ONLINE SUPPLEMENTARY DATA

Supplementary Data 1. Descriptors with definitions of terms

During step 3, besides or instead of the ILAR descriptors, a list of additional items will be collected in order to better characterize the distinct disorders that are proposed for the new PRINTO JIA Provisional Classification Criteria. The list of these descriptors may be further enlarged as new evidence emerges from the literature.

The list includes the following descriptors listed by type and alphabetical order. These descriptors will be checked at each visit.

- *Demographic:* gender, age at onset, age at first visit, at diagnosis, ethnicity.

Signs and symptoms (in alphabetical order)

- *Arthralgia*: joint pain or tenderness
- *Arthritis (active arthritis)*: Swelling within a joint, or limitation in the range of joint movement with joint pain or tenderness, which persists for at least 6 weeks, observed by a physician, and not due to primarily mechanical disorders or to other identifiable causes. Number of joints with swelling, pain/tenderness, limitation on motion as per the PRINTO/PRCSG joint form (available upon request).
- *Arthritis (asymmetry of arthritis):* One single joint or more than 1 non-paired joints to be considered as asymmetric on the day of the first clinical manifestation consistent with the disease.
- *Arthritis (peripheral arthritis):* arthritis involving peripheral joints: elbow, hands/wrist, hip, knees, feet (e.g. sacroiliitis is not counted as peripheral arthritis).
- *Arthritis (oligoarticular course):* oligoarticular course refers to patients with less than 5 active joints during the entire course of the disease.
- *Arthritis (polyarticular course):* polyarticular course refers to patients with at least 5 active joints during the entire course of the disease.
- *Dactylitis*: Swelling of one or more digits (a finger or toe) or sausage-like digital inflammation involving the entire digit, usually in an asymmetric distribution, which extends beyond the joint margin and which may be painful.
- *Enthesitis*: Tenderness at the insertion of a tendon, ligament, joint capsule, or fascia to bone.
- *Fever (quotidian fever):* Fever that rises to ≥39°C once a day and returns to ≤ 37°C between fever peaks. Fever should be quotidian fever for at least 3 consecutive days and re-occurring over a duration of at least 2 weeks.
- *Inflammatory bowel disease* (IBD) (Crohn or ulcerative colitis): p*resence of IBD.*
- *Joint involvement (Type and number of joints involved)*: Joints that can be individually evaluated clinically are counted as separate joints as per the PRINTO-PRCSG rheumatologic examination form v. 2013.
- *Lymph node (generalized) enlargement and/or hepatomegaly and/or splenomegaly by clinical examination.*
- *Macrophage Activation Syndrome (MAS) ACR/EULAR/PRINTO classification criteria:* A febrile patient with known or suspected systemic JIA is classified as having MAS if the following criteria are met: Ferritin > 684 ng/ml and any 2 of the following: Platelet count ≤ 181 x109/l, Aspartate aminotransferase (AST) > 48 U/l, Triglycerides > 156 mg/dl, Fibrinogen ≤ 360 mg/dl
- *Nail pitting*: A minimum of 2 pits on one or more nails at any time in the absence of other causes of nail pitting.
- *Onycholysis*: painless detachment of the nail from the nail bed, usually starting at the tip and/or sides.
- *Organ involvement* (lung, heart, etc.) as defined by the treating physician.
- *Pain (Inflammatory back pain)*: insidious onset, improves with exercise, does not improve with rest, occurs mostly at night.
- *Psoriasis*: As diagnosed by a physician (not necessarily a dermatologist).

- *Psoriasis (family history): family history of psoriasis* in a first degree relative.
- Rash: evanescent erythematous rash, which is typically non-fixed lasting a few hours, non-pruritic, macular, and salmon colored on the trunk and extremities.
- *Sacroiliac joint tenderness*: Presence of tenderness on direct compression over the sacroiliac joints. To be differentiated from fibromyalgic pain.
- *Sacroiliitis (sacroiliitis* on X-Ray*)*: Grade 2 bilateral radiographic sacroiliitis or grade 3
- *Sacroilitis* (*sacroiliitis* on MRI): MRI criteria: 1. Bone marrow edema (BMO) on a T2-weighted sequence (required criteria) sensitive for free water (such as short tau inversion recovery (STIR) or T2FS) or bone marrow contrast enhancement on a T1-weighted sequence (such as T1FS post-Gadolinium). 2. Inflammation must be clearly present and located in a typical anatomical area (subchondral bone). 3. MRI appearance must be highly suggestive of SpA.
- *Serositis*: Pericarditis and/or pleuritis and/or peritonitis.
- *Sore throat* as defined by the treating physician.
- *Spondyloarthritis:* chronic inflammatory arthritis primarily affecting the axial skeleton, with a characteristic involvement of the spine and sacroiliac joints.
- *Spondyloarthritis (family history):* family history of spondyloarthritis in a first degree relative.
- *Uveitis (chronic)*: Chronic anterior posterior or pan-uveitis as diagnosed by an ophthalmologist.
- *Uveitis (acute):* uveitis characterized by sudden onset and limited duration as diagnosed by an ophthalmologist.
- *Vasculitis and subcutaneous nodules*

Laboratory and other tests

- *Ophthalmologic examination:* as per established recommendation of the American Academy of Pediatrics (AAP) so called SUN criteria
- *Antinuclear antibodies (ANA):* At least 2 positive results (in immunofluorescence on HEp-2 with a titer ≥1/160), at least 3 months apart, during the first 6 months of disease
- *Rheumatoid Factor (RF)*: At least 2 positive results, at least 3 months apart, during the first 6 months of disease.
- *Anti CCP:* to be collected only in non-systemic JIA patients
- *HLA-B27:* in suspected enthesitis/spondylitis related JIA
- ESR, CRP, blood count

Other data to be collected

- *Therapies:* synthetic and biologic Disease Modifying Anti Rheumatic Drugs (DMARDs) for all JIA disorders
- *Therapies (response to therapy in systemic JIA)*: Rapid and complete response to anti-IL1
- *Therapies (*Enthesitis/spondylitis related JIA*)*: good response to NSAIDs
- Others (as per physician decision)

Supplementary Data 2. Methodological approach for the validation of the PRINTO JIA classification criteria.

Step 3. Large-scale data collection for evidence based validation

A prospective cohort of at least 1,000 JIA patients at onset and longitudinal follow-up for up to 5 years will be collected. This process will represent the core of the validation process. Data collection will involve both data needed for the current ILAR criteria, the "provisional new PRINTO JIA classification criteria" and adult criteria.

Data will be collected via web on patients with first visit to a pediatric rheumatologist within 6 months from disease onset. The schedule for data collection will be at onset, within 6 months from onset, at least 3 months thereafter and then annually up 5 years of follow-up.

In silico analysis and prospective validation on real new onset JIA patients

The project will take advantage of two large databases collected within the framework of the following projects of the PRINTO network (1):

- EPidemiology, treatment and Outcome of Childhood Arthritis (EPOCA), which has recently enrolled over 9,000 JIA and over 4,800 healthy children.
- Pharmacovigilance in juvenile idiopathic arthritis patients (Pharmachild) treated with biologic agents and/or methotrexate (MTX) which has enrolled over 8,000 JIA.

The databases contain well-characterized data regarding the current ILAR criteria (category definitions, exclusion criteria, additional laboratory parameters such as rheumatoid factor-RF, anti-nuclear antibodies-ANA, HLA B27, etc.).

The *in silico* study will provide useful information on:

a) how accurately the ILAR criteria are applied in the everyday clinical practice and how many patients were categorized as undifferentiated arthritis (a residual large category which contains patients with arthritis that

fulfill criteria in no category, e.g. because of the presence of exclusion criteria, or arthritis that fulfills criteria in 2 or more categories).

b) the statistical performance (Agreement with the physician diagnosis) of the new PRINTO 2016 criteria and old ILAR criteria for the related JIA categories. It is indeed not possible to test in silico the validity of the new PRINTO JIA classification as a whole since some information that is needed (for instance anti-CCP antibodies, standardized ANA evaluation) is not present in the databases. However, the analysis will provide information on the performance of some of the new criteria and will be valuable to tune the sensitivity of some of them.

Prospective collection of at least 1,000 JIA new patients at disease onset including biologic samples collection.

In parallel with the *in silico* analysis, the project includes the enrollment of a prospective cohort of at least 1,000 new onset JIA patients, evaluated within 6-month from onset, at least 3 months apart and then annually up to year 5. Related biologic samples (serum and DNA) will be collected at 2 time points at least 3-month apart for central evaluation of ANA, anti CCP, RF, HLA B27 determinations with storing of the serum and DNA left over. This phase represents the core of the validation process. It will also include, in selected centres, ultrasound evaluation at each visit in order to examine the potential added value that might be provided to the new classification criteria by the use of imaging techniques.

Inclusion criteria for the patients to be collected prospectively

The patients to be included in the prospective data collection must fit the following inclusion criteria:

1. A diagnosis of JIA according to the ILAR criteria by the treating physician.
2. The availability to provide an evaluation within 6 months after the onset of JIA. The onset evaluation can also be completed retrospectively (based on reliable family history or prior attending physician's reports), but only if joint assessment data can be provided.
3. A clinical evaluation 3 months apart and then annually up to year 5.

The following forms will need to be completed:

AT ONSET EVALUATION

1. Demographic and subject history
2. Clinical history
3. Family history
4. Other clinical information
5. Joint assessment
6. Drug therapy
7. Laboratory form
8. Biologic samples
9. Back assessment
10. Joint injection form (only in case of intra-articular injections)
11. Tender entheseal assessment (only if spondyloarthritis is present)
12. Dactylitis assessment (only if dactylitis is present)
13. Psoriasis assessment (only if psoriasis is present)
14. Uveitis (only if uveitis is present)
15. Ultrasound assessment (for some selected centres only)

FOLLOW-UP EVALUATIONS

1. Clinical information
2. Joint assessment
3. Drug therapy
4. Laboratory form
5. Biologic samples
6. Back assessment

7. Joint injection form (only in case of intra-articular injections)
8. Tender entheseal assessment (only if spondyloarthritis is present)
9. Dactilitis assessment (only if dactilitis is present)
10. Psoriasis assessment (only if psoriasis is present)
11. Uveitis (only if uveitis is present)
12. Ultrasound assessment (for Italian centres only)

Biologic samples to collect at baseline and second visit

3 EDTA samples (5ml each)

1 serum sample (2 ml)

Statistical analysis

The main validation analysis will be performed on the data at 12 months.

The performance of both the ILAR, the new provisional PRINTO classification criteria and adult criteria and others to be identified from literature and from the paediatric rheumatology community (via Delphi questionnaires), will be compared with respect to their identification of more homogenous entities.

Patients included in the "other" category will be analyzed according to the presence of various already identified "descriptors". The purpose will be to see if some of these descriptors cluster together allowing the identification of new homogenous JIA categories.

Furthermore, we will examine whether any of these descriptors, for instance the presence of psoriasis, identifies a defined disease and/or modifies the clinical phenotype of other already well-characterized forms of chronic arthritis.

In addition, currently the old and new classification criteria entail a thorough clinical evaluation and few selected laboratory tests (RF, anti CCP, ANA, HLA B27) which will be centrally performed at the PRINTO headquarters; this will be accomplished by storage of serum on at least 2 occasions 3 months apart and of cells for HLA-B27 determination. Serum samples and DNA will be stored for genetic evaluations as well.

It is expected that it will take 3 months to develop the protocol, case report forms via web database, the biologic samples collection guidance document, the Standard Operating Procedures (SOP), training and training material; 12 months for ethics committee approval for all participating centers; 24 months for training, enrolment of patients and biologic sample central storage; 6 months for analysis of clinical data and central biologic samples evaluation for ANA, anti CCP, RF and HLA B27.

Statistical analysis.

Descriptive statistics will be reported as medians and interquartile ranges for continuous variables, and as absolute frequencies and percentages for categorical variables. Comparisons of quantitative variables between 2 groups will be made by means Mann-Whitney U test, whereas comparisons of quantitative variables between more than 2 groups will be made by nonparametric analysis of variance (Kruskal-Wallis test). Dunn's test will be chosen as an a posteriori test to assess the statistical significance of differences between pairs of groups. Categorical data will be compared by chi-square test, or by Fisher's exact test in case of expected frequencies ≤5. Bonferroni adjustment will be applied as a correction for multiple comparisons to explore post hoc differences between pairs of groups. Multiple logistic regression analysis will be performed, entering explanatory variables that showed significant results in univariate tests ($P < 0.05$) or will be considered a priori to be of foremost importance for the study outcome. Cases with missing variables will be excluded from the analysis. Before the application of logistic regression procedures, some continuous variables will be dichotomized to binary variables through receiver operating characteristic (ROC) curve analysis. The step-down strategy of analysis will be chosen; to examine the effect of removing variables from the saturated model. Possible explanatory variables assessed will be demographic data, joint count, iridocyclitis, etc. The effect will be expressed in terms of odds ratios, and 95% confidence intervals (CI) with statistical significance tested by likelihood ratio test. The area under the ROC of the best fitting model will be used as an indicator of the predictive ability of the model. The association between categorical variables will be further explored by multiple correspondence analysis (5).

Multiple correspondence analysis is an exploratory analysis and descriptive technique that enables analysis of contingency tables with a large number of variables, considering measures of correspondence between rows and columns. It can be considered a variant of factor analysis, but it differs because it evaluates the relationship

between categorical data organized in contingency tables, rather than continuous data. Essentially, multiple correspondence analysis converts a non-negative data matrix into a kind of graphic representation that allows studying the relationship between the categories in a multiway contingency table. The importance of the categories of variables for constructing each axis is measured by their absolute contribution and aids in the interpretation of the axis. Categories with high absolute contribution show greater importance in the factor's formation. The relative contribution provides information as to how much of the category's variability is explained by the axis. Graphic analysis of association of variables is performed by considering their geometric proximity and the separation of categories by quadrants, because the closer the variables, the more interrelated they are, and those separated by quadrants display groups with opposite profiles. The variables that will prove statistically significant in multivariate analysis will be used in the multiple correspondence analysis. Only patients for whom complete information will be available will be included.

All statistical tests will be 2-sided; *P* values less than 0.05 will be considered significant. The statistical packages to be used will be Statistica (version 8.0; StatSoft) for univariate analyses, Stata release 11 (StataCorp) for multivariate and cluster analysis or SAS as appropriate. For multiple correspondence analysis, the software XLSTAT, 6.1.9 (Addinsoft) will be used.

Step 4. Final consensus conference

Once the prospective data collection will be finished and analyzed, a final consensus conference with NGT will be organized to discuss the results and finalize the new, validated PRINTO JIA classification criteria.