. Predicting Which Children with Juvenile Idiopathic Arthritis Will Not Attain Early Remission with Conventional Treatment: Results from the ReACCh-Out Cohort. *J Rheumatol*. 2019.
5. Livermore P, Eleftheriou D, Wedderburn LR. The lived experience of juvenile idiopathic arthritis in young people receiving etanercept. *Pediatr Rheumatol Online J*. 2016;14(1):21. [PubMed: 27068023]
6. Brunner HI, Ruperto N, Tzaribachev N, Horneff G, Chasnyk VG, Panaviene V, et al. Subcutaneous golimumab for children with active polyarticular-course juvenile idiopathic arthritis: results of a multicentre, double-blind, randomised-withdrawal trial. *Ann Rheum Dis*. 2018;77(1):21–9. [PubMed: 28507219]
7. Lovell DJ, Johnson AL, Huang B, Gottlieb BS, Morris PW, Kimura Y, et al. Risk, Timing, and Predictors of Disease Flare After Discontinuation of Anti-Tumor Necrosis Factor Therapy in Children With Polyarticular Forms of Juvenile Idiopathic Arthritis With Clinically Inactive Disease. *Arthritis Rheumatol*. 2018;70(9):1508–18. [PubMed: 29604189]
8. Brunner HI, Ruperto N, Zuber Z, Keane C, Harari O, Kenwright A, et al. Efficacy and safety of tocilizumab in patients with polyarticular-course juvenile idiopathic arthritis: results from a phase 3, randomised, double-blind withdrawal trial. *Ann Rheum Dis*. 2015;74(6):1110–7. [PubMed: 24834925]
9. Brunner HI, Lovell DJ, Finck BK, Giannini EH. Preliminary definition of disease flare in juvenile rheumatoid arthritis. *J Rheumatol*. 2002;29(5):1058–64. [PubMed: 12022323]
10. Weiss PF, Colbert RA, Xiao R, Feudtner C, Beukelman T, DeWitt EM, et al. Development and retrospective validation of the juvenile spondyloarthritis disease activity index. *Arthritis Care Res (Hoboken)*. 2014;66(12):1775–82. [PubMed: 25047959]
11. Zanwar A, Phatak S, Aggarwal A. Prospective validation of the Juvenile Spondyloarthritis Disease Activity Index in children with enthesitis-related arthritis. *Rheumatology (Oxford)*. 2018;57(12):2167–71. [PubMed: 30107576]
12. Ruperto N, Ravelli A, Pistorio A, Malattia C, Cavuto S, Gado-West L, et al. Cross-cultural adaptation and psychometric evaluation of the Childhood Health Assessment Questionnaire (CHAQ) and the Child Health Questionnaire (CHQ) in 32 countries. Review of the general methodology. *Clin Exp Rheumatol*. 2001;19(4 Suppl 23):S1–9.
13. DeWitt EM, Stucky BD, Thissen D, Irwin DE, Langer M, Varni JW, et al. Construction of the eight-item patient-reported outcomes measurement information system pediatric physical function scales: built using item response theory. *J Clin Epidemiol*. 2011;64(7):794–804. [PubMed: 21292444]
14. Thissen D, Liu Y, Magnus B, Quinn H, Gipson DS, Dampier C, et al. Estimating minimally important difference (MID) in PROMIS pediatric measures using the scale-judgment method. *Qual Life Res*. 2016;25(1):13–23. [PubMed: 26118768]
15. Houghton FT. The Child Health Questionnaire (CHQ-PF50) studies: sincere congratulations and a sincere plea for terminological accuracy. *Clin Exp Rheumatol*. 2002;20(3):436; author reply -7. [PubMed: 12102491]
16. Brunner HI, Klein-Gitelman MS, Miller MJ, Barron A, Baldwin N, Trombley M, et al. Minimal clinically important differences of the childhood health assessment questionnaire. *J Rheumatol*. 2005;32(1):150–61. [PubMed: 15630741]
17. H D, Freels S, RM Y. Plausibility of multivariate normality assumption when multiple imputing non-Gaussian continuous outcomes *Journal of Statistical Computation and Simulation*. 2008;78(1):69–84.
18. Lee KJ, Carlin JB. Multiple imputation for missing data: fully conditional specification versus multivariate normal imputation. *Am J Epidemiol*. 2010;171(5):624–32. [PubMed: 20106935]

19. Hayati Rezvan P, Lee KJ, Simpson JA. The rise of multiple imputation: a review of the reporting and implementation of the method in medical research. *BMC Med Res Methodol*. 2015;15:30. [PubMed: 25880850]
20. McLeod LD, Coon CD, Martin SA, Fehnel SE, Hays RD. Interpreting patient-reported outcome results: US FDA guidance and emerging methods. *Expert Rev Pharmacoecon Outcomes Res*. 2011;11(2):163–9. [PubMed: 21476818]
21. J N IB. *Psychometric Theory* 3rd edition. New York, New York: McGraw-Hill; 1994.
22. Taylor J, Giannini EH, Lovell DJ, Huang B, Morgan EM. Lack of Concordance in Interrater Scoring of the Provider's Global Assessment of Children With Juvenile Idiopathic Arthritis With Low Disease Activity. *Arthritis Care Res (Hoboken)*. 2018;70(1):162–6. [PubMed: 28118522]
23. Armbrust W, Kaak JG, Bouma J, Lelieveld OT, Wulffraat NM, Sauer PJ, et al. Assessment of disease activity by patients with juvenile idiopathic arthritis and the parents compared to the assessment by pediatric rheumatologists. *Pediatr Rheumatol Online J*. 2013;11(1):48. [PubMed: 24368009]
24. Brandon TG, Becker BD, Bevans KB, Weiss PF. Patient-Reported Outcomes Measurement Information System Tools for Collecting Patient-Reported Outcomes in Children With Juvenile Arthritis. *Arthritis Care Res (Hoboken)*. 2017;69(3):393–402. [PubMed: 27159889]
25. Mann CM, Schanberg LE, Wang M, von Scheven E, Lucas N, Hernandez A, et al. Identifying clinically meaningful severity categories for PROMIS pediatric measures of anxiety, mobility, fatigue, and depressive symptoms in juvenile idiopathic arthritis and childhood-onset systemic lupus erythematosus. *Qual Life Res*. 2020;29(9):2573–84. [PubMed: 32410143]
26. Consolaro A, Negro G, Chiara Gallo M, Bracciolini G, Ferrari C, Schiappapietra B, et al. Defining criteria for disease activity states in nonsystemic juvenile idiopathic arthritis based on a three-variable juvenile arthritis disease activity score. *Arthritis Care Res (Hoboken)*. 2014;66(11):1703–9. [PubMed: 24980508]

Significance and Innovations

- Developed and validated a composite disease flare definition for juvenile spondyloarthritis (JSpAflare) designed to closely approximate the disease activity threshold that triggers re-initiation of systemic therapy after therapy de-escalation.
- JSpAflare will be applicable in clinical practice, observational studies, and therapeutic trials and is the first flare outcome for juvenile spondyloarthritis.
- Performance of JSpAflare in comparison to our reference standard of physician-defined active disease plus re-initiation of systemic medication was strong: in the development cohort the AUROC was 0.91, PPV was 87.5%, and NPV was 98.1% and in the validation cohort, AUROC was 0.85, PPV was 92.3%, and NPV was 96.8%.



- PGA plus ≥ 3 cores with meaningful change in remaining 4 cores
- ≥ 2 cores with meaningful change (5 core definition)
- ...●... ≥ 3 cores with meaningful change (5 core definition) - JSpAflare

Figure 1.

Visual representation of the diagnostic test statistic performance of select candidate definitions in the development cohort. AUROC = Area under the receiver operating characteristic curve; PPV = positive predictive value; NPV = negative predictive value; Sens = sensitivity; Spec = specificity.

Table 1.

Patient characteristics in the development and validation cohorts

	Development cohort	Validation cohort
Unique patients	39	36
Therapy de-escalation episodes	45	39
Total visits	180	164
Age (years) at index/reference visit, median (IQR)	14.1 (11.9-17.3) [N=45]	14.3 (12-15.8) [N=39]
Male, Freq. (%)	25 (64.1%) [N=39]	23 (63.9%) [N=36]
bDMARD use at reference visit, Freq. (%)	34 (75.6%) [N=39]	33 (84.6%) [N=36]
TNFi use at reference visit, Freq. (%)	33 (73.3%) [N=39]	29 (74.4%) [N=36]
cDMARD use at reference visit, Freq. (%)	22 (48.9%) [N=39]	22 (56.4%) [N=36]
Core measures at reference visit	Mean (SD)	
Caregiver/patient assessment of well-being (0-10)	0.74 (1.11) [N=45]	0.91 (1.42) [N=39]
Physician assessment of disease activity (0-10)	0.09 (0.36) [N=45]	0.08 (0.29) [N=39]
Caregiver/patient assessment of pain (0-10)	0.84 (1.19) [N=45]	1.08 (1.81) [N=39]
CHAQ (0-3)	0.02 (0.04) [N=16]	0.09 (0.19) [N=24]
PROMIS mobility t-score (14-59) *	56.33 (4.08) [N=29]	48.88 (8.01) [N=8]
PROMIS upper extremity t-score (10-57) *	55.66 (3.62) [N=29]	51.63 (6.52) [N=8]
Active joint count	0 (0) [N=45]	0 (0) [N=39]
Tender entheses count	0.2 (0.73) [N=45]	0.05 (0.22) [N=39]

* PROMIS short forms are converted to T-scores, with higher scores representing more of the trait (i.e. more mobility and higher upper extremity function).

Abbreviations: Freq = Frequency; bDMARD = biologic disease-modifying antirheumatic drug; cDMARD = conventional synthetic disease-modifying antirheumatic drug; CHAQ = Child Health Assessment Questionnaire; PROMIS = Patient Reported Outcomes Measurement Information System.

Table 2.

Model diagnostics (95% CI) for candidate definitions of juvenile spondyloarthritis flare (JSpAflare)

Composite definition in...	Meaningful change	AUROC	PPV	NPV	Sensitivity	Specificity
Development cohort						
PGA plus 2 5 core measures		0.88 (0.78-0.98)	92.9 (66.1-99.8)	97.6 (93.9-99.3)	76.5 (50.1-93.2)	99.4 (96.6-100)
PGA plus 3 5 core measures		0.77 (0.64-0.9)	100 (63.1-100)	95.9 (91.7-98.3)	53.3 (26.6-78.7)	100 (97.8-100)
2 6 core measures		0.96 (0.93-0.98)	56.7 (37.4-74.5)	100 (97.3-100)	100 (80.5-100)	91.3 (85.6-95.3)
3 6 core measures		0.91 (0.81-1)	87.5 (61.7-98.4)	98.1 (94.4-99.6)	82.4 (56.6-96.2)	98.7 (95.4-99.8)
PGA plus 2 4 core measures [^]		0.88 (0.78-0.98)	92.9 (66.1-99.8)	97.6 (93.9-99.3)	76.5 (50.1-93.2)	99.4 (96.6-100)
PGA plus 3 4 core measures [^]		0.83 (0.69-0.97)	100 (63.1-100)	97.6 (93.9-99.3)	66.7 (34.9-90.1)	100 (97.7-100)
2 of the 5 core measures [^]		0.93 (0.87-0.99)	57.1 (37.2-75.5)	99.3 (96.1-100)	94.1 (71.3-99.9)	92.1 (86.5-95.8)
3 of the 5 core measures ^{^**}		0.91 (0.81-1)	87.5 (61.7-98.4)	98.1 (94.4-99.6)	82.4 (56.6-96.2)	98.7 (95.4-99.8)
Validation cohort						
JSpAflare		0.8 (0.64-0.96)	100 (54.1-100)	97.4 (93.4-99.3)	60 (26.2-87.8)	100 (97.6-100)

Legend. Test properties of candidate disease flare definitions. AUROC = Area under the receiver operating characteristic curve; PPV = positive predictive value; NPV = negative predictive value; PGA = Physician global assessment. Core measures tested include: Physician global assessment of disease activity VAS, caregiver/patient assessment of well-being VAS, caregiver/patient assessment of pain VAS, caregiver-/patient-reported physical function, active joint count, and tender entheses count. Meaningful change was considered 2 for all VAS measures, 3 unit change in the PROMIS T-scores, 0.125 unit change in the CHAQ, 1 unit change in active joint count, and 2 unit change in the tender entheses count.

[^] Tender entheses count not included as core measure.

^{**} This is the definition for the "JSpAflare" that was tested in the validation cohort.

Table 3.

Juvenile spondyloarthritis flare (JSpAflare) core measures and frequency of flare visits where each core met the clinically meaningful change from index visit.

Core measures	Clinically meaningful (absolute) change from reference/ index visit	Frequency meeting flare criteria N (%)	
		Development Cohort	Validation Cohort
1. Caregiver/patient assessment of overall well-being	2	12 (75.0)	4 (80.0)
2. Physician assessment of disease activity	2	13 (81.3)	5 (83.3)
3. Caregiver/patient assessment of pain	2	14 (87.5)	3 (50.0)
4. Function: PROMIS mobility or upper extremity function or CHAQ	PROMIS: 3 CHAQ: 0.125	8 (80.0)	2 (40.0)
5. Active joint count	1	12 (75.0)	6 (100.0)

Legend. Flare defined as clinically meaningful worsening in absolute values of 3 core measures from the reference visit. PROMIS = Patient Reported Outcomes Measurement Information System; CHAQ = Child Health Assessment Questionnaire.