Arthritis Care & Research
Vol. 71, No. 5, May 2019, pp 579–590
DOI 10.1002/acr.23834

American College of Rheumatology
Empowering Rheumatology Professionals

American College of Rheumatology Provisional Criteria for Clinically Relevant Improvement in Children and Adolescents With Childhood-Onset Systemic Lupus Erythematosus

Hermine I. Brunner,[1] Michael J. Holland,[1] Michael W. Beresford,[2] Stacy P. Ardoin,[3] Simone Appenzeller,[4] Clovis A. Silva,[5] Francisco Flores,[1] Beatrice Goilav,[6] Pinar Ozge Avar Aydin,[1] Scott E. Wenderfer,[7] Deborah M. Levy,[8] Angelo Ravelli,[9] Raju Khubchandani,[10] Tadej Avcin,[11] Marisa S. Klein-Gitelman,[12] Nicolino Ruperto,[9] Brian M. Feldman,[8] and Jun Ying,[13] for the Paediatric Rheumatology International Trial Organisation and Pediatric Rheumatology Collaborative Study Group

Objective. To develop a Childhood Lupus Improvement Index (CHILI) as a tool to measure response to therapy in childhood-onset systemic lupus erythematosus (cSLE), with a 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 following cSLE core response variables (CRVs): global assessment of patient well-being (patient-global), physician assessment of cSLE activity (MD-global), disease activity index score (here, we used the Systemic Lupus Erythematosus Disease Activity Index), urine protein-to-creatinine ratio, and Child Health Questionnaire physical summary score. Percentage and absolute changes in these cSLE-CRVs (baseline versus follow-up) were considered in order to develop candidate algorithms and validate their performance (sensitivity, specificity, area under the receiver operating characteristic curve [AUC]; range 0–1).

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

Conclusion. The CHILI is a new, seemingly highly accurate index for measuring CRI in cSLE over time. This index is useful to categorize the degree of response to therapy in children and adolescents with cSLE.

INTRODUCTION

Systemic lupus erythematosus (SLE) is a complex, chronic, multisystem autoimmune inflammatory disease, with up to 20% of patients diagnosed during childhood (cSLE) (1,2). When lupus disease commences early in life rather than during adulthood, the prognosis is poorer, particularly due to multiorgan and kidney involvement (3,4). The course of cSLE is characterized by

Supported by the NIH (grants 5U01-AR-51868, P30-AR-AR47363, and 2UL1-RR-026314). Dr. Silva's work was supported by grants from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP 215/03756-4), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq 303422/2015-7 and CNPq 304255/2015-7), and by Núcleo de Apoio à Pesquisa "Saúde da Criança e do Adolescente" da USP (NAP-CriAd). Dr. Ardoin's work was supported by CNPq 7/2016-9.

[1]Hermine I. Brunner, MD, MSc, Michael J. Holland, MD, Francisco Flores, MD, Pinar Ozge Avar Aydin, MD: University of Cincinnati and Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio; [2]Michael W. Beresford, MBBS: Institute of Translational Medicine and Alder Hey Children's NHS Foundation Trust, Liverpool, UK; [3]Stacy P. Ardoin, MD: Ohio State University, Nationwide Children's Hospital, and Wexner Medical Center, Columbus, Ohio; [4]Simone Appenzeller, MD: University of Campinas, Campinas, Brazil; [5]Clovis A. Silva, MD: Children's Institute, Hospital das Clinicas HCFMUSP, Universidade de São Paulo, São Paulo, Brazil; [6]Beatrice Goilav, MD: Children's Hospital at Montefiore, Albert Einstein College of Medicine, Bronx, New York; [7]Scott E. Wenderfer, MD: Baylor College of Medicine and Texas Children's Hospital, Houston, Texas; [8]Deborah M. Levy, MD, Brian M. Feldman, MD: University of Toronto and Hospital for Sick Children, Toronto, Ontario, Canada; [9]Angelo Ravelli, MD, Nicolino Ruperto, MD: Clinica Pediatrica e Reumatologia, Istituto Giannina Gaslini, PRINTO, Genoa, Italy, [10]Raju Khubchandani, MD: Jaslok Hospital Mumbai, India; [11]Tadej Avcin, MD: University Children's Hospital, University Medical Centre, Ljubljana, Slovenia;

SIGNIFICANCE & INNOVATIONS

- International consensus regarding a definition of clinically relevant improvement in children and adolescents with lupus has been achieved.
- The Pediatric Rheumatology International Trials Organization/American College of Rheumatology (ACR) provisional criteria for response to therapy in children with lupus have only fair accuracy for capturing clinically relevant improvement in 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 in children and adolescents with lupus.

episodes of disease flares followed by periods of improvement, generally due to more intensive drug therapy. There is international consensus that core response variables (cSLE-CRVs) should be considered when assessing response to therapy and cSLE disease flare (5,6). Considering changes in cSLE-CRVs, a provisional American College of Rheumatology (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 in cSLE-CRVs are the basis for the Pediatric Rheumatology International Trials Organization (PRINTO)/ACR provisional criteria for response to therapy (9).

We have previously shown, albeit in a rather small data set, that the PRINTO/ACR provisional criteria for response to therapy and, to a lesser extent, the Systemic Lupus Erythematosus Responder Index (SRI) are both very well suited to capture major improvement in cSLE; however, both the PRINTO/ACR provisional criteria for response to therapy and the SRI 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 in cSLE, and consensus is lacking regarding 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% improvement (ACR20) response, or in juvenile idiopathic arthritis (JIA) an ACR30 response (JIA-ACR30), provides such a measure of clinically relevant improvement. ACR20 and JIA-ACR30 responses, respectively, are regarded as improvement thresholds that can support labeling of new medications by the US 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 to develop, as well as initially validate, criteria to measure CRI_{cSLE}. Further, we sought to measure minor, moderate, and major responses to therapy in cSLE.

PATIENTS AND METHODS

The overall approach to this project was based on the methodologic 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 shown 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 steps 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 data set was randomly split into a training data set and a validation data set (step 4). The training data set was used to develop candidate criteria for CRI_{cSLE} (step 5). These candidate criteria were tested using the validation data set (step 6). As in step 3, agreement was achieved around a

[12]Marisa S. Klein-Gitelman, MD: Northwestern University Feinberg School of Medicine and Ann and Robert Lurie Children's Hospital of Chicago, Chicago, Illinois; [13]Jun Ying, PhD: University of Cincinnati, Cincinnati, Ohio.

Dr. Brunner has received consulting fees from 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 (less than $10,000 each), speaking fees from Genentech and Novartis (more than $10,000 each), and Scientific grant support from Pfizer and Bristol-Myers Squibb for research studies in adult lupus and juvenile idiopathic arthritis. 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 Pediatric Rheumatology Collaborative Study Group network. Dr. Ravelli has received consulting and/or speaking fees from AbbVie, Bristol-Myers Squibb, Pfizer, Hoffman-LaRoche, Novartis, Centocor, "Francesco Angelini," and Reckitt Benckiser (less than $10,000 each). Dr. Avcin has received consulting and/or speaking fees from AbbVie and Boehringer Ingelheim (less than $10,000 each). Dr. Ruperto has received consulting and/or speaking fees from Abbott, AbbVie, Amgen, Biogenidec, Astellas, Alter, AstraZeneca, Baxalta Biosimilars, Boehringer, Bristol-Myers Squibb, CD Pharma, Celgene, Crescendo Bioscience, EMD Serono, Hoffman-La Roche, Italfarmaco, Janssen, MedImmune, Medac, Novartis, Novo Nordisk, Pfizer, R/Pharma, Sanofi Aventis, Servier, Sinergie, Takeda, Vertex, UCB Biosciences (less than 10,000 each). IRCCS Istituto Giannina Gaslini, the public hospital where Dr. Ruperto is a full-time public employee, has received contributions from Abbott, Bristol-Myers Squibb, "Francesco Angelini," GlaxoSmithKline, Hoffman-La Roche, Italfarmaco, Janssen, Novartis, Pfizer, Sanofi-Aventis, Schwarz Biosciences, Sobi, Xoma, and Wyeth (more than $10,000 each). No other disclosures relevant to this article were reported.

Address correspondence to Hermine I. 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. E-mail: hermine.brunner@cchmc.org.

Submitted for publication May 24, 2018; accepted in revised form January 8, 2019.

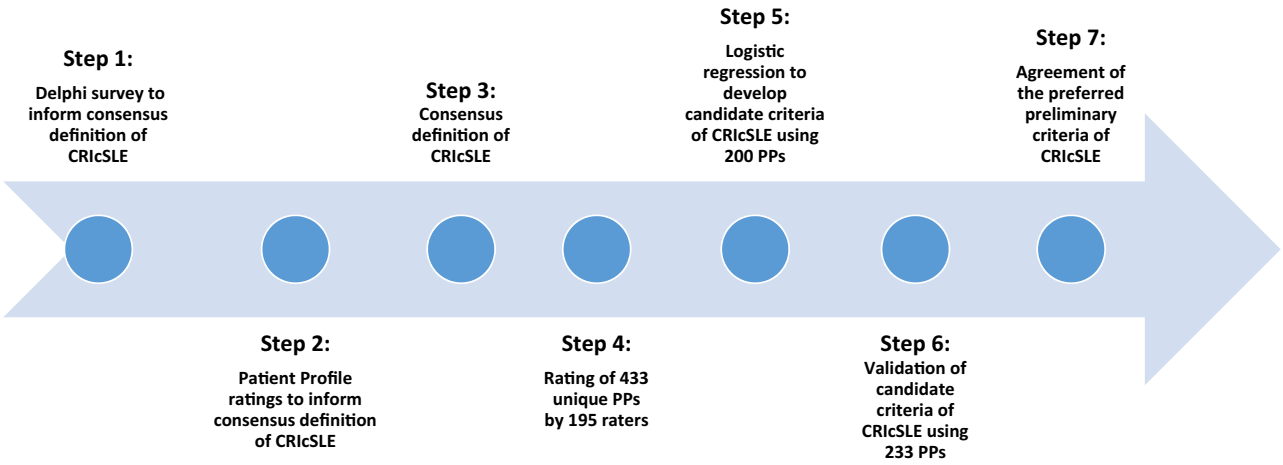

Figure 1. Process of developing the criteria for clinically relevant improvement of childhood-onset systemic lupus erythematosus (cSLE). Consensus formation methodology was used, starting with a Delphi survey, a series of online ratings of patient profiles (PPs) by physicians experienced in the care of children with cSLE, a consensus conference, and final consensus about preferred criteria to measure clinically relevant improvement of cSLE (CRI_{cSLE}).

preferred Childhood Lupus Improvement Index (CHILI) algorithm among cSLE experts with voting rights who had participated in the consensus conference (step 7). See Appendix A for a list of the members of the Pediatric Rheumatology Collaborative Study Group who contributed to steps 1 and 6.

Step 1: Delphi survey regarding CRI_{cSLE}. The 13 expert participants in the consensus conference and 100 of the pediatric rheumatologists who contributed to the development of other cSLE criteria sets (6,8) received a Delphi survey inquiring about cSLE (1) characteristics and changes in 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: patient profile ratings prior to the consensus conference. Using prospective data for cSLE patients in the Cincinnati Children's Hospital Medical Center Lupus Registry (17), the PRINTO Lupus Cohort (6), and a multicenter North American cSLE cohort (grant U01-AR-5868 to Dr. Brunner, Principal Investigator), 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 in 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 (MD-global) as measured on a visual analog scale (VAS) (0 = inactive disease and 10 = very active disease), 2) parent assessment of patient overall well-being (patient-global) as measured on a VAS (0 = very poor and 10 = very well), 3) proteinuria, measured by timed urine collection or protein-to-creatinine ratio in a spot urine specimen, 4) erythrocyte sedimentation rate (ESR), 5) levels of complement C3 and C4, 6) item and summary scores of the Systemic Lupus Erythematosus Disease Activity Index 2000 (SLEDAI-2K) (21), and 7) the Child Health Questionnaire (version P50) physical summary score (CHQ-PhS) (5,6). Information about complete blood cell counts with differential, serum chemistry, ESR, urinalysis, and anti–double-stranded DNA 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 200 pediatric rheumatologists who previously participated in a similar patient profile rating exercise (6–8) were asked to rate 50 patient profiles each that were randomly selected from the pool of 433 patient profiles. Each patient profile rater was asked to assess the disease course (Question A), using the following response options: 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 or not improvement constituted CRI_{cSLE} (Question B). In this context, minor improvement can be considered equivalent to "any improvement" in cSLE. The survey source data were batch-processed, and open source online survey software, REDCap, 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 for a given patient profile, the "true" overall course of cSLE for a

given patient profile was adjudicated using majority rule (i.e., the majority of the raters of a patient profile agreed on a given disease course). Other rules were calculated, including the 67% rule (i.e., at least two-thirds 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 used to develop a series of candidate criteria to measure CRI_{cSLE}. We considered the PRINTO/ACR provisional criteria for response to therapy (9), which have been previously validated to measure improvement in cSLE (strategy 1). Furthermore, we developed algorithms that considered absolute change (strategy 2) or relative or percentage changes (strategy 3) in the cSLE-CRVs between baseline and follow-up, using multinomial logistic regression. Strategies 2 and 3 yield a numeric "CHILI score" (or log odds of improvement) calculated from the combined changes in the cSLE-CRV predictors between baseline and follow-up (9,22).

The accuracy of the PRINTO/ACR provisional criteria for response to therapy was tested using kappa statistics. With respect to the criterion standard (here, adjudicated disease course from the patient profile ratings), kappa values can be interpreted as follows: poor agreement, $\kappa < 0.4$; fair to good agreement, $\kappa = 0.4$–0.75; and 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 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, or 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, or major).

All analyses were conducted using SAS 9.4 and Systat 12 software. *P* values less than 0.05 were considered significant.

Step 3: consensus definition of CRI_{cSLE}. Participants in the consensus conference were experienced pediatric rheumatologists and nephrologists from South America, North America, Asia, and Europe who had 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 in previous similar studies (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 patient profile ratings from step 2, considering the Outcome Measures in Rheumatology 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 2 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 have 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 steps 1 and 2, the 433 patient profiles were then sent to another 100 pediatric rheumatologists who previously participated in a similar study (7). Therefore, a total of 213 patient profile raters received 50 randomly selected patient profiles each; formats, response options, and adjudication were described in step 2. The resulting data set was divided, in the sequence of acquisition, into a training data set and a validation data set.

Steps 5 and 6: development and preliminary validation of the CHILI. Using the training data set (step 4), we newly developed candidate algorithms to measure improvement (CRI_{cSLE}; minor, moderate, or 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 data set (step 5) were validated using the validation data set to derive 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 who had voting rights. These cSLE experts were asked whether, in the setting of a clinical trial, 1) use of CRI_{cSLE} algorithms from multinomial logistic regression was preferable to the use of the PRINTO/ACR provisional criteria for response to therapy 2) absolute differences in 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, or major) in cSLE.

Table 1. Description of 433 patient profiles used in step 4 using the majority rule*

	Training data set (n = 200)		Validation data set (n = 233)	
Patient profile details	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 ± .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)

* SLEDAI = Systemic Lupus Erythematosus Disease Activity Index; anti-dsDNA = anti–double-stranded DNA; ESR = erythrocyte sedimentation rate; UPCR = proteinuria, measured by timed urine collection or protein-to-creatinine ratio in a spot urine specimen; MD-global = physician assessment of childhood-onset SLE as measured on a visual analog scale (VAS) (0 = inactive disease, 10 = very active disease); patient-global = global assessment of patient well-being, measured on a VAS (0 = very poor; 10 = very well); CHQ-PhS: Child Health Questionnaire (parent version P50) physical function summary score

† Values are the number (%).

‡ Values are the mean ± SD/median (minimum, maximum).

RESULTS

Definition of CRI_{cSLE}. The survey (step 1) inquired about changes in cSLE-CRVs, signs, and symptoms that would support the presence of CRI_{cSLE}. Among the 113 pediatric rheumatologists and nephrologists approached for survey participation, 92 (81%) responded. Survey participants from different regions or less versus 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 for a disease activity index must be better or unchanged and that patients with CRI_{cSLE} could experience new organ involvement as long as it did not involve the neuropsychiatric, hematologic, gastrointestinal, renal, ophthalmologic, or cardiopulmonary organ system. The initial ratings of 200 patient profiles provided additional data regarding the measurement of CRI_{cSLE} (see Supplementary Table 1 [available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23834/abstract] 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 system. 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 for response to therapy in cSLE was warranted, and that multinomial logistic regression modeling should be pursued to measure CRI_{cSLE} with threshold choice at the statistical optimal point on the AUC (≥92% agreement for both).

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 of 213), and all 433 patient profiles qualified for adjudication. The resulting data set was split into a training data set (200 patient profiles) and a validation data set (n = 233). Baseline characteristics of the patients represented in these data sets are shown in Table 1. When the majority rule was applied, there were 95 (47.5%) patient profiles without CRI_{cSLE} and 105 patient profiles (52.5%) with CRI_{cSLE} in the training data set. 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 in cSLE.

Performance of individual cSLE-CRVs to measure CRI_{cSLE}. Based on univariate logistic regression in the training data set, absolute changes and percentage (or relative) changes in the cSLE-CRVs had similar discriminative properties to detect CRI_{cSLE} (Table 2). However, only absolute changes in the urine protein-to-creatinine ratio ($P < 0.001$) between baseline and follow-up but not percentage changes

Table 2. Discriminative properties of absolute and relative (percentage) changes in the cSLE core response variable for capturing CRI_{cSLE} in the training data set*

Core response variables	CRI_{cSLE}			AUC	Sensitivity†	Specificity†	Threshold score‡
	No	Yes	*P*				
Absolute difference§							
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
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
Percentage difference¶							
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
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 for sensitivity and specificity are the percent. cSLE = childhood-onset systemic lupus erythematosus; CRI_{cSLE} = clinically relevant improvement in cSLE (see Table 1 for other definitions).
† Sensitivity and specificity calculated at the threshold score.
‡ Optimal score from univariate logistic regression to discriminate between the presence versus absence of CRI_{cSLE}.
§ Values are the mean ± SD from absolute differences between the baseline and follow-up time points.
¶ Values are the percentage change at the time of the follow-up visit relative to the baseline visit.

(P = 0.132) significantly differed between patients with and those without CRI_{cSLE}. Compared with the other cSLE-CRVs but irrespective of the type of change (absolute, relative) considered, the urine protein-to-creatinine ratio had only fair accuracy (AUC ≤ 0.67) for capturing CRI_{CSLE}. Individually, the MD-global and the SLEDAI had the highest accuracy (AUC ≥ 0.90 for both) for identifying CRI_{CSLE} status.

Performance of the PRINTO/ACR provisional criteria for response to therapy to measure CRI_{cSLE}. As shown in Table 3 and Supplementary Table 2 (available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.23834/abstract), in both the training data set and validation data set, the PRINTO/ACR provisional criteria for response to therapy had, at most, fair accuracy for capturing CRI_{cSLE} status ($\kappa \leq 0.3$ for the majority rule and $\kappa \leq 0.43$ for the 67% rule). The same was also true for measuring various levels of improvement ($\kappa < 0.34$ for both the majority rule and the 67% rule).

Development of the CHILI to measure CRI_{cSLE}. As part of the step 5 analyses (Table 4), we used multinomial regression to generate candidate algorithms using the cSLE-CRVs that have

Table 3. Performance of the PRINTO/ACR/EULAR criteria for cSLE improvement*

Reference standard (vs. no change)	PRINTO/ACR/EULAR provisional criteria for response to therapy†	Development data set (n = 200)			Validation data set (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

* The table was adapted from reference 9 (Ruperto et al, 2006). Values for sensitivity and specificity are the percent. PRINTO = Pediatric Rheumatology International Trials Organization; ACR = American College of Rheumatology; EULAR = European League Against Rheumatism; cSLE = childhood-onset systemic lupus erythematosus; CRI_{cSLE} = clinically relevant improvement in cSLE (see Table 1 for other definitions).
† The 4 highest-ranking algorithms (DI11, DI08, DI17, DI12) are shown. Ratio (n/m) designates the number of the 5 cSLE core response variables CRVs (MD-global, patient-global, Disease Activity Score, urine protein-to-creatinine ratio from spot urine or estimated by 24-hour timed urine collection) with improvement.
‡ The kappa value provides agreement between the provisional PRINTO/ACR/EULAR criteria for response to therapy and patient profile ratings from step 2, adjudicated using the majority rule.

been considered relevant for capturing improvement of cSLE (6,22). Irrespective of the type of change, i.e., absolute or percentage differences in the cSLE-CRVs between baseline and follow-up, algorithms had similar accuracy (AUC) for measuring CRI_{cSLE}. For example, using the algorithm that considered absolute changes in the cSLE-CRVs, a logit score of 0.16 or, after transformation to a scale of 0 to 100, a CHILI score of 54 was 89.5% sensitive and 92.6% specific for capturing CRI_{cSLE} status correctly (AUC 0.97) in the development data set. When considering percentage changes in 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).

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

Use of the CHILI to identify minor, moderate, and major response to cSLE therapy. As shown in Figure 2 and Table 4, the CHILI algorithms developed and validated to measure CRI_{cSLE} were also excellent for discriminating patients with various levels of improvement (minor, moderate, or major) between baseline and follow-up. Again, algorithms considering absolute differences and percentage differences in the cSLE-CRVs between baseline and follow-up performed similarly well in both the training data set and the validation data set.

Ranking of the candidate CHILI algorithms. The results for performance of the PRINTO/ACR provisional criteria for response to therapy, CHILI algorithms considering percentage change, and

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	Development data set					Validation data set		
		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)]}								
CHILI‡ 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)]}								

* Predictors used in a multivariate logistic model can be either from absolute differences between the baseline and follow-up time points OR percentage changes at the time of follow-up relative to the baseline visit in the development data set. An improvement logit score can be converted into an improvement score (range 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.
† Majority rule; Δ change score between baseline and follow-up visits.
‡ Improvement logit scores were calculated using multivariate logistic models. 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.

CHILI algorithms considering absolute changes in the cSLE-CRVs were presented to the consensus conference participants who had voting rights. There was consensus (100%) that the CHILI algorithms were preferable to the PRINTO/ACR provisional criteria for response to therapy to measure CRI_{cSLE} as well as various levels of improvement in clinical trials of cSLE. Furthermore, CHILI algorithms using absolute changes were favored over those using percentage changes in the cSLE-CRVs, given their ease of use. Although scaling to a range of 0 to 100 was favored, there were some concerns that transformation might be mathematically challenging.

DISCUSSION

This international study investigated clinically important improvement in children with cSLE. In addition to a consensus definition of CRI_{cSLE}, we developed and initially validated the CHILI to serve as provisional criteria to measure CRI_{cSLE}. A composite measure to capture CRI_{cSLE} is necessary, because there is no single sign, clinical test, or patient symptom that is adequately sensitive and simultaneously specific for the presence of CRI_{cSLE}. Furthermore, 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 in 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 in the cSLE-CRVs to be easier to compute, hence preferable. This is consistent with

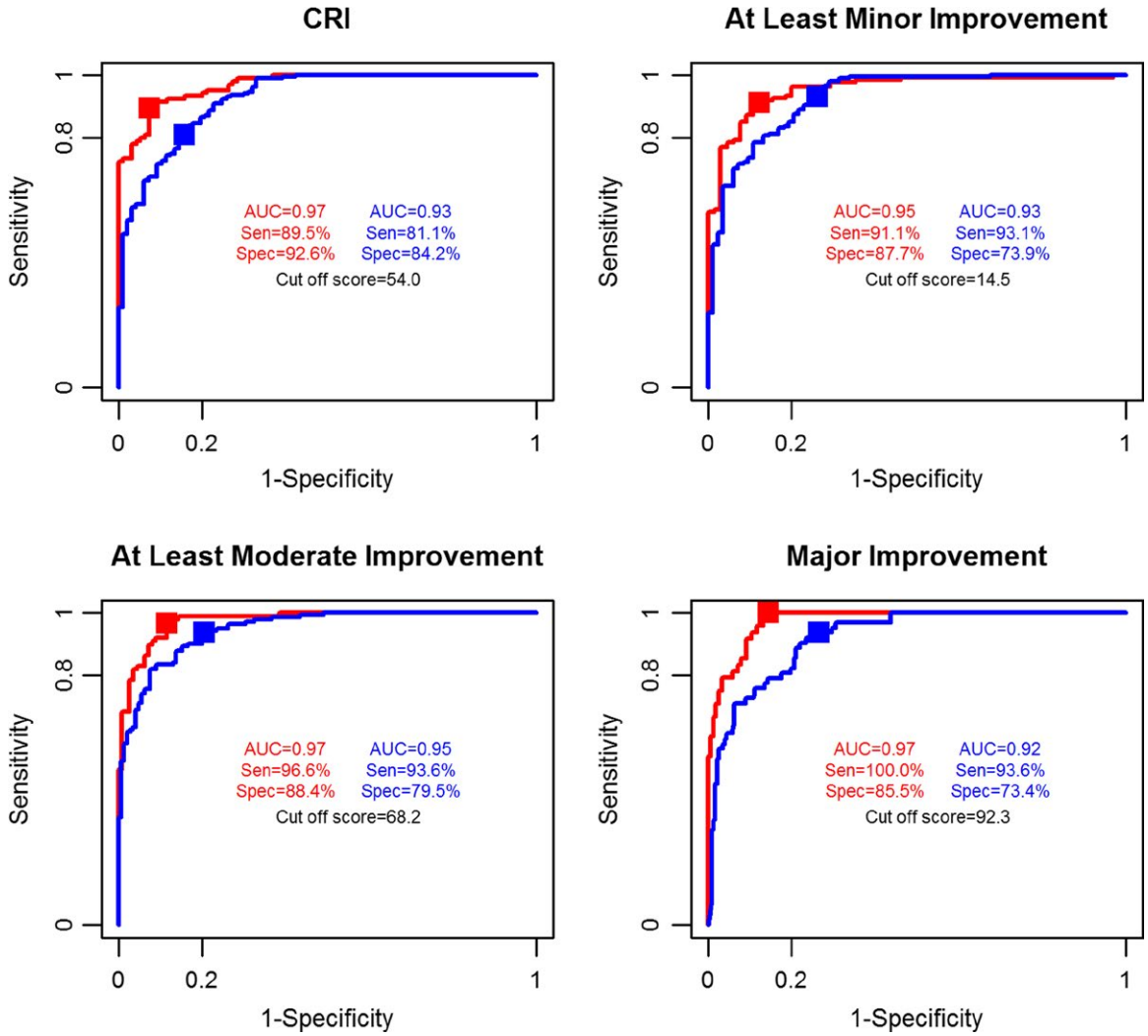

Figure 2. Performance of Childhood Lupus Improvement Index (CHILI) when considering absolute of the cSLE core response variables between baseline and follow-up in the development data set and the validation data set. The accuracy of the CHILI algorithm is largely maintained in the validation data set (blue) as compared to the training data set (red). As shown in panel labeled CRI a CHILI score of 54 (range 0–100) was 81.1% sensitive (Sen) and 84.5% specific (Spec) in the validation data set to capture CRI_{cSLE}, with an overall accuracy (area under the receiver operating characteristic curve [AUC]) of 0.93. CHILI algorithms developed and validated to measure CRICSLE were also excellent in discriminating patients with various levels of improvement (minor, moderate, major) between baseline and follow-up. CRI = clinically relevant improvement (see Figure 1 for other definitions).

the recently published ACR provisional criteria for global flares in cSLE (7,28). Indeed, more complex mathematical maneuvers beyond addition and multiplication are avoided, which is different from the Disease Activity Score in 28 joints, which includes a square root calculation, for example (29). For reasons associated with scaling, we transformed the CHILI scores to range from 0 to 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 for global flares in cSLE (7), the CHILI considers patient perspectives more comprehensively. Specifically, changes in patient overall well-being and physical function (CHQ-PhS) are included in the algorithm. This is consistent with the results of earlier discussions of how to capture response to therapy in cSLE (6,8).

Currently, the SRI is the principal outcome measure used in clinical trials in adults with SLE. We have shown that the PRINTO/ACR provisional criteria for response to therapy in cSLE (9) are more accurate than the SRI in capturing improvement in cSLE (10). In the current study, we confirmed that the PRINTO/ACR provisional criteria for response to therapy seem to have, at best, fair accuracy for capturing the true course of cSLE, including CRI_{cSLE}. Different from the CHILI, in the PRINTO/ACR provisional criteria for response to therapy all cSLE-CRV changes are considered equally (same percentage changes) relevant for measuring response to therapy. However, from a measurement point of view, as supported by the consensus definition for CRI_{cSLE} and our univariate analysis, the cSLE-CRVs have differential importance to clinicians when judging the disease course in a child with cSLE (10). Taken together, the PRINTO/ACR provisional criteria for response to therapy—and by extension the SRI—can be used in clinical trials of cSLE but likely require sample sizes larger than those 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 (30) or other disease activity indices instead of the SLEDAI as a measure of cSLE activity would have allowed us to accurately identify cSLE patients who experienced CRI_{cSLE}. Indeed, the cSLE-CRVs do not specify which validated measures of cSLE activity are 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. In addition, we did not provide patient profile raters with consensus definitions of what constitutes minor, moderate, and major improvement. Nonetheless, the accuracy of the CHILI algorithm performed well in the data sets used in this study. Last, we focused on the majority rule to adjudicate the disease course presented in the various patient profiles, which might have introduced bias. However, use of the 67% rule yielded results comparable to those using 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 to develop response criteria. However, such data from interventions that impact cSLE activity currently are unavailable. Thus, we used the patient profile raters' perceptions of the course of cSLE instead. Given the prospective character of our data and the expertise of the patient profile raters, we considered the quality of the training data set and the validation data set to be high, and the number of patient profiles to assess CRI_{cSLE} yielded a robust CHILI.

In summary, a methodologically stringent process was used 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.

ACKNOWLEDGEMENTS

We thank 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) from Cincinnati Children Hospital Medical Center (CCHMC) as well as 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 regarding study implementation, conduct, and statistical analysis: Drs. Tuhina Neogi, Ian Bruce, David Isenberg, Nicola Ruperto, and James Witter.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. Brunner had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Brunner, Ying.

Acquisition of data. Brunner, Holland, Beresford, Adroit, Appenzeller, Silva, Flores, Goliad, Avar Ayden, Wanderer, Levy, Ravelli, Khubchandani, Avcin, Klein-Gitelman, Ruperto, Feldman, Ying.

Analysis and interpretation of data. Brunner, Ying.

REFERENCES

1. Silva CA, Avcin T, Brunner HI. Taxonomy for systemic lupus erythematosus with onset before adulthood. Arthrit Care Res (Hoboken) 2012;64:1787–93.
2. Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 1997;40:1725.
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;58:556–62.

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. J Pediatr 2008;152:550–6.
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;52:2854–64.
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;42:1452–9.
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;70:813–22.
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;62:811–20.
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;55:355–63.
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;73:401–6.
11. U.S. Food and Drug Administration. Guidance for industry: clinical development programs for drugs, devices, and biological products for the treatment of rheumatoid arthritis (RA). In: Center for Drug Evaluation and Research (CDER) and Center for Devices and Radiologic Health (CDRH); 1999.
12. European Medicines Agency, Committee for Medicinal Products for Human Use. Guideline on pharmaceutical development of medicines for paediatric use; 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;33:789–95.
14. Giannini EH, Ruperto N, Ravelli A, Lovell DJ, Felson DT, Martini A. Preliminary definition of improvement in juvenile arthritis. Arthritis Rheum 1997;40:1202–9.
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;55:348–52.
16. Dillman DA, Smyth JD, Christian LM. Internet, phone, mail, and mixed mode surveys: the tailored design method (4th ed). Hoboken (NJ): John Wiley & Sons; 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. Arthritis Care Res (Hoboken) 2016;68:179–86.
18. Little RJ. Pattern-mixture models for multivariate incomplete data. J Am Stat Assoc 1993;88:125–34.
19. Schafer JL. Multiple imputation: a primer. Stat Methods Med Res 1999;8:3–15.
20. Rubin DB. Multiple imputation for nonresponse in surveys. Hoboken (NJ): Wiley-Interscience; 2004.
21. Gladman DD, Ibanez D, Urowitz MB. Systemic lupus erythematosus disease activity index 2000. J Rheumatol 2002;29:288–91.
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;59:4–13.
23. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29–36.
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;30:603–17.
25. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for outcome measures in rheumatology. J Rheumatol 1998;25:198–9.
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;41:1025–30.
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;67:745–53.
28. 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;63:1213–23.
29. 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;31:40–6.
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;44:902–6.
31. Johnson SR, Grayson PC. Use of "provisional" designation for response criteria. Arthritis Care Res (Hoboken) 2018;70:811–2.

APPENDIX A. Members of the Pediatric Rheumatology Collaborative Study Group and PRINTO Who Served as Patient Profile Raters

(Argentina) Cristina Battagliotti, Maria Isabel Brusco, Rubén Cuttica, Carmen De Cunto, Graciela Espada, Maximiliano Farfan, Stella Garay, Maria Marcantoni, Alvarez Marcela, Silvia Meiorin, Maria Elena Rama, Ricardo Russo, Carolina Torre Walsh, Celso Zamparo; (Australia) Navid Adib, Jonathan Akikusa,Christina Boros, Senq J. Lee, Damien Mckay, Susan Piper; (Belgium) Rik Joos; (Brazil) Blanca Bica, Leonardo Campos, André Cavalcanti, Rogerio do Prado, Amanda Donner-Maliki, Taciana Fernandes, Adriana Fonseca, Rozana Gasparello de Almeida, Andressa Guariento, Catherine Gusman, Fernanda Jusan Fiorot, Sheila Knupp Oliveira, Claudio Len, Lucia M. Arruda Campos, Sandra Machado, Luciana Marques, Luciana Martins de Carvalho, Rodrigo Moulin, Soraya Pedroso, Gecilmara Pileggi, Paulo Roberto S. Romanelli, Claudia Saad-Magalhaes, Ana Sakamoto, Maria Carolina Santos, Marco Felipe Silva, Paulo Spelling, Slavio Sztajnbok, Maria Terreri; (Canada) David Cabral, Gaëlle Chédeville, Janet Ellsworth, Adam Huber, Lori Tucker, Kristin Houghton; (Chile) Arturo Borzutzky, Mabel Ladino, Ximena Norambuena; (Colombia) Ruth Eraso, Angela Mosquera, Monica Velasquez; (Croatia) Miroslav Harjacek, Marija Jelusic; (Cuba) Cecilia Coto Hermosilla; (Czech Republic) Pavla Dolezalova; (Denmark) Susan Nielsen; (Dominican Republic) Carmen Tineo; (El Salvador) Mauricio Alegria; (Germany) Frank Dressler, Dirk Foell, Gerd Ganser, Claas Hinze, Markus Hufnagel, Thomas Lutz, Ralf Trauzeddel; (Greece) Sorina Boiu, Maria Trachana, Elena Tsitsami; (Guatemala) Mayra Cifuentes; (Hungary) Ilonka Orbán;

(India) Amita Aggarwal, Sujata Sawhney; (Israel) Yonatan Butbul Aviel; (Italy) Rolando Cimaz, Maria Cristina Maggio; (Latvia) Ingrid Rumba-Rozenfelde; (Libya) Soad Hashad; (Malaysia) Sern Chin Lim; (Mexico) Carlos Abud, Ruben Burgos-Vargas, Roberto Carreño-Manjarrez, Sandra Enciso Pelaez, Hayde Hernandez-Huirache, Rocio Maldonado Velázquez, Javier Orozco, Ana Luisa Rodriguez-Lozano, Omar Ernesto Rojas Pacheco, Luz Maria Suárez Larios, Gabriel Vega, Julia Verónica Ramírez Miramontes, Ivon Karina Ruíz Lopez; (Netherlands) Sylvia Kamphuis, Dieneke Schonenberg-Meinema; (New Zealand) Anthony Concannon, Jaqueline Yan; (Nicaragua) Martha Jarquin Jaime; (Oman) Safiya Al Abrawi; (Paraguay) Cynthia Vega, Jorge Lopez-Benitez; (Peru) Amparo Ibáñez Estrella, Tatiana Miraval; (Philippines) Leonia Dans, Karen Joy Kimseng; (Poland) Violetta Opoka-Winiarska, Lidia Rutkowska-Sak, Elzbieta Smolewska; (Portugal) Marta Conde, Margarida Guedes; (Puerto Rico) Enid del Valle, Ana Quintero-Del Rio; (Romania) Constantin Ailioaie, Mihaela Sparchez; (Russian Federation) Ekaterina Alekseeva, Vladimir Keltsev; (Saudi Arabia) Sulaiman Al-Mayouf, Abdurhman Asiri, Wafaa Suwairi; (Serbia) Gordana Susic, Gordana Vijatov-Djuric; (Singapore) Elizabeth Ang, Thaschawee Arkachaisri; (Spain) Alina-Lucica Boteanu, Juan Carlos Lopez-Robledillo, Marta Medrano San Ildefonso, Consuelo Modesto, Inmaculada Calvo; (Sweden) Jorge Sotoca-Fernandez; (Switzerland) Isabel Bolt; (Thailand) Soamarat Vilaiyuk; (United Kingdom) Eslam Al-Abadi, Eileen Baildam, Lampros Fotis, Clare Pain, Clarissa Pilkington; (United States) Khalid Abulaban, Fatima Barbar-Smiley, Bryce Binstadt, John Bohnsack, Alexis Boneparth, Diane Brown, Peter Chira, Randy Cron, Fatma Dedeoglu, Anne Eberhard, Abraham Gedalia, Alexei Grom, Beatrice Goilav, Michael Henrickson, Christine Hom, Jennifer Huggins, Rita Jerath, Jordan Jones, Lawrence Jung, Daniel Kingsbury, Jamie Lai, Daniel Lovell, Kabita Nanda, James Nocton, Judyann Olson, Kathleen O'Neil, Karen Onel, Lynn Punaro, Andreas Reiff Kelly Rouster-Stevens, Natasha Ruth, Ken Schikler, Kara Murphy Schmidt, Grant Schulert, Bracha Shaham, Nora Singer, Judith Smith, Robert Sundel, Grant Syverson, Patricia Vega-Fernandez, Richard Vehe, Linda Wagner-Weiner; (Uruguay) Juan Cameto, Rosario Jurado; (Venezuela) Irama Maldonado.