






# New Features for Measuring Disease Activity in Pediatric Localized Scleroderma

Suzanne C. Li , Xiaohu Li , Elena Pope , Katie Stewart, Gloria C. Higgins, C. Eglia Rabinovich, Kathleen M. O'Neil, Kathleen A. Haines, Ronald M. Laxer, Marilynn Punaro, Heidi Jacobe, Tracy Andrews, Knut Wittkowski , Themba Nyirenda , Ivan Foeldvari, and Kathryn S. Torok

**ABSTRACT. Objective.** To identify clinical features that define disease activity in pediatric localized scleroderma (LS), and determine their specificity and importance.

**Methods.** We conducted a multicenter prospective study of patients with active and inactive LS skin lesions. A standardized evaluation of a single designated study lesion per subject was performed at 3 visits. We evaluated the pattern and correlation between assessed features and physician's global assessments of activity (PGA-A).

**Results.** Ninety of 103 subjects had evaluable data; 66 had active and 24 inactive disease. Subjects had similar age of onset, sex, and disease patterns. Linear scleroderma was the most common subtype. Features specific for active disease included erythema, violaceous color, tactile warmth, abnormal skin texture, and disease extension. Scores for these variables changed over time and correlated with PGA-A of the lesion. Active and inactive lesions could not be distinguished by the presence or level of skin thickening, either of lesion edge or center. However, in active lesions, skin thickening scores did correlate with PGA-A scores. Regression analysis identified the combination of erythema, disease extension, violaceous color, skin thickening, and abnormal texture as predictive of PGA-A at study entry. Damage features were common irrespective of activity status.

**Conclusion.** We identified variables strongly associated with disease activity, expanding upon those used in current measures, and determined their relative importance in physician activity scoring. Skin thickening was found to lack specificity for disease activity. These results will help guide development of a sensitive, responsive activity tool to improve care of patients with LS. (J Rheumatol First Release September 15 2018; doi:10.3899/jrheum.171381)

## Key Indexing Terms:

LOCALIZED SCLERODERMA PHYSICIAN'S GLOBAL ASSESSMENT DISEASE ACTIVITY

From the Joseph M. Sanzari Children's Hospital, Hackensack University Medical Center, Department of Pediatrics, Hackensack; Stevens Institute of Technology, Hoboken, New Jersey, USA; Hospital for Sick Children, Toronto, Ontario, Canada; Texas Scottish Rite Hospital, University of Texas (UT) Southwestern, Dallas, Texas; The Ohio State University and Nationwide Children's Hospital, Columbus, Ohio; Duke University, Durham, North Carolina; Riley Hospital for Children at Indiana University Health, Indianapolis, Indiana; Rockefeller University, New York, New York, USA; Hamburger Zentrum für Kinder- und Jugendrheumatologie, Hamburg, Germany; Children's Hospital of Pittsburgh, Pittsburgh, Pennsylvania, USA.

Financial support from the Arthritis Foundation New Jersey Chapter Grant; Childhood Arthritis and Rheumatology Research Alliance.

S.C. Li, MD, PhD, Joseph M. Sanzari Children's Hospital, Hackensack University Medical Center, Department of Pediatrics; X. Li, PhD, Stevens Institute of Technology; E. Pope, MD, Hospital for Sick Children; K. Stewart, MD, Texas Scottish Rite Hospital, UT Southwestern; G.C. Higgins, MD, PhD, The Ohio State University and Nationwide Children's Hospital; C.E. Rabinovich, MD, MPH, Duke University; K.M. O'Neil, MD, Riley Hospital for Children at Indiana University Health; K.A. Haines, MD, Joseph M. Sanzari Children's Hospital, Hackensack University Medical Center, Department of Pediatrics; R.M. Laxer, MD, Hospital for Sick Children; M. Punaro, MD, Texas Scottish Rite Hospital, UT Southwestern; H. Jacobe, MD, UT Southwestern; T. Andrews, MA, Hackensack University Medical Center; K. Wittkowski, PhD, Rockefeller University; T. Nyirenda, PhD, Hackensack University Medical Center; I. Foeldvari, MD, Hamburger Zentrum für Kinder- und Jugendrheumatologie; K.S. Torok, MD, Children's Hospital of Pittsburgh.

Address correspondence to Dr. S.C. Li, Joseph M. Sanzari Children's Hospital, Hackensack University Medical Center, 30 Prospect Ave., Imus 348, Hackensack, New Jersey 07601, USA.

E-mail: suzanne.li@hackensackmeridian.org

Accepted for publication June 25, 2018.

Localized scleroderma (LS) is the most common childhood form of scleroderma, an autoimmune disease whose pathology includes inflammation, vasculopathy, and fibrosis<sup>1,2,3</sup>. The disease often lasts throughout childhood either with a persistently active or a remitting and relapsing course, and many continue to have active disease as adults<sup>4,5,6,7</sup>. Deep tissue and extracutaneous involvement is common<sup>2,8</sup>, and pediatric-onset LS is associated with higher damage levels than adult-onset disease<sup>7</sup>. In contrast to the widely held impression that pediatric LS is a "benign disease" with a "good outcome," its chronicity and extracutaneous involvement cause important longterm morbidity<sup>2,6,9</sup>. Sequelae include limb and facial hemiatrophy, severe disfigurement, arthritis, seizures, uveitis, psychosocial complications, and rarely, death<sup>2,10,11,12</sup>.

There are few options for treating fibrosis, so treatment

aims to minimize progression of fibrosis and other damage by controlling inflammation. While systemic immunosuppression is considered the standard of care for moderate to severe pediatric LS<sup>13,14,15</sup>, comparative effectiveness and other large treatment studies are needed to identify the most effective regimens<sup>16</sup>. Longterm disease monitoring is needed because active disease can persist for decades<sup>4,6</sup>, and relapses are common (15–53% of cases)<sup>17,18,19,20,21</sup>. A sensitive disease activity tool to track disease status and measure treatment response is therefore needed.

Assessment of pediatric LS disease activity is difficult because routine laboratory tests are not reliable biomarkers of active disease in most patients<sup>22</sup>. While imaging modalities can aid evaluation, especially of deeper tissue involvement<sup>8,23,24,25,26</sup>, routine evaluation relies on clinical assessment. Clinical measures to evaluate disease activity or severity (activity and damage) include the Modified Skin Score (MSS, modeled after the modified Rodnan skin score); Dyspigmentation, Induration, Erythema, and Telangiectasia measure (DIET); LS Severity Index (LoSSI, and its modification, mLoSSI), and Computerized Skin Score (CSS)<sup>27,28,29,30,31,32</sup>. All these measures score skin thickening (ST) or induration. Additional scored features include erythema (DIET, LoSSI), disease extension (LoSSI, CSS), dyspigmentation (DIET), telangiectasia (DIET), and the extent of the scored feature(s). The CSS precisely assesses the extent of a single lesion using serial tracings, while other measures include a limited extent assessment across all affected anatomic regions (MSS, LoSSI)<sup>27,29,31</sup>. All these measures can track treatment response, but the specificity, sensitivity, and relative importance of individual lesion features as indicators of activity have not been assessed.

To identify features associated with activity, we conducted a multicenter study of 90 pediatric LS subjects classified by their treating physician as having either active or inactive disease. Our study focused on detailed clinical assessment of a single lesion (study lesion) per subject that was prospectively tracked in 84 subjects to determine the correlation of features with physician's global assessment (PGA) of disease activity. Our study design allowed us to identify features specific for active disease, and to determine each one's relative importance in physician activity assessment.

## MATERIALS AND METHODS

This prospective observational multicenter cohort study was conducted by Localized scleroderma Clinical and Ultrasound Study group (LOCUS), a multidisciplinary collaboration organized through the Childhood Arthritis and Rheumatology Research Alliance (CARRA). Ten pediatric rheumatologists and 2 dermatologists from 9 academic centers participated. Numerous face-to-face meetings and conference calls were held to establish and standardize scoring methods. These included reviews of photographs of pediatric LS lesions that had erythema and violaceous color, and skin thickness assessment workshops. Inclusion criteria were a diagnosis of pediatric LS confirmed by a pediatric rheumatologist or dermatologist, and classified according to the Padua Preliminary Classification Criteria<sup>33</sup>, with disease onset before the 16th birthday.

Physicians classified their subjects as having active or inactive disease,

and enrolled them at a minimum ratio of 2:1 active:inactive. Treatment was at the discretion of the subject's physician, without restriction. Study visits were performed at visits for clinical care every  $3 \pm 1$  months (baseline, 3, and 6 months). Each site obtained institutional ethics approval for the study, assent, and consent forms, which included our intent to publish the results of the study and measures to protect confidentiality. Data identified only by subject number were analyzed at the coordinating center (Hackensack University Medical Center, ethics approval number 07.02.055).

Because LS lesions vary in their features, a single lesion per subject was selected to serve as the study lesion for all visits. For active subjects, the investigator designated the most active, readily evaluable lesion, and specified the features that indicated that the study lesion was active. For subjects with inactive disease, any lesion could be designated as the study lesion.

PGA are reliable, quantitative measures that are sensitive to change when physicians are well-trained and experienced<sup>34,35</sup>. The study physicians have extensive experience evaluating and treating pediatric LS, collectively following over 500 patients. Because of the lack of an objective measure to determine disease activity, we used PGA assigned by this group of physicians as the gold standard. To avoid interrater variability in scoring, the same investigator examined a given subject's study lesion at all visits, scoring 11 features (Table 1) and 4 PGA (visual analog scales, 0–100 mm). The scored PGA were (1) activity of study lesion (anchors: not active, very active), (2) activity of subject's overall disease (anchors: not active, very active), (3) level of damage of study lesion (anchors: no loss, severe loss), and (4) level of subject's overall disease damage (anchors: no chronic change, severe chronic change).

Abnormal skin texture was defined as representing lesions that had an altered appearance and texture, such as an abnormally smooth, shiny, and/or waxy appearance. A composite variable, STmax, was created to represent the maximum score of skin thickening of the edge or center in some analysis. For each variable, a set of dichotomous variables was created for the regression analysis [normal or none (yes/no), mild (yes/no), moderate (yes/no), and severe (yes/no)]. Medical history, family history, and demographics were also collected.

The 90 evaluated subjects had complete data at the first visit; some lacked data for later visits. Subjects with missing values were excluded from the analysis of those visits, but included in analyses for which they had complete data to maximize our available study sample. The sample size for Visit 1 analysis was 90; Visit 2, 83; and Visit 3, 79. From Visit 2 to 3, 78 subjects were evaluated.

**Statistical analysis.** Disease activity was assessed by initial physician classification of the subject as active or inactive, and PGA of activity (PGA-A). Disease activity scores were compared between active and inactive group at each visit, and within each group across visits. For categorical variables, changes between groups at each visit were assessed using chi-square or Fisher's exact test; comparisons within groups across time were assessed using Friedman's test. Continuous data were assessed for normality using the Shapiro-Wilks test. Two-sided 2-sample t tests and/or Wilcoxon rank sum tests (non-normal analog) were used to evaluate differences in variables between groups at each visit, and 2-sided paired t tests or Wilcoxon signed-rank tests (non-normal analog) were used to evaluate differences within groups across visits. The relationship between variables was assessed by Pearson's correlation coefficient. The correlation of variable scores with PGA-A scores was examined in bee-swarm box plots.

Multivariate linear regression was performed to examine the relationship between lesion features and PGA-A, the dependent variable. Incremental model building was used to determine the final model (evaluated by F test and adjusted R<sup>2</sup>). Regression coefficients allowed us to determine the relative ability (or weight) of each variable to explain the variation in PGA-A scores. For all analyses,  $p < 0.05$  was considered significant.

## RESULTS

**Characteristics of the study subjects.** Of the 103 patients with pediatric LS enrolled, 13 were excluded because of incomplete data. Of the remaining 90 subjects, 66 had an active study lesion and 24 had an inactive study lesion. Most study

Table 1. Scored lesion features.

Lesion Feature	Scoring Range
Visible	
Erythema	0 = none, 1 = mild, 2 = moderate, 3 = severe
Violaceous or blue color	0 = none, 1 = present, 2 = moderate, 3 = severe
Dyspigmentation (hyper or hypo)	0 = none, 1 = mild, 2 = moderate, 3 = severe
Dermal atrophy	0 = none, 1 = visible veins or cliff-drop atrophy
Subcutaneous atrophy	0 = none, 1 = mild fat loss, 2 = moderate loss, 3 = severe (absent subcutaneous fat)
Palpable	
Abnormal skin texture	0 = none, 1 = shiny, waxy, or abnormally smooth
Tactile lesion warmth	0 = none, 1 = increased warmth
STE	0 = none, 1 = mild, 2 = moderate, 3 = severe (no horizontal mobility)
STC	0 = none, 1 = mild, 2 = moderate, 3 = severe (no horizontal mobility)
Visible or palpable	
Distinct lesion border	0 = none, 1a = palpable border, 1b = visible (colored) border, 2 = both a palpable and visible border are present
Size	
New lesion	0 = none, 1 = study lesion is a new lesion (developed within prior 3 mos); can only be scored at baseline visit
Larger size (includes deeper)	0 = no, 1 = increase in study lesion size from prior study visit, or within prior 3 mos if baseline visit

Each lesion feature and its scoring range are shown. New lesions could only be scored at Visit 1. At every visit, the study lesion was scored for the worst (highest) level of each characteristic that was present. The selection of features was based upon published studies and group discussions during the study design. Abnormal skin texture was defined as representing lesions that had an altered appearance and texture, with lesions having an abnormally smooth, shiny, and/or waxy appearance. For scoring larger lesion size, physicians could base this assessment on either serial photographs or serial measurements of the lesion. STE: skin thickening lesion edge; STC: skin thickening lesion center.

lesions were located on a limb, with linear scleroderma the most common subtype (Table 2). Features most commonly cited by physicians as indicative of activity in the active study lesions at enrollment were erythema 41/66 (62%), induration

34/66 (51.5%), tactile warmth 24/66 (36%), and lesion extension 22/66 (33%). Distinct border 16/66 (24%), development of a new lesion 13/66 (20%), violaceous color 12/66 (18%), white color or hypopigmentation 12/66 (18%), and

Table 2. Study population characteristics.

Characteristics	All Subjects, n = 90	Active Subjects, n = 66	Inactive Subjects, n = 24	p
Age at study entry, yrs, median (IQR)	11.5 (8.8, 14.5)	11.0 (8.2, 13.9)	12.9 (10.1, 16.1)	NS
Age at disease onset, yrs, median (IQR)	7.9 (5.0, 10.0)	8.0 (5.1, 10)	7.0 (5.0, 9.1)	NS
Disease duration at study onset, yrs, median (IQR)	2.8 (1.3, 5.0)	2.1 (0.8, 4.4)	4.4 (2.3, 7.2)	0.004
Duration in study, mos (IQR)	5.6 (4.4, 6.8)	5.5 (4.3, 6.5)	6.3 (5.4, 7.2)	NS
Sex, F:M	71:19	53:13	18:6	NS
White	70 (78)	52 (79)	18 (75)	NS
Linear scleroderma subtype	55 (61)	41 (62)	14 (58)	NS
Limb/trunk	44 (51.1)	33 (50.0)	11 (45.8)	
Head	11 (11.1)	8 (12.1)	3 (12.5)	
Circumscribed morphea subtype	35 (38.9)	25 (37.9)	10 (41.7)	NS
Superficial	19 (21.1)	14 (21.2)	5 (20.8)	
Deep	16 (17.8)	11 (16.7)	5 (20.8)	
Anatomic sites: Head	14 (15.6)	10 (15)	4 (16.7)	NS
Trunk	28 (31)	21 (32)	7 (29)	NS
Limb	48 (53)	35 (53)	13 (54)	NS
ANA positivity	38 (42)	28 (42)	10 (42)	NS
Extracutaneous morbidity	50 (56)	37 (56)	13 (54)	NS
Prior treatment with MTX	64 (71)	42 (64)	22 (92)	0.001
Prior treatment with corticosteroids	55 (61)	39 (59)	16 (67)	NS

Data are n (%) unless otherwise indicated. P values refer to results of tests of differences between active and inactive subjects reported by 2-sided Wilcoxon rank sum test or 2-sample t test. IQR: interquartile range; NS: not significant; MTX: methotrexate; ANA: antinuclear antibodies.

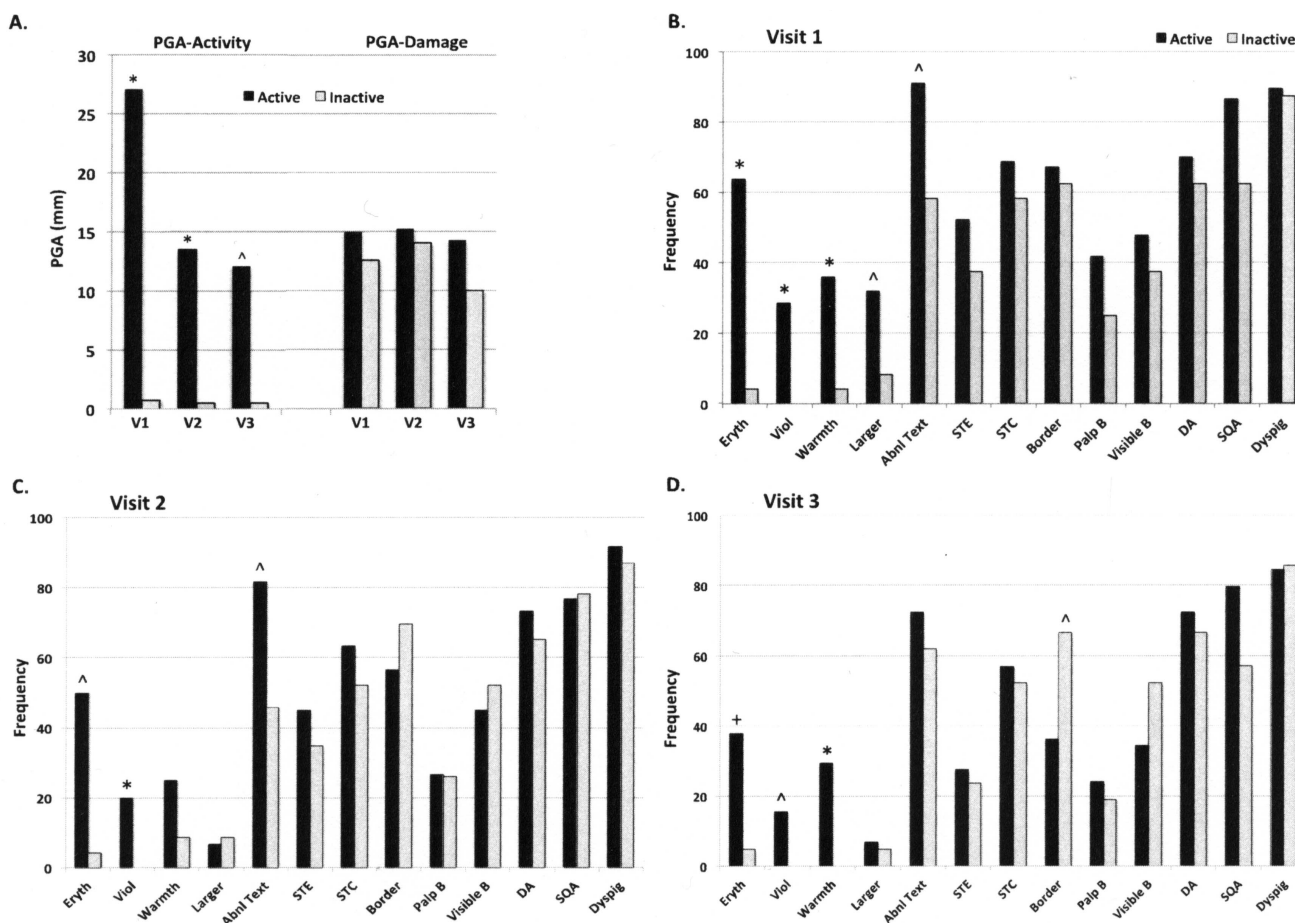
hyperpigmentation 10/66 (15%) were also cited.

Seventy-eight subjects completed 3 visits, 6 two visits, and 6 one visit. Most subjects were white (70/90) and female (71/90; Table 2). Mean age at disease onset was 7.7 years (SD 3.6). Groups did not differ in age at disease onset, race, antinuclear antibody (ANA) positivity, or extracutaneous morbidity (Table 2).

Active subjects had a shorter median disease duration than inactive subjects, lower frequency of prior treatment with methotrexate (MTX; Table 2), and shorter duration of prior MTX treatment (0.17 vs 2 yrs,  $p < 0.001$ ). They were more likely to receive corticosteroid treatment during the study (active 67%, inactive 25%,  $p < 0.001$ ). Active subjects had higher scores for PGA-A of the study lesion, while scores for PGA of damage (PGA-D) of the study lesion did not differ significantly among subjects with active versus inactive disease (Figure 1A). An expected similar difference in overall

disease activity scores was also found at all visits (data not shown) in PGA-A for overall disease activity (median active, inactive scores: Visit 1, 30.0, 0.25 mm; Visit 2, 18.1, 0.52 mm; Visit 3, 13.8, 0.50 mm;  $p < 0.001$  all visits). As was the case for damage assessment in the individual study lesions, PGA-D of overall disease at Visit 1 were similar (active 25.1, inactive 21.0 mm). However, in active subjects, PGA-D of overall disease scores remained stable across visits (Visit 2, 27.7; Visit 3, 24.8), but declined in inactive subjects (Visit 2, 15.1,  $p$  NS; Visit 3, 11.5,  $p = 0.007$ ).

**Frequency of lesion features.** The frequencies of assessed lesion features were calculated for active and inactive study lesions, with differences found for the following features: development of a new lesion, enlargement of study lesion, erythema, violaceous color, tactile warmth, abnormal skin texture, and distinct lesion border (Figure 1, B–D). Development of a new lesion, erythema, and violaceous color were



**Figure 1.** Physician's global assessment (PGA) and lesion feature scores for active and inactive lesions. The median PGA of study lesion (Panel A, mm, range 0–100) and frequency of assessed lesion features at each visit (Panels B–D) are shown for active (dark columns) and inactive (light columns) study lesions. Significant differences between the active and inactive lesions are indicated with an \* ( $p < 0.001$ ), ^ ( $p < 0.01$ ), or + ( $p < 0.05$ ) above the column. Not shown is new lesion, which could only be scored at Visit 1 (active 13, inactive 0,  $p < 0.001$ ). Subject numbers at visits (V): Visit 1 — 66 active, 24 inactive; Visit 2 — 60 active, 23 inactive; Visit 3 — 58 active, 21 inactive. PGA-Activity: PGA of activity of study lesion; PGA-Damage: PGA of damage of study lesion; Eryth: erythema; Viol: violaceous color; Warmth: tactile lesion warmth; Larger: enlargement of study lesion; Abnl Text: abnormal skin texture; STE: skin thickening of lesion edge; STC: skin thickening of lesion center; Border: distinct lesion border (visible or palpable); Palp B: distinct palpable lesion border; Visible B: distinct visible lesion border; DA: dermal atrophy; SQA: subcutaneous atrophy; Dyspig: dyspigmentation (hyper- or hypopigmentation).



more frequent in active lesions at all visits (a new lesion could be scored only at Visit 1), with new lesion and violaceous color exclusive to active lesions (Figure 1, B–D; new lesion at Visit 1, active 13, inactive 0,  $p < 0.001$ ). Tactile warmth and abnormal skin texture were more frequent in active lesions at 2 visits, with tactile warmth approaching significance at the remaining visit (tactile warmth visit 1 and 3,  $p < 0.001$ ; Visit 2,  $p = 0.052$ ). Enlargement of the study lesion was more frequent in active lesions at Visit 1 (Figure 1B,  $p < 0.01$ ), while distinct lesion border, either palpable or visible, occurred at a lower frequency in active lesions at Visit 3 (Figure 1D,  $p < 0.01$ ).

The frequency of skin thickening of the lesion edge (STE), skin thickening of lesion center (STC), dermal atrophy, subcutaneous atrophy, and dyspigmentation did not differ between active and inactive lesions at any visit (Figure 1, B–D). STC was present in most study lesions (active 56.9–68.7%, inactive 52.4–58.3%), while STE was scored in up to half the study

lesions (active 27.6–52.2%, inactive 23.8–37.5%). Dermal atrophy, subcutaneous atrophy, and dyspigmentation were present in the majority of lesions at all visits (Figure 1, B–D). *Lesion feature scores in active and inactive study lesions.* We evaluated whether there were differences between groups in the distribution of scores for features scored on an ordinal (0–3) scale at each visit (Table 3). At all visits, active and inactive lesion groups differed in the distribution of erythema and violaceous color scores (Table 3, erythema  $p = 0.016$  to  $< 0.001$ ; violaceous  $p = 0.013$  to  $0.042$ ). Groups did not differ in the distribution of STE, STC, subcutaneous atrophy, or dyspigmentation scores at any visit.

Changes in the mean value of lesion feature scores across visits were compared between active and inactive study lesions. Active and inactive lesions differed in the level of change in scores for erythema, violaceous color, tactile warmth, abnormal skin texture, and distinct lesion border ( $p < 0.001$  to  $0.043$ ). Active and inactive lesions did not differ

Table 3. Distribution of lesion feature scores by visit and activity status.

Lesion Feature and Score	Visit 1			Visit 2			Visit 3		
	Inactive	Active	p	Inactive	Active	p	Inactive	Active	p
Erythema			< 0.001			0.001			0.016
None	23 (95.8)	23 (34.9)		22 (95.7)	30 (52.6)		20 (95.2)	36 (62.1)	
Mild	1 (4.2)	30 (45.5)		1 (4.4)	25 (43.9)		1 (4.8)	20 (34.5)	
Moderate	0 (0.0)	13 (19.7)		0 (0.0)	2 (3.5)		0 (0.0)	2 (3.5)	
Severe	0 (0.0)	0 (0.0)		0 (0.0)	0 (0.0)		0 (0.0)	0 (0.0)	
Violaceous color			0.013			0.017			0.042
None	24 (100)	47 (71.2)		23 (100)	45 (79.)		21 (100.)	48 (82.8)	
Mild	0 (0.0)	16 (24.2)		0 (0.0)	12 (21.1)		0 (0.0)	10 (17.3)	
Moderate	0 (0.0)	3 (4.6)		0 (0.0)	0 (0.0)		0 (0.0)	0 (0.0)	
Severe	0 (0.0)	0 (0.0)		0 (0.0)	0 (0.0)		0 (0.0)	0 (0.0)	
STE			NS			NS			NS
None	15 (62.5)	34 (51.5)		15 (65.2)	31 (54.4)		16 (76.2)	44 (75.9)	
Mild	9 (37.5)	22 (33.3)		6 (26.1)	22 (38.6)		5 (23.8)	11 (19.0)	
Moderate	0 (0.0)	8 (12.1)		2 (8.7)	4 (7.0)		0 (0.0)	3 (5.2)	
Severe	0 (0.0)	2 (3.03)		0 (0.0)	0 (0.0)		0 (0.0)	0 (0.0)	
STC			NS			NS			NS
None	10 (41.7)	22 (33.3)		11 (47.8)	22 (38.6)		11 (52.4)	24 (41.4)	
Mild	7 (29.2)	17 (25.8)		8 (34.8)	19 (33.3)		7 (33.3)	19 (32.8)	
Moderate	7 (29.2)	24 (36.4)		4 (17.39)	15 (26.3)		3 (14.3)	15 (25.9)	
Severe	0 (0.0)	3 (4.6)		0 (0.0)	1 (1.8)		0 (0.0)	0 (0.0)	
SQ atrophy			NS			NS			NS
None	9 (37.5)	23 (34.9)		5 (21.7)	11 (19.3)		9 (42.9)	16 (27.6)	
Mild	9 (37.5)	29 (43.9)		15 (65.2)	33 (57.9)		8 (38.1)	29 (50.0)	
Moderate	5 (20.8)	11 (16.7)		2 (8.7)	8 (14.0)		4 (19.1)	11 (19.0)	
Severe	1 (4.2)	3 (4.6)		1 (4.4)	5 (8.8)		0 (0.0)	2 (3.5)	
Dyspigmentation			NS			NS			NS
None	3 (12.5)	7 (10.6)		3 (13.)	5 (8.8)		3 (14.3)	7 (12.3)	
Mild	8 (33.3)	31 (45.0)		12 (52.2)	29 (50.9)		11 (52.4)	26 (45.6)	
Moderate	12 (50.0)	21 (31.8)		6 (26.1)	21 (36.8)		6 (28.6)	21 (36.8)	
Severe	1 (4.2)	7 (10.6)		2 (8.7)	2 (3.5)		1 (4.76)	3 (5.3)	

The distribution of scores for lesion features scored on a 0 to 3 range were examined in active and inactive lesions. The left column lists the lesion features and their scoring levels. The n (%) of lesions with each of the scoring level scores are shown for active and inactive lesion groups at each visit. P values were calculated for the distribution of the scores in active compared to inactive group at each visit. NS: not significant; STE: skin thickening of the lesion edge; STC: skin thickening of the lesion center; SQ atrophy: subcutaneous atrophy.

in the level of change in scores for STE, STC, dermal atrophy, subcutaneous atrophy, or dyspigmentation).

*Lesion feature scores in active study lesions.* Within the active study lesion group, scores for erythema, violaceous color, tactile warmth, abnormal skin texture, and presence of a distinct border changed across visits ( $p < 0.001$  to  $0.035$ ). In contrast to the lack of difference between active and inactive lesions in the level of skin thickening score change over time, scores for both STE and STC changed from Visit 1 to Visit 3 within the active study lesion group ( $p = 0.008$ ,  $0.011$ , respectively). Scores for dermal atrophy, subcutaneous atrophy, and dyspigmentation did not change across visits in the active study lesion group (data not shown). None of the lesion features had a change in scores across visits in the inactive study lesion group (data not shown).

We examined the correlations among disease features to evaluate their uniqueness. In active lesions, erythema and violaceous color had low correlation coefficients at all visits, suggesting they represent distinct features (visits 1, 2, and 3, Pearson's  $r$  0.237,  $-0.122$ ,  $-0.176$ , respectively). Moderate correlations were found between skin thickening of lesion edge and center at all visits (visits 1, 2, and 3, Pearson's  $r$  0.440, 0.416, 0.452, respectively), suggesting that they are related. We evaluated the correlation of STE, STC, and a composite skin thickening variable representing the maximum score of STE or STC (STmax) with PGA-A of the study lesion. All 3 variables had similar levels of correlation with PGA-A of the study lesion at all visits (STmax, STE, STC Pearson's  $r$ : Visit 1 — 0.222, 0.227, 0.151; Visit 2 — 0.389, 0.284, 0.272; Visit 3 — 0.455, 0.285, 0.462, respectively). We therefore used STmax to evaluate the correlation between ST and PGA-A.

At all visits, active study lesions with erythema, abnormal skin texture, or skin thickening (STmax) had higher PGA-A scores than those without these features ( $p < 0.05$  to  $< 0.001$ , Figure 2). Active study lesions with disease extension (new or larger lesion), violaceous color, tactile warmth, or distinct border had higher PGA-A scores than lesions without these features for at least 1 visit (Figure 2; violaceous color, Visit 1,  $p = 0.003$ ; warmth, Visit 2,  $p = 0.088$ , and Visit 3,  $p = 0.012$ ; distinct lesion border, Visit 1,  $p < 0.001$ , and Visit 2,  $p = 0.020$ ). Presence or degree of dyspigmentation or subcutaneous atrophy were not associated with PGA-A scores (data not shown).

*Ability of lesion features to predict PGA-A of the study lesion.* Multivariate linear regression was used to predict PGA-A of the study lesion at each visit. We used STmax to evaluate the contribution of ST, and disease extension to represent the contribution of development of a new lesion and enlargement of an existing lesion. The best model contained the following variables: erythema, disease extension, violaceous color, maximal skin thickening, and abnormal skin texture [ $F(5,90) = 15.1$ ,  $p < 0.0001$ ; adjusted  $R^2 = 0.588$  at Visit 1, Table 4].

At Visit 1, the variables in the regression model explained 58.8% of the variation in PGA-A scores. More severe levels

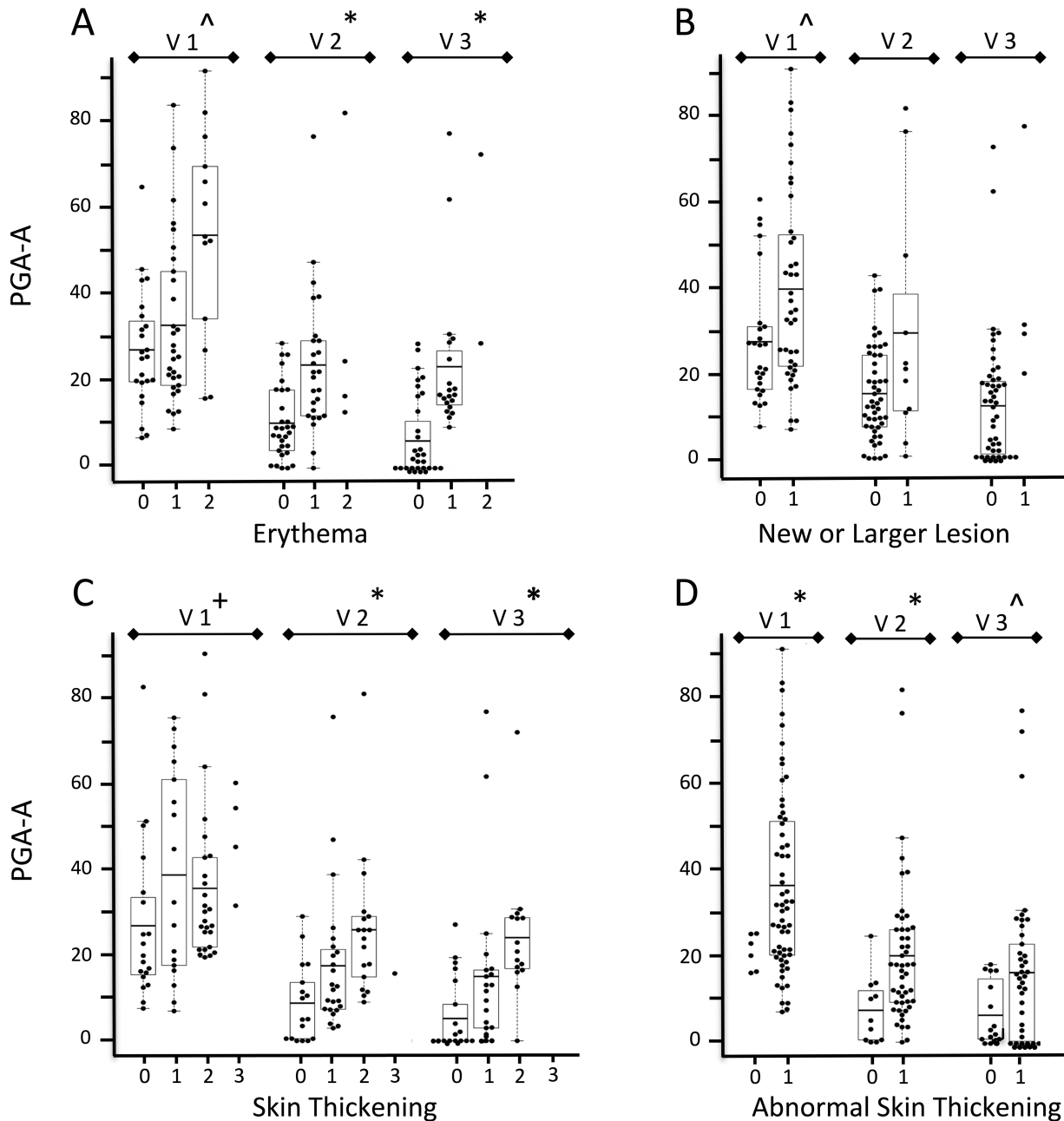
of disease features were associated with larger increases in PGA-A scores (Table 4). The effect of erythema and skin thickening were disproportionately higher at more severe levels of disease. Erythema had the largest effect on PGA-A score; mild and moderate disease levels were associated with 12- and 28-point increases in PGA-A scores, respectively (Table 4). Patients with mild and moderate violaceous color had PGA-A scores that were, on average, 13 and 19 points higher than patients lacking violaceous color, while disease extension, abnormal skin texture, and severe skin thickening were associated with 12-, 11-, and 22-point increases in PGA-A scores, respectively.

At Visit 2, the variables in the regression model explained 32.5% of the variation in PGA-A scores. Erythema was the strongest predictor of PGA-A scores, followed by disease extension and skin thickening (Table 4). At Visit 3, the variables in the regression model explained 49.3% of the variation in PGA-A scores. Erythema remained the strongest predictor of PGA-A, with more severe levels associated with a disproportionately higher effect, followed by disease extension (Table 4). The effect of moderate levels of skin thickening on PGA-A scores increased slightly from Visit 2.

## DISCUSSION

This is the first study, to our knowledge, to prospectively evaluate LS skin features for their association with active disease. Because we had physicians classify subjects at study entry as having active or inactive disease, we could identify features specific for activity. In addition, because scoring focused on a single lesion per subject, we could more accurately determine correlation between features and lesion PGA-A. No single feature was found ubiquitous to all active lesions. Instead, erythema, new disease extension, violaceous color, tactile warmth, and abnormal skin texture were identified as specific activity features, each present in a subset of active lesions. These features were more prevalent and scored at higher levels (for erythema and violaceous color) in active lesions, and similar to PGA-A scores, their scores declined over time. Scores for these features correlated with PGA-A, suggesting they can be used to track changes in activity level. As expected, dyspigmentation, dermal atrophy, and subcutaneous atrophy were not specific to active disease, and their scores did not correlate with PGA-A.

Skin thickening, a feature used in all current LS measures, was not found to be specific for active disease. It is possible our study period was too short to allow us to identify differences between active and inactive lesions in the level of skin thickening score change over time, which warrants future study. However, we were also not able to differentiate active from inactive lesions by either the level of skin thickening scores, or the frequency of skin thickening of the lesion edge or center, even at the first visit when inactive subjects had a median 2 years longer disease duration than active subjects. We suspect that the lack of specificity reflects the dual nature



**Figure 2.** Box plots of lesion feature scores versus PGA-A of the study lesion. The lesion feature scores for all active lesions were plotted against their PGA-A of the study lesion (mm, range 0–100) for each visit. Erythema (Panel A) and skin thickening (maximal skin thickening, Panel C) could be scored from 0 to 3, while new or larger lesion [new lesion or larger lesion size (disease extension), Panel B] and abnormal skin texture (Panel D) were scored as 0 (none) or 1 (present). Boxes indicate 25% and 75%, horizontal line marks the mean, and T brackets indicate the 5% and 95% PGA-A of the study lesion scores. Significant differences for PGA-A between lesion feature scores of 0 versus > 0 are indicated with an \* ( $p < 0.001$ ), ^ ( $p < 0.01$ ), or + ( $p < 0.05$ ) on the visit number (V). There were 66 active lesions at Visit 1, 60 at Visit 2, and 58 at Visit 3. PGA-A: physician's global assessment of activity (of study lesion).

of skin thickening, representing induration in the active, inflammatory phase of LS, and fibrosis in the later damage stage<sup>36</sup>. We therefore suggest that skin thickening should be used in conjunction with more specific activity features when assessing disease activity level.

At the highest activity levels, the combination of erythema,

disease extension, violaceous color, abnormal skin texture, and skin thickening explained more than half of the variation in PGA-A in multiple regression models. More severe levels of lesion features were related to higher PGA-A scores. Erythema was the strongest predictor across all visits. The regression model was less successful in explaining the variation in PGA-A

Table 4. Regression model for physician's global assessment of activity (PGA-A) of study lesion.

Variable	Visit 1		Visit 2		Visit 3	
	Estimate (SE)	p	Estimate (SE)	p	Estimate (SE)	p
Intercept	-4.21 (4.16)	0.3145	1.32 (2.67)	0.6229	0.77 (2.51)	0.7610
Erythema, ref = none						
Mild	11.85 (3.57)	<b>0.0014</b>	10.32 (3.01)	<b>0.0010</b>	14.18 (3.00)	<b>&lt; 0.0001</b>
Moderate	28.13 (5.04)	<b>&lt; 0.0001</b>	14.96 (8.28)	0.0750	32.59 (8.63)	<b>0.0003</b>
Disease extension	12.18 (3.38)	<b>0.0005</b>	9.72 (3.57)	<b>0.0082</b>	12.20 (5.54)	<b>0.0312</b>
Violaceous color, ref = none						
Mild	13.03 (4.56)	<b>0.0054</b>	-0.20 (3.71)	0.9579	5.69 (3.78)	0.1370
Moderate	19.18 (9.18)	<b>0.0397</b>				
STmax, ref = none						
Mild	4.85 (4.19)	0.2497	6.07 (3.19)	0.0612	5.20 (3.06)	0.0940
Moderate	5.83 (4.10)	0.1586	9.15 (3.77)	<b>0.0178</b>	10.42 (3.65)	<b>0.0057</b>
Severe	22.38 (8.10)	<b>0.0071</b>	1.79 (12.02)	0.8822		
Abnormal skin texture	10.56 (4.57)	<b>0.0234</b>	2.37 (3.20)	0.4611	0.38 (2.99)	0.8997
Adj R <sup>2</sup>	0.5881		0.3254		0.4934	
F	15.12		5.76		11.57	
N	90		80		77	

General linear regression and incremental model building were performed to identify the best model for predicting PGA-A of the study lesion. The best model had erythema, disease extension, violaceous color, skin thickening, and abnormal skin texture. For each visit, the estimates (regression coefficients) are shown of the lesion features and their p values, and adjusted R<sup>2</sup> values and F-tests for the model. Significant p values are in bold face. Adj R<sup>2</sup>: adjusted R<sup>2</sup>; disease extension: new lesion or larger lesion size; F: F-test; N: no. subjects analyzed; STmax: maximum skin thickening.

scores for the second visit compared to the first and third.

Limitations of our study include that there is no gold standard for activity. We relied on PGA scores, as has been done for the development of other rheumatic disease measures<sup>34,35,37,38,39</sup>. While the global assessments of activity from all physicians clearly differentiated active from inactive subjects, physicians may have differed in their weighing of certain features at different activity levels. This may have contributed to loss of significance of some features in regression analysis at later visits, especially for uncommon features such as violaceous color and tactile warmth. Given their specificity, further study of these features for assessing activity is warranted.

Poorer regression performance at later visits may be related to our scoring methodology. The abnormal skin texture feature was intended to capture waxy lesions, but our definition could have led us to include some types of dermal atrophy. At later visits, the frequency of waxy lesions, unlike that for dermal atrophy, could have declined in response to treatment, resulting in poorer model performance. For future studies, more specific descriptors should be tested.

Other potential scoring problems were our scoring for the worst rather than average level of each feature in the study lesion, which may have led to large stepwise drops in scores rather than gradual decreases, thereby limiting our ability to detect intermediate levels of activity. We did not assess the overall extent of active skin involvement, nor did we use disease extent as a score multiplier as is done in the Psoriasis Area and Severity Index measure<sup>40,41</sup>. These decisions may have produced lower model performance at later visits.

Our study provides important information for clinicians

caring for patients with pediatric LS, including the lack of a universal activity feature and lack of specificity of skin thickness for active disease. Accurate identification and assessment of activity requires the evaluation of several lesion features. Damage, both atrophy and dyspigmentation, was found to commonly coexist with active disease even at study entry. The presence of damage features should therefore not be interpreted as indicating inactive disease. Because relapses are common, ongoing careful assessment of these patients for disease activity is an essential part of clinical care. The findings from this study should facilitate development of a sensitive, specific, and responsive activity tool. Such a tool should improve care, enable controlled treatment trials to be conducted, and improve longterm outcomes for this often severely damaging disease.

## ACKNOWLEDGMENT

We thank Drs. Thaschawee Arkachaisri and Thomas Medsger for providing formal instruction in methodology for assessing skin thickness at our pre-study meetings. We thank Eseng Lai for help with data analysis and manuscript preparation, Benjamin Li for title suggestions, and Mary Ellen Riordan and Justine Zasa for their help in conducting our study.

## REFERENCES

1. Fett N, Werth VP. Update on morphea: part I. Epidemiology, clinical presentation, and pathogenesis. *J Am Acad Dermatol* 2011; 64:217-28.
2. Zulian F, Vallongo C, Woo P, Russo R, Ruperto N, Harper J, et al. Localized scleroderma in childhood is not just a skin disease. *Arthritis Rheum* 2005;52:2873-81.
3. Vasquez R, Sendejo C, Jacobo H. Morphea and other localized forms of scleroderma. *Curr Opin Rheumatol* 2012;24:685-93.
4. Peterson L, Nelson A, Su W, Mason T, O'Fallon W, Gabriel S. The epidemiology of morphea (localized scleroderma) in Olmstead



- County 1960-1993. *J Rheumatol* 1997;24:73-80.
5. Marzano AV, Menni S, Parodi A, Borghi A, Fuligni A, Fabbri P, et al. Localized scleroderma in adults and children. Clinical and laboratory investigations on 239 cases. *Eur J Dermatol* 2003;13:171-6.
6. Saxton-Daniels S, Jacobe HT. An evaluation of long-term outcomes in adults with pediatric-onset morphea. *Arch Dermatol* 2010;146:1044-5.
7. Condie D, Grabell D, Jacobe H. Comparison of outcomes in adults with pediatric-onset morphea and those with adult-onset morphea: a cross-sectional study from the morphea in adults and children cohort. *Arthritis Rheumatol* 2014;66:3496-504.
8. Li SC, Liebling MS, Haines KA, Weiss JE, Prann A. Initial evaluation of an ultrasound measure for assessing the activity of skin lesions in juvenile localized scleroderma. *Arthritis Care Res* 2011;63:735-42.
9. Piram M, McCuaig CC, Saint-Cyr C, Marcoux D, Hatami A, Haddad E, et al. Short- and long-term outcome of linear morphoea in children. *Br J Dermatol* 2013;169:1265-71.
10. Wu EY, Rabinovich EC, Torok KS, Li SC, Fuhlbrigge RC, CARRAnet Investigators. Description of the localized scleroderma subgroup of the CARRAnet registry [abstract]. *Arthritis Rheumatol* 2011;63 Suppl 10:S787.
11. Diaz-Perez JL, Connolly SM, Winkelmann RK. Disabling pansclerotic morphea of children. *Arch Dermatol* 1980;116:169-73.
12. Sathornsumetee S, Schanberg L, Rabinovich E, Lewis D Jr., Weisleder P. Parry-Romberg syndrome with fatal brain stem involvement. *J Pediatr* 2005;146:429-31.
13. Zulian F, Martini G, Vallongo C, Vittadello F, Falcini F, Patrizi A, et al. Methotrexate treatment in juvenile localized scleroderma: a randomized, double-blind, placebo-controlled trial. *Arthritis Rheum* 2011;63:1998-2006.
14. Li SC, Torok KS, Pope E, Dedeoglu F, Hong S, Jacobe HT, et al. Development of consensus treatment plans for juvenile localized scleroderma: a roadmap toward comparative effectiveness studies in juvenile localized scleroderma. *Arthritis Care Res* 2012;64:1175-85.
15. Kreuter A, Krieg T, Worm M, Wenzel J, Moynadeh P, Kuhn A, et al. German guidelines for the diagnosis and therapy of localized scleroderma. *J Dtsch Dermatol Ges* 2016;14:199-216.
16. Li SC, Feldman BM, Higgins GC, Haines KA, Punaro MG, O'Neil KM. Treatment of pediatric localized scleroderma: results of a survey of North American pediatric rheumatologists. *J Rheumatol* 2010;37:175-81.
17. Mirsky L, Chakkittakandiyil A, Laxer RM, O'Brien C, Pope E. Relapse after systemic treatment in paediatric morphoea. *Br J Dermatol* 2012;166:443-5.
18. Zulian F, Vallongo C, Patrizi A, Belloni-Fortina A, Cutrone M, Alessio M, et al. A long-term follow-up study of methotrexate in juvenile localized scleroderma (morphea). *J Am Acad Dermatol* 2012;67:1151-6.
19. Weibel L, Sampaio MC, Visentin MT, Howell KJ, Woo P, Harper JJ. Evaluation of methotrexate and corticosteroids for the treatment of localized scleroderma (morphea) in children. *Br J Dermatol* 2006;155:1013-20.
20. Mertens JS, Seyger MM, Kievit W, Hoppenreijns EP, Jansen TL, van de Kerkhof PC, et al. Disease recurrence in localized scleroderma: a retrospective analysis of 344 patients with paediatric- or adult-onset disease. *Br J Dermatol* 2015;172:722-8.
21. Pequet MS, Holland KE, Zhao S, Drolet BA, Galbraith SS, Siegel DH, et al. Risk factors for morphoea disease severity: a retrospective review of 114 paediatric patients. *Br J Dermatol* 2014;170:895-900.
22. Zulian F. New developments in localized scleroderma. *Curr Opin Rheumatol* 2008;20:601-7.
23. Horger M, Fierlbeck G, Kuemmerle-Deschner J, Tzaribachev N, Wehrmann M, Claussen CD, et al. MRI findings in deep and generalized morphea (localized scleroderma). *Am J Roentgenol* 2008;190:32-9.
24. Shaw LJ, Shipley J, Newell EL, Harris N, Clinch JG, Lovell CR. Scanning laser Doppler imaging may predict disease progression of localized scleroderma in children and young adults. *Br J Dermatol* 2013;169:152-5.
25. Wortsman X, Wortsman J, Sazunic I, Carreno L. Activity assessment in morphea using color Doppler ultrasound. *J Am Acad Dermatol* 2011;65:942-8.
26. Moore TL, Vij S, Murray AK, Bhushan M, Griffiths CE, Herrick AL. Pilot study of dual-wavelength (532 and 633 nm) laser Doppler imaging and infrared thermography of morphoea. *Br J Dermatol* 2009;160:864-7.
27. Seyger MM, van den Hoogen FH, de Boo T, de Jong EM. Low-dose methotrexate in the treatment of widespread morphea. *J Am Acad Dermatol* 1998;39:220-5.
28. Cunningham BB, Landells ID, Langman C, Sailer DE, Paller AS. Topical calcipotriene for morphea/linear scleroderma. *J Am Acad Dermatol* 1998;39:211-15.
29. Arkachaisri T, Pino S. Localized scleroderma severity index and global assessments: a pilot study of outcome instruments. *J Rheumatol* 2008;35:650-7.
30. Arkachaisri T, Vilaiyuk S, Li S, O'Neil KM, Pope E, Higgins GC, et al. The localized scleroderma skin severity index and physician global assessment of disease activity: a work in progress toward development of localized scleroderma outcome measures. *J Rheumatol* 2009;36:2819-29.
31. Zulian F, Meneghesso D, Grisan E, Vittadello F, Belloni Fortina A, Pigozzi B, et al. A new computerized method for the assessment of skin lesions in localized scleroderma. *Rheumatology* 2007;46:856-60.
32. Torok KS, Arkachaisri T. Methotrexate and corticosteroids in the treatment of localized scleroderma: a standardized prospective longitudinal single-center study. *J Rheumatol* 2012;39:286-94.
33. Laxer R, Zulian F. Localized scleroderma. *Curr Opin Rheumatol* 2006;18:606-13.
34. Anderson JJ, Chernoff MC. Sensitivity to change of rheumatoid arthritis clinical trial outcome measures. *J Rheumatol* 1993;20:535-7.
35. Buchbinder R, Bombardier C, Yeung M, Tugwell P. Which outcome measures should be used in rheumatoid arthritis clinical trials? Clinical and quality-of-life measures' responsiveness to treatment in a randomized controlled trial. *Arthritis Rheum* 1995;38:1568-80.
36. Fleischmajer R, Gay S, Meigel WN, Perlsh JS. Collagen in the cellular and fibrotic stages of scleroderma. *Arthritis Rheum* 1978;21:418-28.
37. Uribe AG, Vila LM, McGwin G Jr., Sanchez ML, Reveille JD, Alarcon GS. The Systemic Lupus Activity Measure-revised, the Mexican Systemic Lupus Erythematosus Disease Activity Index (SLEDAI), and a modified SLEDAI-2K are adequate instruments to measure disease activity in systemic lupus erythematosus. *J Rheumatol* 2004;31:1934-40.
38. Whiting-O'Keefe QE, Stone JH, Hellmann DB. Validity of a vasculitis activity index for systemic necrotizing vasculitis. *Arthritis Rheum* 1999;42:2365-71.
39. Eberl G, Studnicka-Benke A, Hitzelhammer H, Gschnait F, Smolen JS. Development of a disease activity index for the assessment of reactive arthritis (DAREA). *Rheumatology* 2000;39:148-55.
40. Schmitt J, Wozel G. The psoriasis area and severity index is the adequate criterion to define severity in chronic plaque-type psoriasis. *Dermatology* 2005;210:194-9.
41. Fredriksson T, Pettersson U. Severe psoriasis-oral therapy with a new retinoid. *Dermatologica* 1978;157:238-44.