Published in final edited form as:

Arthritis Care Res (Hoboken). 2019 May ; 71(5): 579–590. doi:10.1002/acr.23834.

THE AMERICAN COLLEGE OF RHEUMATOLOGY PROVISIONAL CRITERIA FOR CLINICALLY RELEVANT IMPROVEMENT IN CHILDREN & ADOLESCENTS WITH CHILDHOOD-ONSET SYSTEMIC LUPUS ERYTHEMATOSUS

Hermine I. Brunner, MD[1], **Michael J. Holland, MD**[1], **Michael W. Beresford, MBBS**[2], **Stacy P. Ardoin, MD**[3], **Simone Appenzeller, MD**[4], **Clovis A. Silva, MD**[5], **Francisco Flores, MD**[6], **Beatrice Goilav, MD**[7], **Pinar Ozge Avar Aydin, MD**[1], **Scott E. Wenderfer, MD**[8], **Deborah M.**

Address reprint requests to: Hermine Brunner MD, MSc., Cincinnati Children's Hospital Medical Center, University of Cincinnati, William S. Rowe Division of Rheumatology, E 4010, 3333 Burnet Avenue, Cincinnati, OH 45229-3039, USA; hermine.brunner@cchmc.org.

AUTHOR DISCLOSURES

Hermine I. Brunner:
Consultant: AbbVie, Ablynx, Amgen AstraZeneca, Baxalta Biosimilars, Biogen Idec, Boehringer Ingelheim, Bristol-Myers Squibb, Celgene, Eli-Lilly, EMD Serono, Gilead Sciences, Janssen, MedImmune, Novartis, Pfizer, R-Pharm, Roche, Servier, Takeda (all < 10K) //// Speaker's bureaus: Genentech, Novartis (> 10k) ///// Scientific grant support from Pfizer and Bristol-Myers Squibb for research studies in adult lupus and JIA
Dr. Brunner is a full-time employee of Cincinnati Children's Hospital which has received contributions from: Bristol-Myers Squibb, Hoffman-La Roche, Janssen, Novartis, and Pfizer for the coordination activity of the PRCSG network.

Michael J. Holland None declared

Michael W. Beresford None declared

Stacy P. Ardoin None declared

Simone Appenzeller None declared

Clovis A. Silva None declared

Francisco Flores None declared

Beatrice Goilav None declared

Pinar Ozge Avar Aydin None declared

Scott E. Wenderfer None declared

Deborah M. Levy None declared

Angelo Ravelli <$10,000: Personal fees (Consultancies & speaking fees) from: AbbVie, BMS, Pfizer, Hoffman LaRoche, Novartis, Centocor, "Francesco Angelini" and Reckitt Benckiser.

Raju Khubchandani None declared

Tadej Avcin <$10,000: AbbVie (consultancies, speaking fees) Octapharma (consultancies, speaking fees) Boehringer Ingelheim (consultancies)

Marisa S. Klein-Gitelman None declared

Nicolino Ruperto
Dr Nicolino Ruperto received honoraria (<10.000 USD each) for consultancies or speaker's bureau from the following pharmaceutical companies since the last 5 years: Abbott, AbbVie, Amgen, Biogenidec, Astellas, Alter, AstraZeneca, Baxalta Biosimilars, Boehringer, BMS, CD-Pharma,Celgene, CrescendoBio, EMD Serono, Hoffman-La Roche, Italfarmaco, Janssen, MedImmune, Medac, Novartis, Novo Nordisk, Pfizer, Rewind Arms, R-Pharma, Sanofi Aventis, Servier, Sinergie, Takeda, Vertex, UCB Biosciences GmbH. IRCCS Istituto Giannina Gaslini, which is the public Hospital where NR works as full time public employee, has received contributions (>10.000 USD each) from the following industries: Abbott, BMS, "Francesco Angelini", GlaxoSmithKline (GSK), Hoffman-La Roche, Italfarmaco, Janssen, Novartis, Pfizer, Sanofi Aventis, Schwarz Biosciences, Sobi, Xoma, Wyeth. This money was reinvested for the research activities of the hospital in a fully independent manner without any commitment with third parties.

Brian M. Feldman None declared

Jun Ying None declared

CONFLICT OF INTERESTS

Levy, MD[9], **Angelo Ravelli, MD**[10], **Raju Khubchandani, MD**[11], **Tadej Avcin, MD**[12], **Marisa S. Klein-Gitelman, MD**[13], **Nicolino Ruperto, MD**[10], **Brian M. Feldman, MD**[9], **Jun Ying, PHD**[14], and **Paediatric Rheumatology International Trial Organisation & the Pediatric Rheumatology Collaborative Study Group**

[1]Department of Pediatrics, University of Cincinnati College of Medicine and Division of Rheumatology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH

[2]Department of Paediatric Rheumatology, Institute of Translational Medicine, University of Liverpool, Liverpool, UK; Alder Hey Children's NHS Foundation Trust, Liverpool, UK

[3]Department of Pediatrics and Internal Medicine, Ohio State University, Division of Rheumatology, Nationwide Children's Hospital and Wexner Medical Center, Columbus, OH

[4]Rheumatology Unit-School of Medical Science; University of Campinas, Campinas, Brazil

[5]Children's Institute, Hospital das Clinicas HCFMUSP; Faculdade de Medicina da Universidade de São Paulo, Brazil

[6]Department of Pediatrics, University of Cincinnati College of Medicine and Division of Nephrology & Hypertension, Cincinnati Children's Hospital Medical Center, Cincinnati, OH

[7]Department of Pediatrics, The Children's Hospital at Montefiore, Division of Nephrology and Albert Einstein College of Medicine, Bronx, NY

[8]Department of Pediatrics, Baylor College of Medicine; Renal Section, Texas Children's Hospital Houston, TX

[9]Department of Pediatrics, University of Toronto and Division of Rheumatology, The Hospital for Sick Children, Toronto, Canada

[10]Istituto G. Gaslini, Clinica Pediatrica e Reumatologia, PRINTO, Genoa, Italy

[11]Pediatric Rheumatology; Jaslok Hospital Mumbai, India

[12]Department of Allergology, Rheumatology and Clinical Immunology, University Children's Hospital, University Medical Centre Ljubljana, Slovenia

[13]Department of Pediatrics, Northwestern University Feinberg School of Medicine and Division of Rheumatology, Ann and Robert Lurie Children's Hospital of Chicago, Chicago, IL

[14]Department of Environmental Health Sciences, University of Cincinnati, Cincinnati, OH

Abstract

Objectives: To develop a *Childhood Lupus Improvement Index (CHILI)* as a tool to measure response to therapy in childhood-onset systemic lupus erythematosus (cSLE), with focus on clinically relevant improvement (CRI_{cSLE}).

Methods: Pediatric nephrology and rheumatology subspecialists (n=213) experienced in cSLE management were invited to define CRI_{cSLE} and rate a total of 433 unique patient-profiles for the presence/absence of CRI_{cSLE}. Patient-profiles included the cSLE core response variables [cSLE-CRVs: global assessment of patient well-being (Patient-global), physician assessment of cSLE activity (MD-global), disease activity index score (here: Systemic Lupus Erythematosus Disease

Activity Index), urine protein-to-creatinine ratio (UPCR), Child Health Questionnaire physical summary score (CHQ-PhS)]. Percentage and absolute changes of these cSLE-CRVs (baseline vs. follow-up) were considered to develop candidate algorithms and validate their performance [sensitivity, specificity, area under the receiver operating characteristic curve (AUC; range: 0 −1)].

Results: Using an international consensus conference, unanimous agreement on a definition of CRI_{cSLE} was achieved; cSLE expects (n=13) concurred (100%) that the preferred CHILI algorithm considers absolute changes of the cSLE-CRVs. After transformation to range from 0 −100, a CHILI score of ≥54 had outstanding accuracy for identifying CRI_{cSLE} (AUC=0.93; sensitivity=81.1%; specificity=84.2%); CHILI scores also reflect minor, moderate and major improvement for values exceeding 15, 68 and 92 (all: AUC ≥ 0.92, sensitivity: ≥93.1%; specificity: ≥73.4%).

Conclusions: The CHILI is a new, seemingly highly accurate index for measuring clinically important improvement in cSLE over time. This index is useful to categorize the degree of response to therapy in children and adolescent with cSLE.

Keywords

lupus; childhood-onset SLE; SLE; pediatric SLE; juvenile SLE; improvement; criteria; children; cSLE

INTRODUCTION

Systemic lupus erythematosus is a complex, chronic multi-system autoimmune inflammatory disease, with up to 20% of patients diagnosed during childhood (cSLE) (1, 2). When disease commences early in life rather than during adulthood, lupus has a poorer prognosis, particularly due to multi-organ and kidney involvement (3, 4). The course of cSLE is characterized by episodes of disease flares; followed by periods of improvement, generally due to more intensive drug therapy. There is international consensus around a core set of variables (cSLE-CRVs) that should be considered when assessing response to therapy and flare of cSLE (5, 6). Considering changes in cSLE-CRVs, a Provisional ACR/cSLE Flare Score can be calculated to identify patients who experienced a minor, moderate or severe flare of cSLE (7, 8). Likewise, percentage changes of the cSLE-CRVs are the basis for the Pediatric Rheumatology International Trials Organization, American College of Rheumatology (ACR) Provisional Criteria of Response to Therapy (9). We have previously shown, albeit in a rather small dataset, that the PRINTO/ACR Provisional Criteria for Response to Therapy and, to a lesser extent, the Systemic Lupus Responder Index are both very well-suited to capture major improvement of cSLE; however, both the PRINTO/ACR Provisional Criteria of Response to Therapy and the Systemic Lupus Responder Index appeared less apt to identify patients who experienced moderate or minor improvement of cSLE (10). At present, there are no generally accepted criteria or algorithms to measure various degrees of improvement with cSLE, and consensus is lacking of what constitutes clinically relevant improvement (CRI_{cSLE}) in children and adolescents with cSLE. The latter is especially relevant because in studies of rheumatoid arthritis an ACR 20% level (ACR20) response, or in juvenile idiopathic arthritis an ACR 30% level (JIA-ACR30) response, provide such a measure of clinically relevant improvement. ACR20 and JIA-ACR30

responses, respectively, are regarded improvement thresholds that can support labeling of new medications by the Food and Drug Administration or the European Medicines Agency (11, 12). Prior to developing criteria, or algorithms to measure $CRI_{cSLE,}$ it is necessary to achieve consensus around a definition of CRI_{cSLE}.

Building on prior international consensus around the cSLE-CRVs that are needed to capture response to therapy in cSLE (9), the objectives of this study were to define CRI_{cSLE}, and develop as well as initially validate criteria to measure CRI_{cSLE}. Further, we sought to measure minor, moderate and major response to therapy in cSLE.

PATIENTS AND METHODS

The overall approach to this project was based on the methodological framework successfully employed in pediatric rheumatology criteria development in the past (9, 13, 14), which is aligned with recommendations of the ACR Criteria Subcommittee and the Quality of Care Committee (15). As is summarized in Figure 1, an initial Delphi survey was conducted among 114 pediatric rheumatologists and nephrologists with expertise in cSLE (1) to delineate key features for judging whether a patient experienced CRI_{cSLE} [**Step 1**]. Subsequently, participants in a Consensus Conference rated 200 Patient-Profiles [**Step 2**]. During a Consensus Conference the results of Step 1 and 2 were reviewed to support consensus formation around a definition of CRI_{cSLE} [**Step 3].** This was followed by a second round of patient-profiles sent to 200 pediatric rheumatologists and the cSLE experts who participated in the Consensus Conference. The resulting dataset was randomly split in a training-dataset and a validation-dataset [**Step 4]**. The training-dataset was used to develop candidate criteria for CRI_{cSLE} [**Step 5**]. These candidate criteria were tested using the validation-dataset **[Step 6**]. As done in Step 3, agreement was achieved around a preferred ***Chi****ldhood* ***L****upus* ***I****mprovement Index (CHILI)* algorithm among cSLE experts with voting rights who had participated in Consensus Conference [**Step 7**].

Step 1 – Delphi Survey regarding CRI_{cSLE}

The 13 expert participants in the Consensus Conference and 100 of the pediatric rheumatologists who contributed in the development of other cSLE criteria sets (6, 8) received a Delphi Survey inquiring about cSLE (1) characteristics and changes of cSLE-CRVs that would support the presence of CRI_{cSLE}. The Delphi survey was piloted (HBR, PÖA). Principles and recommendations for the design and conduct of online surveys were followed (16).

Step 2 –PP ratings prior to the Consensus Conference

Using prospective data of cSLE patients of the CCHMC Lupus Registry (17), the PRINTO Lupus Cohort (6), and a multicenter North American cSLE Cohort (U01-AR5868; PI Brunner), we developed 1,482 unique patient-profiles. After omitting patient-profiles with >2 missing data elements and some patient-profiles without changes in cSLE-CRVs between visits, there were 433 unique patient-profiles. Missing observations of these 433 patient-profiles were imputed using multiple imputation methods and expectation–maximization algorithms in computation (18–20).

Each patient-profile provided the following patient data at the time of a baseline visit and a follow-up visit: [1] physician assessment of cSLE activity as measured on a visual analog scale (VAS) (MD-global; 0 = inactive disease; 10 = very active disease); [2] parent assessment of patient overall well-being, measured on a VAS (Patient-global; 0 = very poor; 10 = very well); [3] proteinuria, measured by timed urine collection or UPCR from spot urine; [4] erythrocyte sedimentation rate; [5] levels of complement C3 and C4; [6] item and summary scores of the Systemic Lupus erythematosus Disease Activity Index (SLEDAI; version 2k] (21); and [7] the Child Health Questionnaire Physical summary score (version P50; CHQ-PhS) (5, 6). Information about complete blood counts and differential, serum chemistry, erythrocyte sedimentation rate, urinalysis and anti-dsDNA antibody concentrations were also provided.

Thirteen cSLE experts (HIB, MWB, SPA, SA, CAS, FF, BG, SEW, DML, AR, RK, TA and MKG) who were voting participants at a Consensus Conference were asked to rate 200 of the 433 patient-profiles prior to the meeting. After the Consensus Conference, these cSLE experts plus to 200 pediatric rheumatologists who previously participated in a similar patient-profiles rating exercise (6–8) were asked to each rate 50 patient-profiles which were randomly selected from the pool of 433 patient-profiles. Each patient-profiles rater was asked to assess the disease course, using the following response options: (*Question A*) major improvement; moderate improvement; minor improvement; unchanged or worse or “I do not have enough information to make this assessment”. Further, if a **patient-profile** rater considered improvement to be present, then he/she was asked whether improvement constituted CRI_{cSLE} or not (*Question B*). In this context, minor improvement can be considered equivalent to ‘any improvement’ with cSLE. The survey source data were batch-processed, and open source online survey software, RedCap survey, was used for response management and as a presentation layer (see https://www.project-redcap.org/).

The minimum number of rater responses to each patient-profile was 16, and all patient-profiles were considered in the subsequent adjudication process. Considering that patient-profile raters may not necessarily agree on the interpretation of the disease course of a given patient-profile, the “true” overall course of cSLE for a given patient-profile was adjudicated using the *Majority-Rule,* i.e. the majority of the raters of a patient-profile agreed on a given disease course. Other Rules calculated, including the *67%-Rule*, i.e. at least 2/3 of the raters agreed on a given disease course. Irrespective of the Rule used, results were similar to the Majority-Rule. Hence, we present mainly the results from Majority Rule analyses.

Three statistical strategies were employed to develop a series of candidate criteria to measure CRI_{cSLE}: **(a)** we considered the PRINTO/ACR Provisional Criteria of Response to Therapy (9) which have been previously validated to measure improvement in cSLE. Furthermore, we developed algorithms that considered **(b)** absolute change or **(c)** relative or percentage changes of the cSLE-CRVs between baseline and follow-up using multinomial logistic regression. Strategies (b) and (c) yield a numeric “*CHILI score*” (or log odds of improvement) calculated from the combined changes of the cSLE-CRV predictors between baseline and follow-up (9, 22).

Accuracy of the PRINTO/ACR Provisional Criteria of Response to Therapy was tested using kappa (κ) statistics. With respect to the criterion standard (here: adjudicated disease course from the patient-profiles ratings), κ values can be interpreted as follows: poor agreement: $\kappa <$ 0.4; fair to good agreement: κ = 0.4– 0.75; substantial to excellent agreement: $\kappa >$ 0.75. For each of the candidate CRI_{cSLE} algorithms from multinomial regression analysis, diagnostic accuracy was assessed by receiver's operating characteristic (ROC) curve analysis. The area under the ROC curve (AUC, range 0 −1) was calculated, and the diagnostic *accuracy* was considered outstanding, excellent, good, fair, and poor if the AUC was in the range of 0.9 – 1.0, 0.81 – 0.90, 0.71 – 0.80, 0.61 – 0.70, and <0.60, respectively (23).

Based on prior consensus [Step 3], threshold CHILI scores reflect the highest conditional AUC among all candidate thresholds on the ROC curve, i.e. the point on the ROC curve with the highest precision of correctly classifying the degree of cSLE improvement level (CRI_{cSLE}; minor, moderate, major,).

All analyses were done using SAS 9.4 (SAS, Cary, NC) software and SYSTAT 12 (Systat Software Inc., Chicago, IL) software. P-values < 0.05 were considered statistically significant.

Step 3 – Consensus Definition of CRI_{cSLEE}

Participants in the Consensus Conference were experienced pediatric rheumatologists and nephrologists from South America, North America, Asia, and Europe with substantial clinical and research experience in cSLE. PRINTO leadership (NR) participated in the discussions during the Consensus Conference as a non-voting content expert. *A priori*, the consensus level at the consensus conference was set at 75%, i.e. comparable or even somewhat higher than that chosen for similar studies in the past (6, 13, 24). Using nominal group technique guided by an experienced moderator (BMF), the expert panel reached agreement around the definition of CRI_{cSLE}. The panel also reviewed the performance of the Provisional PRINTO/ ACR Criteria of Response to Therapy and candidate improvement algorithms derived by multinomial logistic regression using PP-ratings from Step 2, considering the OMERACT Filter (25–27): [1] feasibility, i.e. practicability: can the items be measured easily?; [2] reliability, i.e. reproducibility: can the items be measured precisely?; [3] redundancy: are there two or more items included in the candidate criteria measuring the same aspect of the disease?); [4] face validity, i.e. credibility: are the criteria sensible?; [5] content validity, i.e. comprehensiveness: do the criteria sample all of the domains of the disease?; [6] criterion validity: based on AUC, do the criteria accurately approximate the "gold standard", i.e. the adjudicated disease course as Majority-Rule?; [7] sensitivity and specificity: do the criteria effectively identify patients with CRI_{cSLE} and/or various levels of improvement and distinguish them from patients who do not experienced CRI_{cSLE} and/or various levels of improvement?; and [8] discriminant validity: do the criteria detect the smallest clinically important change?; i.e. discriminate patients with one of the following disease courses: CRI_{cSLE}; minor improvement, moderate improvement, major improvement, unchanged or worse.

Step 4 – Second round of Patient-Profile ratings

Besides individuals who were invited to participate in Step 1 and 2, the 433 patient-profiles were then sent to another 100 pediatric rheumatologists who previously participated in a similar study (7). Hence, a total of 213 patient-profile raters received 50 randomly selected patient-profiles each; formats, response options, adjudication were described in Step 2. The resulting dataset was divided in the sequence of acquisition into a training-dataset and a validation-dataset.

Steps 5 and 6 – Development and preliminary validation of the CHILI

Using the training-dataset [Step 4], we newly developed candidate algorithms to measure improvement (CRI_{cSLE}, minor, moderate, major) as described in Step 2. In these algorithms, CRI_{cSLE} was considered to be a special threshold score among many possible improvement scores. Threshold scores were transformed to range between 0 and 100. The algorithms and thresholds developed in the training-dataset [Step 5] were validated using the validation-dataset to derive at preliminary CHILI criteria [Step 6].

Step 7 - Ranking of Preliminary CHILI algorithms after the Consensus Conference

The analyses from Steps 5 and 6 were presented to the Consensus Conference participants with voting rights. These cSLE experts were asked whether, in the setting of a clinical trial, (1) CRI_{cSLE} algorithms from multinomial logistic regression were preferable to the use of the PRINTO/ACR Provisional Criteria for Response to Therapy; (2) absolute differences of the cSLE-CRVs were superior to percentage changes when measuring $CRI_{cSLE;}$ and; (3) these algorithms were useful for categorizing the degree of improvement (minor, moderate, major) in cSLE.

RESULTS

Definition of CRI_{cSLE}

The survey [Step 1] inquired about changes in cSLE-CRVs, signs and symptoms with CRI_{cSLE}. Among the 113 pediatric rheumatologists and nephrologists approached for survey participation 92 responded (81%). Survey participants from different regions or less vs. more than 10 years of experience in treating cSLE did not differ significantly in their responses (data not shown). There was ≥ 80% agreement that, with CRI_{cSLE}, the MD-global and/or the score of a disease activity index must be better or unchanged; and that patients with CRI_{cSLE} could experience new organ involvement as long as this did not involve the neuropsychiatric hematological, gastrointestinal, renal, ophthalmological, or cardiopulmonary organ systems. The initial ratings of 200 patient-profiles provided additional data regarding the measurement of the CRI_{cSLE} (see Supplemental table 1 for adjudication results). After review of this information during the Consensus Conference (Step 3), there was 100% agreement for the following consensus definition of CRI_{cSLE}: *"a clinically relevant improvement has occurred in a child with lupus if there are reduced signs of disease from active lupus. Although there may not be improvement of lupus activity in all organ systems, there cannot be increased lupus activity in a major organ system, i.e. neuropsychiatric hematological, gastrointestinal, renal, ophthalmological, or*

cardiopulmonary organ systems. Patient symptoms will be at least stable, and immunosuppressive therapy should be unchanged or decreased". Further, cSLE experts concluded that further testing of the PRINTO/ACR Provisional Criteria of Response to Therapy in cSLE was warranted, and that a multinomial logistic regression modeling should be pursued to measure CRI_{cSLE} with threshold choice at the statistical optimal point on the AUC (both ≥92 % agreement).

Post Consensus Conference patient profile ratings

As part of Step 4, the 433 patient-profiles were sent to 213 patient-profile raters. The response rate was 91% (194/213, see Appendix 1), and all 433 patient-profiles qualified for adjudication. The resulting dataset was split in a training-dataset (200 patient-profiles) and a validation-dataset (n=233). Baseline characteristics of the patients represented in these datasets are summarized in Table 1.

Using the Majority Rule, there were 95 (47.5%) patient-profiles without CRI_{cSLE} and 105 patient-profiles (52.5%) with CRI_{cSLE} in the training-dataset. Among patient-profiles adjudicated to reflect CRI_{cSLE}, 83% were considered to represent moderate or major improvement of cSLE, while 99% of patient-profiles without CRI_{cSLE} were adjudicated to reflect at most minor improvement of cSLE.

Performance of individual cSLE-CRVs to measure CRI_{cSLE}

Based on univariate logistic regression in the training-dataset (Table 2), absolute changes and percentage (or relative) changes of the cSLE-CRVs had similar discriminative properties to detect CRI_{CSLE} (Table 2). However, only absolute changes of the UPCR ($P< 0.001$) between baseline and follow-up but not percentage changes ($p=0.132$) significantly differed among patients with vs. without $CRI_{CSLE.}$ Different from the other cSLE-CRVs but irrespective of the type of change (absolute, relative) considered, the UPCR had only fair accuracy (AUC≤0.67) in capturing CRI_{CSLE}. Individually, the MD-global and the SLEDAI had the highest accuracy (both AUC≥0.90) for identifying the CRI_{CSLE} status.

Performance of the PRINTO/ACR Provisional Criteria of Response to Therapy to measure CRI_{cSLE}

As shown in Table 3 and Supplemental Table 2, in both the training-dataset or validation-dataset, the PRINTO/ACR Provisional Criteria for Response to Therapy had at most fair accuracy for capturing CRI_{CSLE} status ($k \leq 0.3$ for Majority Rule; $\kappa \leq 0.43$ for 67% Rule**)**. The same was also true for measuring various levels of improvement (all: $\kappa < 0.34$ for Majority Rule and 67% Rule).

Development of the CHILI to measure CRI_{cSLE}

As part of Step 5 analyses (Table 4), we used multinomial regression to generate candidate algorithms that considered the cSLE-CRVs that were identified to relevant for capturing improvement of cSLE (6, 22). Irrespective of the type of change, i.e. absolute or percentage differences of the cSLE-CRVs between baseline and follow-up, algorithms were similar in their accuracy (AUC) to measure CRI_{cSLE}. For example, using the algorithm that considered absolute changes of the cSLE-CRVs, a logit score 0.16 or, after transformation to a scale

between 0 and 100, a CHILI score 54 was 89.5% sensitive and 92.6% specific for capturing CRI_{CSLE} status correctly (AUC= 0.97) in the development-dataset. When considering percentage changes of the cSLE-CRVs between visits in the algorithm instead, a CHILI score of 60 had similar measurement properties (AUC=0.96, sensitivity=87.6%, specificity=92.6) for capturing CRI_{CSLE} status (Figure 2, **panel a**).

Initial validation of the CHILI algorithms

Algorithms considering absolute changes rather than percentage changes of the cSLE-CRVs were similarly robust, i.e. they maintained their accuracy (AUC) similarly well the validation-dataset. Using the model parameters and threshold scores obtained from the training-dataset, the AUC of discrimination between patients who had CRI_{cSLE} as compared to those who did not was 0.93 (Figure 2, **panel a**). Hence, a CHILI score of 54 (absolute changes of the cSLE-CRVs are considered) represents the optimal threshold score based on the training-dataset. This CHILI score of 54 is 81.1% sensitive and 84.2% specific for CRI_{cSLE} in the validation-dataset.

Use of the CHILI to identify minor, moderate and major response to cSLE therapy

As is summarized in Table 4 and in Figure 2, **Panels b-d**, the CHILI algorithms developed and validated to measure CRI_{CSLE} were also excellent in discriminating patients with various levels of improvement (minor, moderate, major) between baseline and follow-up. Again, algorithms considering absolute differences and percentage differences of the cSLE-CRVs between baseline and follow-up performed similarly well in both the trainings dataset and the validation dataset.

Ranking of the candidate CHILI algorithms

The results of the performance of the PRINTO/ACR Provisional Criteria for Response to Therapy, CHILI algorithms considering percentage and CHILI algorithms considering absolute changes of the cSLE-CRVs were presented to the Consensus Conference participants with voting rights. There was consensus (100%) that the CHILI algorithms were preferable to the PRINTO/ACR Provisional Criteria for Response to Therapy to measure CRIcSLE as well as various levels of improvement in clinical trials of cSLE. Further, CHILI algorithms using absolute changes were favored over those using percentage changes of the cSLE-CRVs, given their ease of use. Although scaling to a range between 0 −100 was favored, there were some concerns about that transformation might be mathematically challenging.

DISCUSSION

This international study investigated clinically important improvement of children with cSLE. In addition to a consensus definition of CRIcSLE, we developed and initially validated the CHILI to serve as provisional criteria to measure CRIcSLE. A composite measure to capture CRIcSLE is necessary because there is no single sign, clinical test, or patient symptom that is adequately sensitive and simultaneously specific to the presence of CRIcSLE. Further, we confirm that the CHILI is able to accurately describe the degree of cSLE improvement.

Several pediatric rheumatology response measures, such as the JIA-ACR30 Criteria to capture response to therapy with JIA, consider relative – or percentage – changes of core response variables. While CHILI algorithms using percentage and absolute changes performed similarly in terms of accuracy, sensitivity and specificity at the proposed threshold scores, we consider CHILI scores calculated from absolute changes of the cSLE-CRVs to be easier to compute, hence preferable. This is in keeping with the recently published ACR Provisional Criteria of Global Flare of cSLE (7). Indeed more complex mathematical maneuvers beyond addition and multiplication are avoided which is different from the DAS28 score which includes a square root calculation, for example (28). For reasons of scaling, we transformed the CHILI scores to range between 0 – 100, with higher scores reflecting a larger degree of improvement. Whether such mathematical transformation maneuvers improve the ease of use of the CHILI will need to be studied in the future.

Different from the ACR Provisional Criteria of Global Flare of cSLE (7), the CHILI considers patient perspective more comprehensively, specifically changes in patient overall well-being and physical function (CHQ-PhS) are included in the algorithm. This is in keeping with the results of earlier discussions of the how to capture response to therapy in cSLE (6, 8).

Currently, the Systemic Lupus Responder Index is the principal outcome measure used in clinical trials of adults with SLE. We have shown that the PRINTO/ACR Provisional Criteria for Response to Therapy in cSLE (9) are more accurate than the Systemic Lupus Responder Index in capturing improvement with cSLE (10). In this study, we confirmed that the PRINTO/ACR Provisional Criteria of Response to Therapy seems to have at best fair accuracy for capturing the true course of cSLE, including CRIcSLE. Different from the CHILI, in the PRINTO/ACR Provisional Criteria for Response to Therapy, all cSLE-CRVs changes are considered equally (same percentage changes) relevant for measuring response to therapy. However, from a measurement point of view, supported by the consensus definition for CRIcSLE, and our univariate analysis, the cSLE-CRVs have differential importance to clinicians when judging the disease course of a child with cSLE (10). Taken together, the PRINTO/ACR Provisional Criteria of Response to Therapy – and by extension the Systemic Lupus Responder Index – can be used in clinical trials of cSLE but likely require larger sample sizes than when using the CHILI to capture response to therapy.

A limitation of our study might be that we were unable to test whether consideration of the British Isles Lupus Activity Group index (BILAG) index (30) or other disease activity indices instead of the SLEDAI as a measure of cSLE activity, would have allowed us to identify cSLE patients who experienced CRIcSLE accurately. Indeed, the cSLE-CRVs do not specify which validated measures of cSLE activity is considered for the assessment of patients' response to therapy (9). We used the SLEDAI, given its ease of use and widespread acceptance around the world. Additional research will be required to assess whether other disease activity index scores can be used interchangeably. Further, we did not provide patient-profile raters with consensus definitions of what constitutes minor, moderate or major improvement. Nonetheless, the accuracy of the CHILI algorithm performed well in the datasets used in this study. Lastly, we focused on the Majority Rule to adjudicate the

disease course presented in the various patient-profiles, which might have introduced bias. However, the 67% Rule yielded comparable results for the CHILI

The ACR has outlined a series of validation steps necessary before new criteria are to be widely used for clinical care or research (15, 31). One step is to use data from clinical trials for developing response criteria. However, such data from interventions that impact cSLE activity presently are unavailable. Thus, we used the patient-profiles raters' perception of the course of cSLE instead. Given the prospective character of our data and the expertise of the PP-raters, we consider the quality of the training-dataset and the validation-dataset to be high; and the number of PPs to assess CRIcSLE yielded a robust CHILI.

In summary, a methodologically stringent process has been employed to develop a novel index to measure global improvement or response to therapy in cSLE. This Provisional CHILI instrument can be used to help identify children with cSLE who have experienced a clinically relevant improvement and to categorize the degree of improvement as minor, moderate, or major. However, additional testing in independent data-sets is required to confirm the performance characteristics of the CHILI when used in cSLE.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

CCHMC: Kasha Wiley (overall study coordination), Susan Priest (consensus conference logistics), Carly Muller, Malea Rolfsen, Allen Watts, Gaurav Gulati and Jamie Meyers-Eaton (patient profile testing); CCHMC Biomedical Informatics (Web-based data management application development). We are indebted to the members of the External Scientific Advisory Committee of this study for their advice in the study implementation, conduction and its statistical analysis: Drs. Tuhina Neogi, Ian Bruce, David Isenberg, Nicola Ruperto and James Witter.

Grant Support: The study is supported by NIH grants 5U01-AR51868, P30-AR AR47363 and 2UL1RR026314 and the PRCSG and PRINTO Coordinating Centers. This study is also supported by grants from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP 215/03756–4 to CAS), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq 303422/2015–7 to CAS; CNPq 304255/2015–7 and 401477/2016–9 to SA) and by Núcleo de Apoio à Pesquisa "Saúde da Criança e do Adolescente" da USP (NAP-CriAd) to CAS.

REFEERENCES

1. Silva CA, Avcin T, Brunner HI. Taxonomy for systemic lupus erythematosus with onset before adulthood. Arthrit Care Res. 2012;64(12):1787–93
2. Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum. 1997;40(9):1725
3. Brunner HI, Gladman DD, Ibanez D, Urowitz MD, Silverman ED. Difference in disease features between childhood-onset and adult-onset systemic lupus erythematosus. Arthritis Rheum. 2008;58(2):556–62 [PubMed: 18240232]
4. Hiraki LT, Benseler SM, Tyrrell PN, Hebert D, Harvey E, Silverman ED. Clinical and laboratory characteristics and long-term outcome of pediatric systemic lupus erythematosus: a longitudinal study. The Journal of pediatrics. 2008;152(4):550–6 [PubMed: 18346514]
5. Ruperto N, Ravelli A, Cuttica R, Espada G, Ozen S, Porras O, et al. The Pediatric Rheumatology International Trials Organization criteria for the evaluation of response to therapy in juvenile systemic lupus erythematosus: prospective validation of the disease activity core set. Arthritis Rheum. 2005;52(9):2854–64 [PubMed: 16142708]

6. Ruperto N, Ravelli A, Murray KJ, Lovell DJ, Andersson-Gare B, Feldman BM, et al. Preliminary core sets of measures for disease activity and damage assessment in juvenile systemic lupus erythematosus and juvenile dermatomyositis. Rheumatology (Oxford). 2003;42(12):1452–9 [PubMed: 12832713]
7. Brunner HI, Holland M, Beresford MW, Ardoin SP, Appenzeller S, Silva CA, et al. American College of Rheumatology Provisional Criteria for Global Flares in Childhood-Onset Systemic Lupus Erythematosus. Arthritis Care Res (Hoboken). 2018;70(6):813–22 [PubMed: 29693328]
8. Brunner HI, Klein-Gitelman MS, Higgins GC, Lapidus SK, Levy DM, Eberhard A, et al. Toward the development of criteria for global flares in juvenile systemic lupus erythematosus. Arthritis Care Res (Hoboken). 2010;62(6):811–20 [PubMed: 20535792]
9. Ruperto N, Ravelli A, Oliveira S, Alessio M, Mihaylova D, Pasic S, et al. The Pediatric Rheumatology International Trials Organization/American College of Rheumatology provisional criteria for the evaluation of response to therapy in juvenile systemic lupus erythematosus: prospective validation of the definition of improvement. Arthritis Rheum. 2006;55(3):355–63 [PubMed: 16739203]
10. Mina R, Klein-Gitelman MS, Nelson S, Eberhard BA, Higgins G, Singer NG, et al. Validation of the systemic lupus erythematosus responder index for use in juvenile-onset systemic lupus erythematosus. Ann Rheum Dis. 2014;73(2):401–6 [PubMed: 23345596]
11. U.S. Food and Drug Administration. Guidance for Inudstry: Clinical Development Programs for Drugs, Devices, and Biological Products for the Treatment of Rheumatoid Arthritis (RA) In: Center for Drug Evaluation and Research (CDER) CfBEaRC, Center for Devices and Radiologic Health (CDRH), editor.; 1999.
12. European Medicines Agency, Committee for Medicinal Products for Human Use Guideline on pharmaceutical development of medicines for paediatric use. In: Agency EM, editor.; 2013.
13. Wallace CA, Ravelli A, Huang B, Giannini EH. Preliminary validation of clinical remission criteria using the OMERACT filter for select categories of juvenile idiopathic arthritis. J Rheumatol. 2006;33(4):789–95 [PubMed: 16482643]
14. Giannini EH, Ruperto N, Ravelli A, Lovell DJ, Felson DT, Martini A. Preliminary definition of improvement in juvenile arthritis. Arthritis Rheum. 1997;40(7):1202–9 [PubMed: 9214419]
15. Singh JA, Solomon DH, Dougados M, Felson D, Hawker G, Katz P, et al. Development of classification and response criteria for rheumatic diseases. Arthritis Rheum. 2006;55(3):348–52 [PubMed: 16739201]
16. Dillman DA, Smyth JD, Christian LM. Internet, phone, mail, and mixed-mode surveys : the tailored design method. 4th edition ed. Hoboken: Wiley; 2014.
17. Mina R, Harris JG, Klein-Gitelman MS, Appenzeller S, Centeville M, Eskra D, et al. Initial Benchmarking of the Quality of Medical Care in Childhood-Onset Systemic Lupus Erythematosus. Arthrit Care Res. 2016;68(2):179–86
18. Little RJ. Pattern-mixture models for multivariate incomplete data. Journal of the American Statistical Association. 1993;88(421):125–34
19. Schafer JL. Multiple imputation: a primer. Statistical methods in medical research. 1999;8(1):3–15 [PubMed: 10347857]
20. Rubin DB. Multiple imputation for nonresponse in surveys. Hoboken, N.J.;: Wiley-Interscience; 2004.
21. Gladman DD, Ibanez D, Urowitz MB. Systemic lupus erythematosus disease activity index 2000. J Rheumatol. 2002;29(2):288–91 [PubMed: 11838846]
22. Ruperto N, Ravelli A, Pistorio A, Ferriani V, Calvo I, Ganser G, et al. The provisional Paediatric Rheumatology International Trials Organisation/American College of Rheumatology/European League Against Rheumatism Disease activity core set for the evaluation of response to therapy in juvenile dermatomyositis: a prospective validation study. Arthritis Rheum. 2008;59(1):4–13 [PubMed: 18163404]
23. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143(1):29–36 [PubMed: 7063747]

24. Rider LG, Giannini EH, Harris-Love M, Joe G, Isenberg D, Pilkington C, et al. Defining Clinical Improvement in Adult and Juvenile Myositis. J Rheumatol. 2003;30(3):603–17 [PubMed: 12610824]

25. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for Outcome Measures in Rheumatology. J Rheumatol. 1998;25(2):198–9 [PubMed: 9489805]

26. Boers M, Kirwan JR, Gossec L, Conaghan PG, D'Agostino MA, Bingham CO 3rd, et al. How to choose core outcome measurement sets for clinical trials: OMERACT 11 approves filter 2.0. J Rheumatol. 2014;41(5):1025–30 [PubMed: 24584913]

27. Boers M, Kirwan JR, Wells G, Beaton D, Gossec L, d'Agostino MA, et al. Developing core outcome measurement sets for clinical trials: OMERACT filter 2.0. J Clin Epidemiol. 2014;67(7): 745–53 [PubMed: 24582946]

28. Balsa A, Carmona L, Gonzalez-Alvaro I, Belmonte MA, Tena X, Sanmarti R, et al. Value of Disease Activity Score 28 (DAS28) and DAS28–3 compared to American College of Rheumatology-defined remission in rheumatoid arthritis. J Rheumatol. 2004;31(1):40–6 [PubMed: 14705217]

29. Brunner HI, Mina R, Pilkington C, Beresford MW, Reiff A, Levy DM, et al. Preliminary criteria for global flares in childhood-onset systemic lupus erythematosus. Arthritis Care Res (Hoboken). 2011;63(9):1213–23 [PubMed: 21618452]

30. Isenberg DA, Rahman A, Allen E, Farewell V, Akil M, Bruce IN, et al. BILAG 2004. Development and initial validation of an updated version of the British Isles Lupus Assessment Group's disease activity index for patients with systemic lupus erythematosus. Rheumatology (Oxford). 2005;44(7):902–6 [PubMed: 15814577]

31. Johnson SR, Grayson PC. Use of "Provisional" Designation for Response Criteria. Arthritis Care Res (Hoboken). 2018

SIGNIFICANCE & INNOVATION

- International consensus regarding a definition of clinically relevant improvement of children and adolescents with lupus has been achieved.
- The PRINTO/ACR Provisional Criteria for Response to Therapy of children with lupus have only fair accuracy for capturing clinically relevant improvement of children with lupus as judged by physicians.
- Using strategies for the development of response measures in line with those suggested by the ACR, we newly developed and initially validated highly accurate criteria to measure clinically relevant improvement of children and adolescents with lupus.

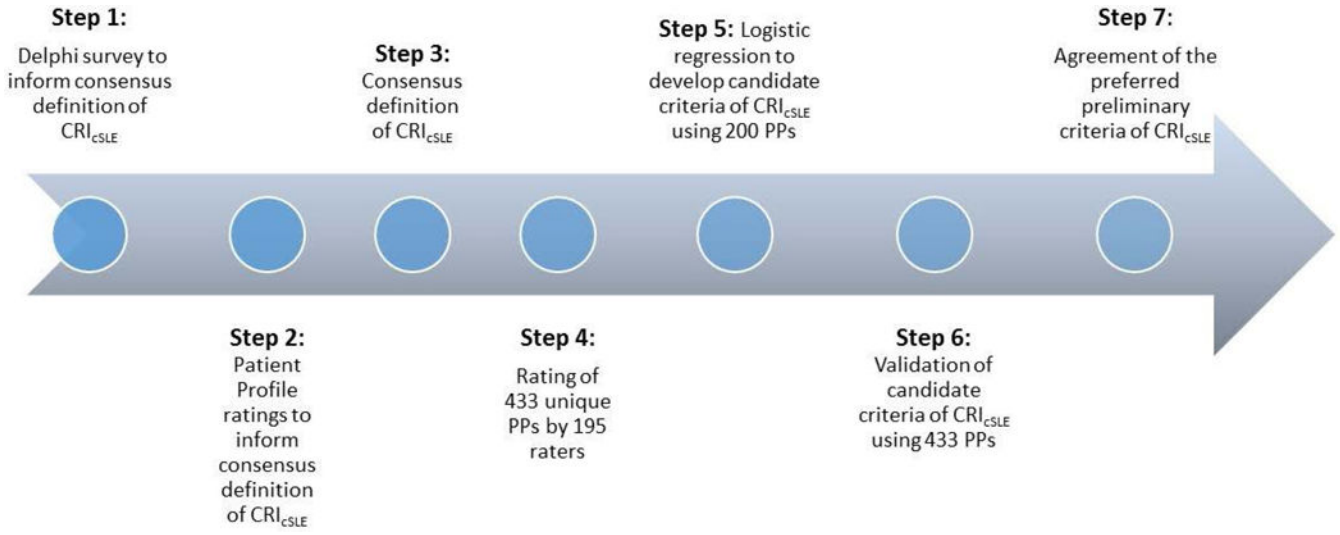

Figure 1.

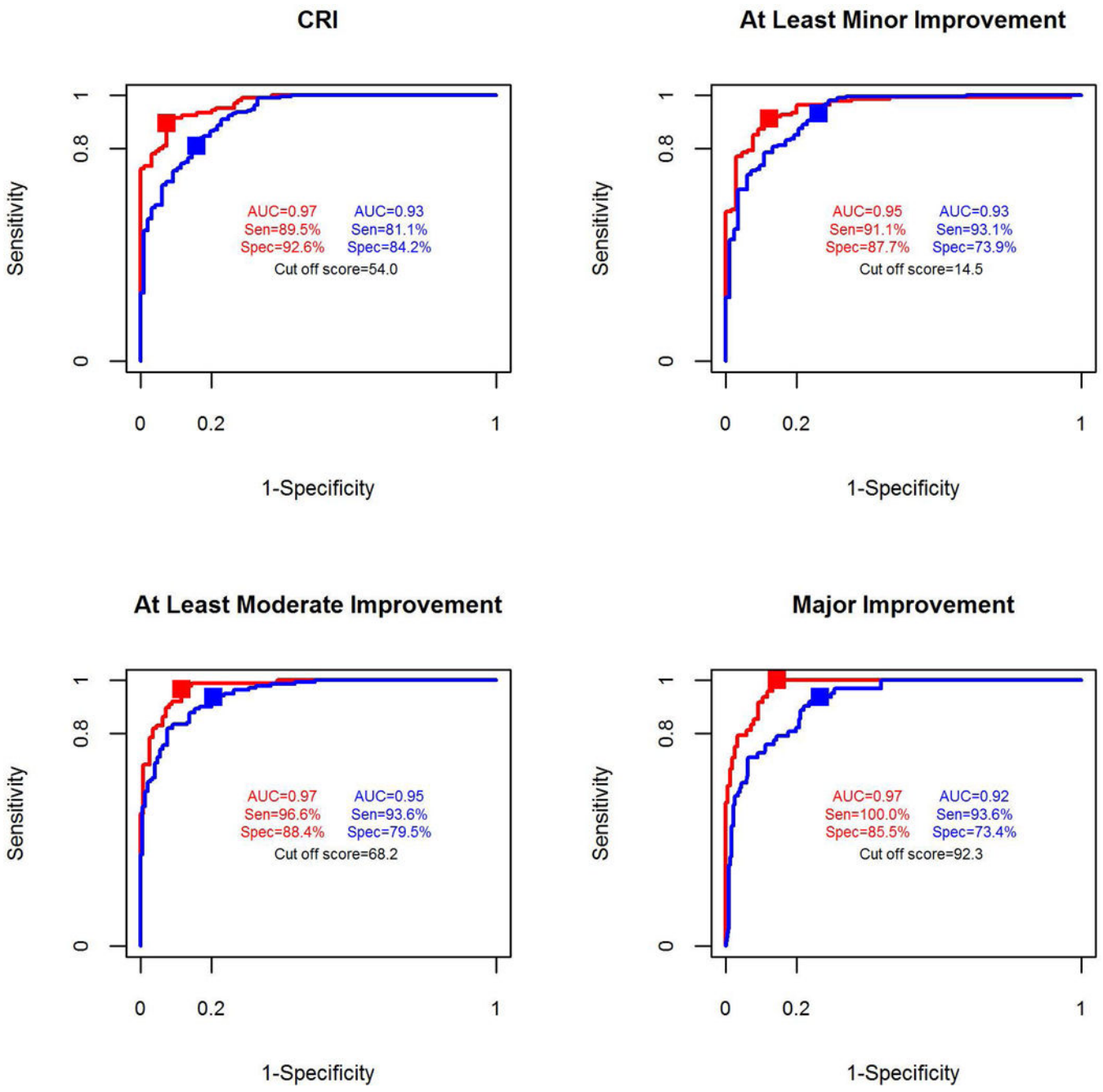

Figure 2.

Table1:

Description of 433 Patient-Profiles used in Step 4 using the Majority Rule

Patient Profile details[¶]	Training-dataset (N=200)		Validation-dataset (M=233)	
	Baseline visit	Follow-up visit	Baseline visit	Follow-up visit
SLEDAI items [*]				
Seizure	3 (1.5%)	0 (0.0%)	4 (1.7%)	0 (0.0%)
Psychosis	4 (2.0%)	0 (0.0%)	5 (2.2%)	0 (0.0%)
Organic brain syndrome	7 (3.5%)	1 (0.5%)	9 (3.9%)	2 (0.9%)
Visual Disturbance	4 (2.0%)	1 (0.5%)	4 (1.7%)	1 (0.4%)
Cranial nerve involvement	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Lupus headaches	12 (6.0%)	2 (1.0%)	15 (6.4%)	2 (0.9%)
Cardiovascular accidents	1 (0.5%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Vasculitis	27 (13.5%)	6 (3.0%)	25 (10.7%)	6 (2.6%)
Arthritis	80 (40.0%)	22 (11.0%)	96 (41.2%)	19 (8.2%)
Myositis	4 (2.0%)	2 (1.0%)	9 (3.9%)	4 (1.7%)
Urinary casts	30 (15.0%)	9 (4.5%)	38 (16.3%)	7 (3.0%)
Hematuria	71 (35.5%)	31 (15.5%)	100 (42.9%)	30 (12.9%)
Proteinuria	90 (45.0%)	49 (24.5%)	82 (35.2%)	52 (22.3%)
Leukocyturia	44 (22.0%)	20 (10.0%)	66 (28.3%)	20 (8.6%)
Rash	81 (40.5%)	26 (13.0%)	100 (42.9%)	27 (11.6%)
Alopecia	42 (21.0%)	13 (6.5%)	50 (21.5%)	15 (6.4%)
Mucosal ulcers	42 (21.0%)	9 (4.5%)	50 (21.5%)	12 (5.2%)
Pleurisy	8 (4.0%)	2 (1.0%)	15 (6.4%)	4 (1.7%)
Pericarditis	8 (4.0%)	1 (0.5%)	14 (6.0%)	0 (0.0%)
Low complement levels	154 (77.0%)	116 (58.0%)	174 (74.7%)	128 (54.9%)
Positive anti-dsDNA antibodies	155 (77.5%)	109 (54.5%)	175 (75.1%)	119 (51.1%)
Fever	42 (21.0%)	5 (2.5%)	46 (19.7%)	3 (1.3%)
Thrombocytosis	12 (6.0%)	2 (1.0%)	13 (5.6%)	4 (1.7%)
Leukopenia	27 (13.5%)	7 (3.5%)	39 (16.7%)	8 (3.4%)
SLEDAI summary score [**]	14.0 ± 8.5 / 13.0 (2.0, 39.0)	5.9 ± 5.1 / 4.0 (0.0, 31.0)	14.2 ± 8.1 / 13.0 (0.0, 39.0)	5.3 ± 4.7 / 4.0 (0.0, 31.0)
Laboratory testing [**]				
ESR	48.8 ± 35.2 / 40.0 (1.0, 180)	25.5 ± 18.9 / 21.0 (2.0, 103)	47.6 ± 37.8 / 40.0 (1.0, 180)	24.3 ± 17.2 / 21.0 (1.0, 101)
UPCR	1.3 ± 2.2 / 0.3 (0.0, 13.2)	0.5 ± 1.1 / 0.2 (0.0, 7.8)	1.2 ± 2.3 / 0.2 (0.0, 13.2)	0.5 ± 1.2 / 0.2 (0.0, 7.8)
Other assessments				
MD-global	4.2 ± 2.9 / 4.1 (0, 10)	1.7 ± 2.1 / 0.8 (0, 10)	5.0 ± 2.6 / 5.0 (0, 10)	1.8 ± 1.8 / 1.1 (0, 8.6)
Patient-global	3.0 ± 3.0 / 1.9 (0, 10)	1.4 ± 2.0 / 0.6 (0, 10)	4.8 ± 3.3 / 5.0 (0, 10)	3.3 ± 3.5 / 1.7 (0, 10)
CHQ-PhS	36.8 ± 15.5 / 40.4 (1.0, 58.7)	45.4 ± 11.5 / 49.4 (5.5, 59.7)	36.9 ± 14.7 / 40.3 (1.0, 58.7)	44.6 ± 11.7 / 48.4 (10.3, 59.7)

[*] Values are n (% N) or

** mean+SD / Median (Min, Max);

¶ MD-global: physician assessment of cSLE Activity as measured on a visual analog scale (VAS), 0 = inactive disease, 10 = very active disease; Patient-global: parent assessment of patient overall well-being, measured on a VAS (Patient-global;0 = very poor; 10 = very well; UPCR: proteinuria, measured by timed urine collection or spot protein to creatinine ratio; ESR: erythrocyte sedimentation rate; CHQ-PhS: Child Health Questionnaire (parent version P50) Physical Function Summary score

Table 2:

Discriminative properties of absolute and relative, i.e. percentage changes of the cSLE core response variable for capturing CRI_{CSLE} in the training-dataset

Core Response variables‡		(1) CRI_{cSLE} NO†	(2) CRI_{cSLE} YES†	p-values (1) vs. (2)	AUC	Sensitivity*	Specificity*	Threshold score∫
	UPCR	−0.21 ± 0.18	−1.16 ± 0.17	<0.0001	0.65	33.33%	90.53%	−0.81
	SLEDAI	−2.71 ± 0.57	−13.01 ± 0.54	<0.0001	0.93	94.29%	82.11%	−5.00
Absolute difference	MD-global	−0.65 ± 0.21	−4.23 ± 0.20	<0.0001	0.90	94.29%	67.37%	−1.00
	Patient global	−0.06 ± 0.31	−2.89 ± 0.29	<0.0001	0.76	70.48%	78.95%	−0.80
	CHQ-PHS	2.01 ± 1.32	14.55 ± 1.26	<0.0001	0.77	70.48%	81.05%	5.15
	UPCR	33% ± 30%	−30% ± 29%	0.132	0.67	60.95%	69.47%	−0.41
	SLEDAI	−24% ± 3%	−72% ± 3%	<0.0001	0.91	84.76%	89.47%	−0.52
Percentage difference	MD-global	−10% ± 3%	−63% ± 3%	<0.0001	0.91	84.76%	86.32%	−0.47
	Patient global	31% ± 12%	−33% ± 12%	<0.0001	0.77	72.38%	77.89%	−0.23
	CHQ-PHS	20% ± 28%	151% ± 27%	0.001	0.76	69.52%	81.05%	0.13

†Values are mean ± SD from absolute differences between the baseline and follow-up time point OR percentage changes at the time of follow-up relative of the baseline visit; AUC: area under the receiver operating characteristic curve;

*sensitivity and specificity at the threshold score

‡UPCR: urine protein to creatinine ratio from post urine or estimated by 24 hour timed urine collection;

SLEDAI: SLE disease activity summary score; MD-global: physician global assessment of cSLE activity; Patient-global: patient assessment of overall well-being; CHQ-PHS Child Health Questionnaire, P50 version physical function summary score

∫Optimal score from univariate logistic regression to discriminate between the presence vs. absence of CRI_{CSLE}

Table 3:

Performance of the ACR/PRINTO Criteria for cSLE Improvement as published by Ruperto et al 2006[†]

Reference standard (vs. no change)	PRINTO/ACR/EULAR Provisional Criteria of Response to Therapy [**]	Development dataset (N=200)			Validation dataset (N=233)		
		Sensitivity	*Specificity*	*Kappa ± SE*[*]	*Sensitivity*	*Specificity*	*Kappa ± SE*[*]
CRI_{CSLE}	2/5 by ≥50% and no more than 1 worse by > 30% (DI11)	91.4%	37.9%	0.30 ± 0.06	84.7%	50.6%	0.37 ± 0.06
	2/5 by ≥40% and no more than 1 worse by > 30% (DI08)	91.4%	37.9%	0.30 ± 0.06	84.7%	50.6%	0.37 ± 0.06
	2/5 by ≥40% and no more than 1 worse by > 30% (DI17)	91.4%	37.9%	0.30 ± 0.06	84.7%	50.6%	0.37 ± 0.06
	3/5 by ≥30% and no more than 2 worse by > 30% (DI12)	100.0%	11.6%	0.12 ± 0.03	97.3%	15.3%	0.15 ± 0.05
At least minor improvement	2/5 by ≥50% and no more than 1 worse by > 30% (DI11)	89.6%	47.7%	0.41 ± 0.07	81.8%	48.7%	0.32 ± 0.07
	2/5 by ≥40% and no more than 1 worse by > 30% (DI08)	89.6%	47.7%	0.41 ± 0.07	81.8%	48.7%	0.32 ± 0.07
	2/5 by ≥40% and no more than 1 worse by > 30% (DI17)	89.6%	47.7%	0.41 ± 0.07	81.8%	48.7%	0.32 ± 0.07
	3/5 by ≥30% and no more than 2 worse by > 30% (DI12)	99.3%	15.4%	0.19 ± 0.06	97.5%	17.1%	0.18 ± 0.05
At least moderate improvement	2/5 by ≥50% and no more than 1 worse by > 30% (DI11)	90.9%	33.0%	0.22 ± 0.05	87.5%	40.5%	0.26 ± 0.05
	2/5 by ≥40% and no more than 1 worse by > 30% (DI08)	90.9%	33.0%	0.22 ± 0.05	87.5%	40.5%	0.26 ± 0.05
	2/5 by ≥40% and no more than 1 worse by > 30% (DI17)	90.9%	33.0%	0.22 ± 0.05	87.5%	40.5%	0.26 ± 0.05
	3/5 by ≥30% and no more than 2 worse by > 30% (DI12)	100.0%	9.8%	0.09 ± 0.03	99.0%	12.2%	0.10 ± 0.03
Major Improvement	2/5 by ≥50% and no more than 1 worse by > 30% (DI11)	93.8%	27.6%	0.12 ± 0.03	91.9%	31.8%	0.10 ± 0.03
	2/5 by ≥40% and no more than 1 worse by > 30% (DI08)	93.8%	27.6%	0.12 ± 0.03	91.9%	31.8%	0.10 ± 0.03
	2/5 by ≥40% and no more than 1 worse by > 30% (DI17)	93.8%	27.6%	0.12 ± 0.03	91.9%	31.8%	0.10 ± 0.03
	3/5 by ≥30% and no more than 2 worse by > 30% (DI12)	100.0%	7.2%	0.04 ± 0.01	100.0%	8.6%	0.03 ± 0.01

[*] kappa provides agreement between Provisional PRINTO/ ACR/EULAR Criteria of Response to Therapy and PP ratings from Step 2 adjudicated by the Majority Rule;

[**] the four highest ranking Provisional PRINTO/ ACR/EULAR Criteria of Response to Therapy algorithms are shown (DI11, DI08, DI17, DI12) and ratio designates (n/m) the number of the five cSLE-CRVs (MD-global, Patient-global, CHQ-PhS, disease activity score, UPCR) with improvement.

[†] Ruperto N, Ravelli A, Oliveira S, et al. The Pediatric Rheumatology International Trials Organization/American College of Rheumatology provisional criteria for the evaluation of response to therapy in juvenile systemic lupus erythematosus: prospective validation of the definition of improvement. Arthritis Rheum. 2006;55(3):355–363.

Table 4:

Absolute versus relative (or percentage) changes of the combination of the cSLE core response variable for capturing CRI_{CSLE} and various levels of improvement †

Level of response to therapy or improvement	Type of CRV change considered	Training-dataset					Validation-dataset		
		AUC	*Sensitivity* *	*Specificity* *	*Threshold improvement logit score*	*Threshold improvement score (0–100)*	*AUC*	*Sensitivity*	*Specificity*
CRI_{CSLE}	Absolute**	0.97	89.5%	92.6%	**0.16**	**54**	0.93	81.1%	84.2%
	Percentage***	0.96	87.6%	92.6%	**0.39**	**60**	0.92	81.1%	87.1%
At least minor improvement	Absolute**	0.95	91.1%	87.7%	**−1.77**	**15**	0.93	93.1%	73.9%
	Percentage***	0.93	85.9%	89.2%	**−1.15**	**24**	0.93	90.2%	76.1%
At least moderate improvement	Absolute**	0.97	96.6%	88.4%	**0.76**	**68**	0.95	93.6%	79.5%
	Percentage***	0.96	88.6%	92.9%	**1.33**	**79**	0.93	85.0%	84.8%
Major improvement	Absolute**	0.97	100 %	85.5%	**2.48**	**92**	0.92	93.6%	73.4%
	Percentage***	0.95	91.7%	84.9%	**2.46**	**92**	0.94	98.4%	77.7%
CHILI‡	Using absolute changes (Δ) of cSLE-CRVs	Improvement Logit Score (y)= - (5.1 + 0.47 x ΔSLEDAI + 0.7 x ΔMD-global + 1.1x Δ UPCR + 0.32 x ΔPatient-global - 0.002x ΔCHQ-PhS) Improvement Score (0–100) = 100● {EXP(y) / [1+EXP(y)]}							
	Using percentage changes ($\Delta_{\%}$) of cSLE-CRVs	Improvement Logit Score (y) = - (6.1 + 6.5x $\Delta_{\%}$SLEDAI + 6.1X $\Delta_{\%}$MD-global + 0.02x $\Delta_{\%}$UPCR + 0.48x $\Delta_{\%}$Patient-global - 0.43x $\Delta_{\%}$CHQ-PhS) Improvement Score (0–100) = 100● {EXP(y) / [1+EXP(y)]}							

† Majority rule;

Δ change score between baseline and follow-up visits

‡ Improvement logit scores were calculated using multivariate logistic models. Predictors using in a multivariate logistic model can be either from absolute differences between the baseline and follow-up time point OR percentage changes at the time of follow-up relative of the baseline visit in the training-dataset. An improvement logit score can be converted into an improvement score (ranging 0–100) using the formula: Improvement Score = 100

* EXP(Improvement logit score) / [1+ EXP(Improvement logit score)]. Either a higher (or lower) improvement logit score or improvement score indicates higher (lower) likelihood of improvement. For a classification purpose, a patient's improvement score can be compared against the threshold score.

UPCR: urine protein to creatinine ratio from random urine sample; SLEDAI: SLE disease activity summary score; MD-global: physician global assessment of cSLE activity; Patient-global: patient assessment of overall well-being; CHQ-PHS Child Health Questionnaire, P50 version Physical function summary score