The Localized Scleroderma Skin Severity Index and Physician Global Assessment of Disease Activity: A Work in Progrss Toward Development of Localized Scleroderma Outcome Measures

THASCHAWEE ARKACHAISRI, SOAMARAT VILAIYUK, SUZANNE LI, KATHLEEN M. O'NEIL, ELENA POPE, GLORIA C. HIGGINS, MARILYNN PUNARO, EGLA C. RABINOVICH, MARGALIT ROSENKRANZ, DANIEL A. KIETZ, PAUL ROSEN, STEVEN J. SPALDING, TERESA R. HENNON, KATHRYN S. TOROK, ELAINE CASSIDY, and THOMAS A. MEDSGER Jr, for the Localized Scleroderma Clinical and Ultrasound Study Group (LOCUS)

ABSTRACT. ***Objective.*** To develop and evaluate a Localized Scleroderma (LS) Skin Severity Index (LoSSI) and global assessments' clinimetric property and effect on quality of life (QOL).

Methods. A 3-phase study was conducted. The first phase involved 15 patients with LS and 14 examiners who assessed LoSSI [surface area (SA), erythema (ER), skin thickness (ST), and new lesion/extension (N/E)] twice for inter/intrarater reliability. Patient global assessment of disease severity (PtGA-S) and Children's Dermatology Life Quality Index (CDLQI) were collected for intrarater reliability evaluation. The second phase was aimed to develop clinical determinants for physician global assessment of disease activity (PhysGA-A) and to assess its content validity. The third phase involved 2 examiners assessing LoSSI and PhysGA-A on 27 patients. Effect of training on improving reliability/validity and sensitivity to change of the LoSSI and PhysGA-A was determined.

Results. Interrater reliability was excellent for ER [intraclass correlation coefficient (ICC) 0.71], ST (ICC 0.70), LoSSI (ICC 0.80), and PhysGA-A (ICC 0.90) but poor for SA (ICC 0.35); thus, LoSSI was modified to mLoSSI. Examiners' experience did not affect the scores, but training/practice improved reliability. Intrarater reliability was excellent for ER, ST, and LoSSI (Spearman's rho = 0.71–0.89) and moderate for SA. PtGA-S and CDLQI showed good intrarater agreement (ICC 0.63 and 0.80). mLoSSI correlated moderately with PhysGA-A and PtGA-S. Both mLoSSI and PhysGA-A were sensitive to change following therapy.

Conclusion. mLoSSI and PhysGA-A are reliable and valid tools for assessing LS disease severity and show high sensitivity to detect change over time. These tools are feasible for use in routine clinical practice. They should be considered for inclusion in a core set of LS outcome measures for clinical trials. (J Rheumatol First Release Oct 15 2009; doi:10.3899/jrheum.081284)

Key Indexing Terms:

LOCALIZED SCLERODERMA   OUTCOME MEASURE
SKIN SCORES   GLOBAL ASSESSMENT   QUALITY OF LIFE

Localized scleroderma (LS) is a group of uncommon, presumably autoimmune disorders primarily affecting skin and deeper structures including subcutaneous tissue and, rarely, muscle and bone[1]. The incidence of LS is estimated to be 2.7 cases per 100,000 persons at risk/year[2]. Based on morphologic appearance and levels of tissue involvement, LS is subdivided into morphea — focal or generalized and others (deep or subcutaneous morphea, morphea profunda) — and linear scleroderma, which includes the *en coup de sabre* variant affecting the scalp and face[3]. LS is not a fatal disease, but many children with LS suffer from emotional (cosmetic disfigurement) and/or physical sequelae (joint contractures, localized growth failure)[2,4,5].

To date, there has not been a reliable and standardized outcome measure for LS; thus, the development and evaluation of current and future therapies has been hampered. Sophisticated tools including thermography, ultrasound, and laser Doppler flowmetry assessing LS disease activity have been recommended, but these require considerable time, expense, and operator experience[6-10]. Histopathologic changes (skin biopsy) accurately reflect LS disease stages but are limited by sampling bias, and repeated biopsies are inconvenient and not well accepted by patients. Development of a reliable clinical tool to measure skin disease activity would facilitate clinical trials and inform treatment decisions.

We recently developed and published a preliminary, semiquantitative clinical skin score to assess LS severity, the LS Skin Severity Index (LoSSI)[11]. In the present study, we assess the LoSSI on 18 cutaneous anatomic sites (instead of 14 anatomic sites in the original version, by separating the hands and feet out of forearms and legs, respectively). This approach will increase the sensitivity of the LoSSI in assessing the extent of disease severity. We intended to evaluate LoSSI reliability more precisely by using a larger group of examiners from different institutions, to assess the validity and responsiveness of the LoSSI and to eliminate any domain from the measure that showed poor reliability. We also formulated the physician global assessment of disease activity clinical variables and assessed its reliability, responsiveness, and content validity in our study. Further, quality of life (QOL) measurement has been studied in many skin diseases and the generic dermatology QOL measure, Children's Dermatology Life Quality Index (CDLQI), has been used and validated in childhood skin diseases[12-14]. We explored the QOL in our patients with LS using the generic CDLQI and assessed its reliability.

MATERIALS AND METHODS

Study design. The study was conducted in 3 phases as follows:
(1) First phase was a 1-day, 2-session study intending to assess the inter- and intrarater reliability of clinical skin scores (LoSSI), assess the intrarater reliability of patient global assessment of disease severity (PtGA-S), and explore the effect of LS on QOL using the CDLQI and assess its intrarater reliability. This phase involved 15 patients with LS and 14 examiners (rheumatologists and dermatologist). (2) Second phase was a study intending to obtain the Localized Scleroderma Clinical and Ultrasound Study Group (LOCUS) consensus for the development of physician global assessment of LS disease activity (PhysGA-A) clinical determinants, and to assess their content validity. This phase involved members of the LOCUS. (3) Third phase as follows: evaluate the effectiveness of training (involving demonstration, feedback, and discussion using real patients in our scleroderma clinic) on improving LoSSI interrater reliability; a pilot study on interrater reliability of PhysGA-A after its determinants were obtained from the second phase; assess the construct validity of LoSSI against PhysGA-A, PtGA-S, and CDLQI; and evaluate the sensitivity to change of the LoSSI. This phase involved 27 patients with LS and 2 examiners. There were 25/27 patients who had completed 2-timepoint data and were analyzed for LoSSI and PhysGA-A sensitivity to change.

Patients. Patients with LS were recruited from the Scleroderma Clinics at Children's Hospital of Pittsburgh and the University of Pittsburgh Medical Center. Diagnosis and classification were according to Peterson, *et al*[3]. The University of Pittsburgh Institutional Review Board approved the study.

Examiners. First phase: 14 examiners (9 rheumatologists, 1 dermatologist, and 4 rheumatology fellows) participated. Five senior rheumatologists had 23.4 + 9.3 years in practice, evaluating on average 72 LS patient visits per year (range 24–144, median 50). Four junior rheumatologists and a pediatric dermatologist had 5.8 ± 2.2 years of practice experience and had an average of 58 LS patient visits per year (range 6–144, median 15). All examiners had a 1 h slide presentation and training on clinical skin score assessment (LoSSI and its domains). Each patient was examined randomly and independently. For intrarater reliability, each examiner evaluated each patient twice (morning and afternoon) at least 3 h apart in order to minimize recall memory. Patients or their caregivers (if < 8 yrs old) also completed PtGA-S and CDLQI twice — in the morning and afternoon sessions.

Second phase: 8 rheumatology/dermatology attendings (members of LOCUS: TA, SL, KMO, MP, GCH, ECR, EP, Ronald Laxer, and TAM) suggested and ranked the importance of clinical and laboratory variables to formulate the PhysGA-A determinants. The physicians participating in these surveys had, on average, 19.9 years in practice (range 8.0–40.0, median 18), evaluating, on average, 88 LS patient visits per year (range 24.0–144.0, median 108).

Third phase: 27 patients with LS were evaluated by 2 independent examiners, a rheumatologist (TA) and a clinical fellow (SV), at the same clinic visit, 1 month after the first phase, without referring to the previous scores. The LoSSI and PhysGA-A were obtained by the examiners and PtGA-S and CDLQI were completed by patients. There were 25/27 patients (14 active disease and 11 inactive disease stages) in this cohort who had complete LS skin scores and PhysGA-A assessed by 1 of the authors (TA) at 2 timepoints, which were used to analyze the sensitivity to change. Active disease was defined as lesions with erythema or enlargement of existing lesion/new lesion development within 1 month. Stable disease without any changes of lesions for > 3 months was defined as inactive disease.

Clinical skin scoring. The LoSSI was designed to be simple, quick, and easy to score, using the information obtained from a brief review of a patient's clinical history and by cutaneous examination. The LoSSI is calculated by summing 4 domain scores based on the extent (surface area: enlargement of existing lesions and new lesion development) and intensity (erythema and skin thickness) of the disease in 18 cutaneous surface anatomic sites (head, neck, chest, abdomen, upper back, lower back, right and left — arms, forearms, hands/fingers, buttocks/thighs, legs and feet)[11].
1. Surface area score (SA): The extent of surface area involvement within each anatomic site was scored from 0 to 3, where 0 = no involvement; 1 = 1/3, 2 = > 1/3–2/3, and 3 = > 2/3–3/3 of surface area of site affected. This was obtained by simple "eyeball" estimation of the whole circumference of the given limbs or entire surface of the anatomic sites of the trunk.
2. Erythema (ER): The degree of erythema at the edge of a lesion was scored 0–3 (0 = normal or postinflammatory hyper/hypopigmentation, 1 =

slight erythema/pink, 2 = red/clearly erythema, and 3 = violaceous/marked erythema).

3. Skin thickness score (ST): We adopted the modified Rodnan skin thickness system as follows: 0 = normal; 1 = mild increase in thickness; 2 = moderate increase in thickness, difficult to move skin; 3 = severe thickness, unable to move skin. ST was determined at the edge of a lesion and compared to the unaffected contralateral, or nearby ipsilateral skin if symmetrical lesions were present, thus minimizing inter-subject variability.

4. New lesion/lesion extension (N/E) — A new lesion and/or enlargement of an existing lesion within the past month was scored 3.

The most severe or highest score of each domain (SA, ER, ST, and N/E) of lesions in a given anatomic site are summed to obtain the LoSSI (range 0–216). For example, a patient has 2 lesions on the abdomen (1 anatomic site) — an old lesion (A) is scored SA1, ER0, ST2, N/E0; and a new lesion (B) detected 3 weeks ago is scored SA1, ER2, ST1, N/E3 — thus LoSSI at this first visit would be $1 + 2 + 2 + 3 = 8$. At a followup visit 2 months later, there are no new lesions or enlargement of the original lesions. Lesion A is scored SA1, ER0, ST1, N/E0 and lesion B has SA1, ER1, ST1, N/E0. Thus LoSSI would be $1 + 1 + 1 + 0 = 3$.

Global assessment using a 100-mm visual analog scale (VAS).
1. PhysGA-A, anchors being "not active" at the 0 point and "very active" at the 100 point. There were 2 aspects for the PhysGA-A study.

The first aspect (happened in the second phase of the study) was to determine the clinical and laboratory measures that give high content validity as assessed by the LOCUS survey consensus. Disease activity was defined as the extent and severity of reversible manifestations, both cutaneous and extracutaneous, due to underlying disease. Physicians suggested the variables (18 clinical signs and 11 laboratory tests) and ranked them on a 0–4 scale (0 = unimportant, 1 = minimally important, 2 = moderately important, 3 = very important, 4 = extremely important). The variables that yielded high content validity were then used to assess PhysGA-A.

The second aspect (happened in the third phase of the study) was to assess the PhysGA-A reliability and construct validity against the clinical skin scores.

2. Patient global assessment of disease severity (PtGA-S-S), anchors being "not severe" at the 0 point and "very severe" at the 100 point, over the past month, completed by patients ≥ 8 years old or by an accompanying parent.

Dermatology Life Quality Index. CDLQI is a reliable and validated QOL measure developed for use in children with skin diseases[12-14]. The CDLQI consists of 10 questions regarding how skin disease has affected the patient's quality of life over the past 1 week, in each of 10 domains, with 4 possible responses graded 0–3 (range of scores 0–30)[14]. For LS, we modified a question on symptoms to read "How TIGHT has your skin been?". We used CDLQI only for our 13 pediatric LS patients in the first phase and 27 in the third phase of the study. This tool can be found on the website http://www.dermatology.org.uk.

Statistical analysis. All analyses were performed using statistical programs SPSS v.15 (SPSS Inc., Chicago, IL, USA) and Stata v.8 (Stata Corp., College Station, TX, USA). Mean and standard deviation (SD) or median and interquartile range (IQR) were used to describe data where appropriate.

Interrater reliability was determined by using intraclass correlation coefficients (ICC; continuous or categorical outcomes from > 2 examiners) or percentage raw agreement and weighted kappa (κw) statistics (categorical outcomes from 2 examiners)[15-17]. Intrarater reliability was assessed by κw and percentage raw agreement for each examiner's paired data and Spearman's rho (r_s) for overall reliability[17]. Interpretation of agreement followed the recommendation of Landis and Koch: 0.00–0.20 = slight, 0.21–0.40 = fair, 0.41–0.60 = moderate, 0.61–0.80 = substantial, and > 0.80 = almost perfect agreement[15]. The Wilcoxon signed-rank test was used to assess differences between clinical skin scores and GA obtained from each rater group.

Content validity concerns the comprehensiveness of items for a given construct. This was assessed by using mean ranked, percentage consensus agreement of each variable and content validity index (CVI). Item-level

CVI was computed as the number of experts giving a rating of moderate to extremely important (score 2–4) divided by total number of experts, and scale-level CVI was the mean of item-level CVI included in the scale[18]. Modified kappa statistic (κ^*) corrected for chance agreement was also calculated for both CVI as follows: $\kappa^* = (CVI - p_c)/(1 - p_c)$ where $p_c = [N!/A!(N-A)!]*0.5^N$ (N = number of experts and A = number agreeing on good relevance)[18]. Item-level CVI ≥ 0.78 and scale-level CVI ≥ 0.9 were taken as evidence of excellent content validity[18]. For modified kappa, $\kappa^* > 0.74$ was taken as excellent agreement[18].

In LS, there is no "gold standard" against which to test the validity of the LoSSI, and therefore it is impossible to assess criterion validity. For this reason, convergent construct validity was investigated[19]. LoSSI, PhysGA-A, PtGA-S, and CDLQI correlations were assessed by r_s ($r_s > 0.7$ were considered strong, 0.4 to < 0.7 moderate, and < 0.4 weak). We predicted that the LoSSI would correlate moderately with PhysGA-A and PtGA-S since the latter 2 outcome measures also assess other constructs (noncutaneous disease activity and psychological concern due to cosmetic effects of the disease). Correlation of the LoSSI with CDLQI was predicted to be low, since CDLQI is a measure primarily assessing QOL or psychosocial influence of skin disease. PhysGA-A was predicted to correlate moderately with PtGA-S, as patients and their families have different perceptions regarding the degree or severity of the physical illness and its psychological effects, and was predicted to correlate poorly with CDLQI since they do not measure the same construct entirely (disease activity and psychological effects vs QOL alone). PtGA-S should correlate moderately or weakly with CDLQI. Agreement of the observed r_s with these predictions was taken as evidence of construct validity[20].

Sensitivity to change (responsiveness). Standardized response mean (SRM) and standardized effect size (SES) were used to detect clinical change[20,21]. SRM is similar to a paired t-test. However, since it avoids the use of standard error of the mean as the denominator, it is less influenced by sample size[21-23]. SRM is considered large (> 0.8), moderate (> 0.5), or small (> 0.2)[21]. The 95% confidence interval (CI) for the SRM was calculated by the method described by Beaton, *et al*[21]. The Wilcoxon signed-rank test was used to demonstrate the difference between baseline and followup LoSSI and PhysGA-A.

All statistical tests were 2-sided, and a p value < 0.05 was considered to be significant.

RESULTS

Reliability. Patient characteristics: Table 1 summarizes the characteristics of the 15 patients with LS involved in the first phase of the study. There were 11 female and 4 male patients [2 morphea (M), 5 linear scleroderma (L), 5 mixed LS (M + L), 1 *en coup de sabre* (E), 1 subcutaneous morphea (SqM), and 1 generalized morphea (GM)]. All were Caucasian. The median number of affected cutaneous sites was 4 (IQR 1–6). One patient had no treatment and 14 patients were receiving a variety of therapies.

Clinical skin scores: A total of 39 cutaneous surface anatomic sites were examined. Since only 4 lesions from 2 patients had mild erythema (ER1, see Figure 1: A1, B1-3), 20 color photographs of skin lesions taken from patients attending our clinic and dermatology clinics were used to assess ER reliability. However, we calculated the LoSSI using patients' actual ER scores. Figure 1 shows the distribution of the mean and 95% CI for SA, ER, ST, and LoSSI. Twenty-two (56%) sites had lesions involving > 1/3 of the surface area involved. Using the photographs of skin lesions for comparison, the distribution of ER responses was almost

Table 1. Demographic and clinical characteristics of patients with localized scleroderma.

Patient	Age, yrs	Sex	Diagnosis*	Affected Area†	Duration of Disease, mo	Duration of Therapy, mo
1	14	F	M, L	L, T	25	24
2	7	F	E	F	9	3
3	10	M	SqM	T	17	13
4	14	F	L	L	36	27
5	14	F	L	L	32	32
6	24	F	M	L	48	45
7	12	F	M, L	L, T	96	24
8	17	F	M, L	L, T	26	22
9	15	M	M, L, PRS	F, L, T	30	22
10	9	F	L	L	15	8
11	9	M	M, L	L	20	5
12	6	F	L	L, T	54	46
13	17	F	L	L	36	22
14	20	F	M	T	31	3
15	53	M	GM	T	12	3

* M: plaque morphea; L: linear scleroderma; E: *en coup de sabre;* PRS: Parry-Romberg syndrome; GM: generalized morphea; SqM: subcutaneous morphea. † F: face/scalp; L: limb; T: trunk.

equal, with 40% having ER = 3 and 30% each ER = 1 and 2. For ST, 13% had no skin thickening, 67% 1+, and 10% each 2+ or 3+. We added the highest scores of each domain (SA, ER, ST, and N/E) from all anatomic sites of each patient together to obtain the patient's LoSSI. The mean LoSSI was 5.82 (SD 4.60) and the median was 4.42 (IQR 2.79–6.76).

Interrater reliability: As shown in Table 2, ER, ST, and LoSSI demonstrated substantial to excellent interrater agreement, but SA had only fair interrater agreement. Practice experience did not influence the ER scores. Although not significant, attendings (both senior and junior) had higher interrater reliability for SA and ST compared to that of fellows. Interrater reliability improved after 1 month of training and practice, as demonstrated by increase in reliability coefficients (κw and percentage raw agreement) between an attending rheumatologist and a fellow for ST, from 0.62 (87%) to 0.88 (98%); for ER from 0.74 (88%) to 0.85 (99%); and for overall LoSSI from 0.59 (86%) to 0.93 (98%).

Since SA had low interrater reliability and was not sensitive to change (see below), omitting SA from the LoSSI did not significantly impair overall LoSSI assessment [LoSSI: ICC 0.80 (95% CI 0.68, 0.91) vs LoSSI without SA: 0.70 (95% CI 0.53, 0.85)]. We therefore omitted SA, making the modified LS Skin Severity Index (mLoSSI), as the sum of ER, ST, and N/E scored in the 18 cutaneous anatomic sites (possible scores range 0–162).

Intrarater reliability. κw and percentage raw agreement are summarized in Table 3. Intrarater agreement between 2 sessions varied widely for the SA domain (κw 0.12–1.00, 76%–100%). Moderate to excellent intrarater agreements were found for ER (κw 0.57–1.00, 85%–100%) and ST (κw

0.55–0.88, 84%–97%). The LoSSI demonstrated excellent intrarater reliability (ICC 0.79–0.99, data not shown), as did the mLoSSI (ICC 0.78–0.99). Overall intrarater reliability between clinical skin scores obtained from the 2 sessions was moderate for SA (r_s = 0.51, p < 0.001), excellent for ER (r_s = 0.89, p < 0.001), substantial for ST (r_s = 0.71, p < 0.001), and excellent for LoSSI (r_s = 0.81, p < 0.001) and mLoSSI (r_s = 0.77, p < 0.001).

Global assessment. PtGA-S: The median PtGA-S was 28.0 (IQR 12.5–51.0) and the mean was 32.21 ± 22.52. Intrarater agreement was substantial (ICC 0.63, 95% CI 0.21, 0.86).

CDLQI: The median CDLQI was 3.0 (IQR 2.0–3.0) and the mean was 3.79 ± 2.61. Intrarater reliability was excellent (ICC 0.80, 95% CI 0.45, 0.93).

PhysGA-A: To determine which clinical variables were important in formulating the PhysGA-A, pediatric rheumatologists and dermatologists, members of the LOCUS, suggested and then ranked the importance of 18 clinical and 11 laboratory variables on a 0–4 scale in the second phase of the study (Table 4). Four clinical variables were considered to be very or extremely important (mean rank 3.7, 75% agreement in grades 3 or 4) and one variable was considered as moderately important (rank 2.6, 75% agreement in grades 2 or 3) in the determination of global assessment of LS disease activity. These same clinical variables also had excellent item-level content validity index (> 0.78) and modified kappa (κ* > 0.8). Nine variables were minimally or unimportant (mean rank 0.5, 75% agreement in grades 0 or 1) and 4 variable showed no consensus (mean rank 1.9, < 75% agreement within 1 rank) on their importance in assessing disease activity. Three clinical variables including arthritis, warmth at the center, and warmth at the border of a lesion had item-level CVI (κ*) of 0.75 (0.72), 0.86 (0.85),
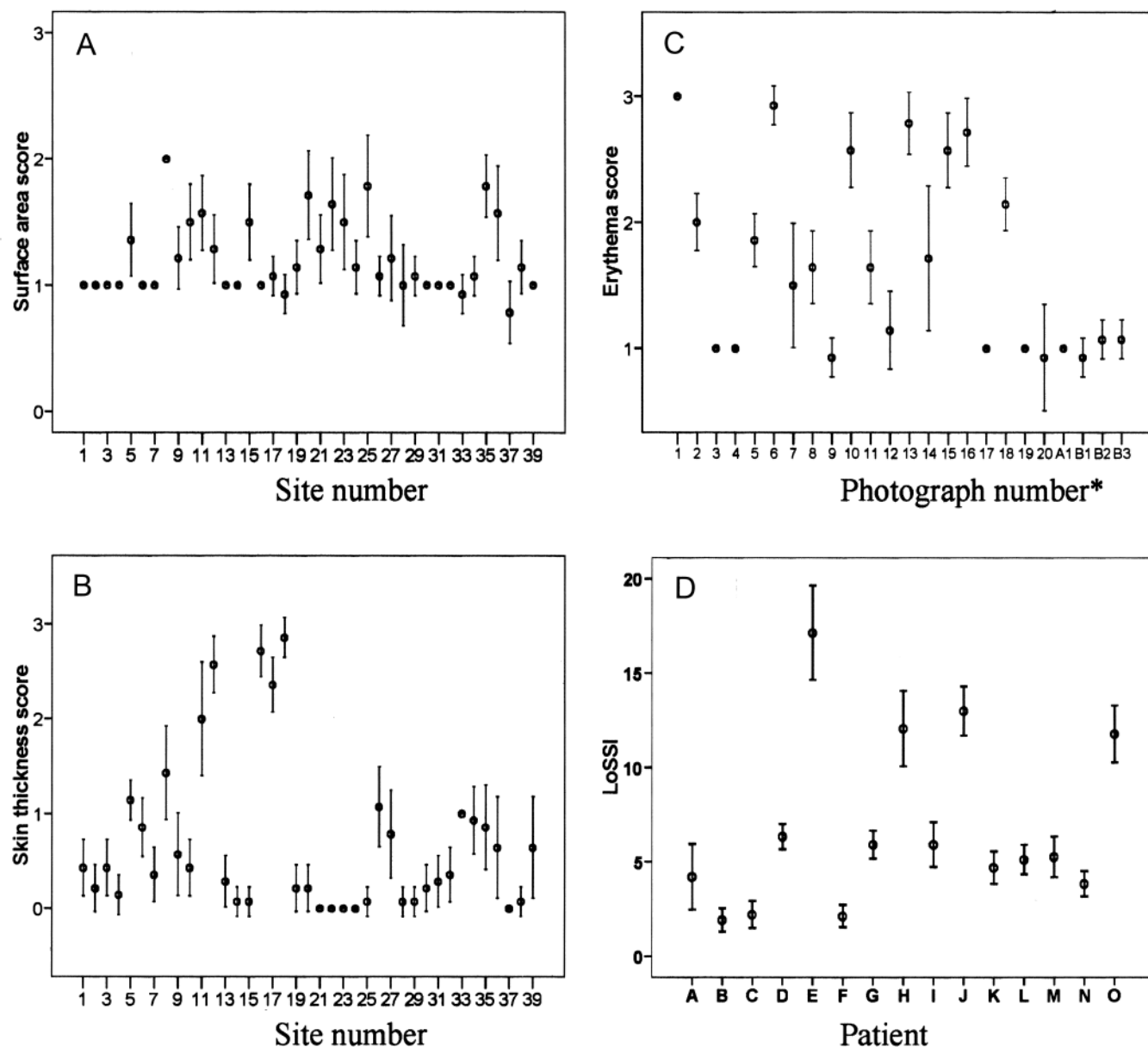
Figure 1. Error bar plot of mean and 95% confidence interval of the Localized Scleroderma Skin Severity Index (LoSSI) and its domains. Each domain [surface area (A) and skin thickness scores (B)] was graded from 0 to 3 by 14 examiners from 15 patients with 39 cutaneous anatomic sites. Erythema score (C) was graded from 0 to 3 by 14 examiners using 20 color photographs of skin lesions. *A1 and B1-B3 were scores from 2 patients with erythema score > 0. LoSSI (D) was assessed from 15 patients by 14 examiners.

and 0.86 (0.85), respectively, but there was no consensus agreement (< 75% agreement within 1 rank), thus they were not included as PhysGA-A clinical variables after our study group discussion. The clinical variables with moderate to extreme importance and item-level CVI ≥ 0.78/modified kappa (κ^*) > 0.74 were chosen to assess PhysGA-A. Elevated erythrocyte sedimentation rate (ESR) and C-reactive protein (CRP) had no consensus agreement and had excellent item-level content validity and κ^* [item-level CVI (κ^*) 1.00 (1.00) for both]. After discussion, both laboratory variables were included in the PhysGA-A assessment. PhysGA-A data were then obtained during the third phase of the study (using LOCUS consensus on PhysGA-A determinants — 5 clinical variables and 2 laboratory variables). The median of PhysGA-A was 1.0 (IQR 0.0–5.75, range 0.0–56.0). Interrater reliability was excellent (ICC 0.90, 95% CI 0.80, 0.96).

Validity. Patient characteristics: Twenty-seven patients with LS (19 female and 8 male; 2 M, 10 L, 8 M + L, 2 E, 4 SqM, and 1 GM) with 87 cutaneous anatomic sites were included

Table 2. Interrater reliability* of Localized Scleroderma Skin Severity Index (LoSSI) and its domains.

Examiners**	Surface Area	Erythema†	Skin Thickness	LoSSI
All examiners (n = 14)	0.35 (0.25, 0.49)	0.71 (0.58, 0.84)	0.70 (0.60, 0.80)	0.80 (0.68, 0.91)
Senior examiners (n = 5)	0.44 (0.29, 0.60)	0.68 (0.50, 0.84)	0.74 (0.62, 0.83)	0.83 (0.68, 0.93)
Junior examiners (n = 5)	0.38 (0.24, 0.55)	0.73 (0.57, 0.86)	0.72 (0.60, 0.82)	0.82 (0.67, 0.93)
Clinical fellows (n = 4)	0.24 (0.10, 0.42)	0.76 (0.58, 0.88)	0.65 (0.51, 0.78)	0.85 (0.68, 0.94)

* Assessed by intraclass correlation coefficient (ICC) and reported as ICC (95% confidence interval). ** Senior examiners have > 10 years, junior examiners ≤ 10 years in practice. † Erythema score reliability was assessed using 20 photographs of skin lesions.

Table 3. Intrarater reliability* of the modified Localized Scleroderma Skin Severity index (mLoSSI) and its domains including surface area domain. All domains were assessed by 14 examiners on 39 cutaneous anatomic sites. Intraclass correlation coefficient (95% confidence interval) was used to assess intrarater agreement.

Examiners	Surface Area	Erythema†	Skin Thickness	mLoSSI
1	0.89 (96.7)	0.95 (97.50)	0.87 (95.96)	0.99 (0.97, 1.00)
2	0.63 (95.16)	1.00 (100.00)	0.72 (91.92)	0.83 (0.55, 0.94)
3	0.44 (82.14)	0.95 (98.33)	0.71 (93.10)	0.93 (0.77, 0.98)
4	0.35 (89.58)	0.89 (95.00)	0.71 (89.39)	0.91 (0.72, 0.97)
5	0.29 (82.69)	0.57 (86.67)	0.58 (83.91)	0.91 (0.69, 0.97)
6	0.27 (87.88)	1.00 (100.00)	0.65 (90.91)	0.91 (0.74, 0.97)
7	0.60 (90.91)	0.79 (90.00)	0.65 (88.89)	0.85 (0.61, 0.95)
8	1.00 (100.00)	0.76 (90.00)	0.73 (91.23)	0.94 (0.75, 0.99)
9	0.15 (88.71)	0.68 (85.00)	0.62 (85.86)	0.83 (0.35, 0.95)
10	0.12 (81.25)	0.84 (82.50)	0.56 (84.52)	0.80 (0.45, 0.94)
11	0.47 (75.76)	0.73 (91.67)	0.63 (86.87)	0.91 (0.67, 0.97)
12	0.63 (93.33)	0.84 (95.00)	0.55 (88.33)	0.78 (0.41, 0.93)
13	0.52 (86.67)	1.00 (100.00)	0.64 (90.00)	0.90 (0.70, 0.97)
14	0.76 (95.45)	0.94 (97.50)	0.88 (96.97)	0.94 (0.83, 0.98)
Overall**	0.51	0.89	0.71	0.77

* Weighted kappa statistic (percentage raw agreement). ** Overall intrarater reliability calculated by Spearman's rho (r_s). † Erythema score reliability was assessed using 20 photographs of skin lesions.

in the third phase of the study. The median patient age was 13.0 years (IQR 8.0–14.0) and age at onset of LS was 8.0 years (IQR 5.0–12.0). The median disease duration was 26.0 months (IQR 19.0–42.0) and duration of therapy 14.0 months (IQR 3.0–25.0). Four patients had no therapy and 23 patients were on different treatment regimens. Of 87 cutaneous anatomic sites, 6 (7%) had ER > 0; 11 (13%), 7 (8%), and 5 (6%) had ST of 1+, 2+, and 3+, respectively. Three (3%) had new lesions. The median mLoSSI was 5.0 (IQR 2.8–7.3).

Content validity: As shown in Table 4, considering only cutaneous signs, ER and N/E were ranked as being very or extremely important, with item-level CVI of 1.00, and ST was ranked as moderately important (item-level CVI 0.88) in assessing LS skin activity. Scale-level and κ* were identical (0.97), suggesting that mLoSSI has excellent content validity.

Convergent construct validity: The correlation coefficient (r_s) results are summarized in Table 5. The findings

were consistent with our a priori predictions. PhysGA-A strongly correlated with PtGA-S (r_s = 0.81). mLoSSI correlated moderately with both global assessments, PhysGA-A (r_s = 0.49) and PtGA-S (r_s = 0.44), but weakly with CDLQI (r_s = 0.25). CDLQI correlated weakly with PhysGA-A (r_s = 0.15) and PtGA-S (r_s = 0.21), as predicted.

Sensitivity to change/ responsiveness. Patient characteristics: Table 6 shows patient demographic data and responsiveness statistics. Fourteen "active" (10 female, 4 male) and 11 "inactive" (8 female, 3 male) LS patients were included. Nine patients with active disease (lesions with erythema or enlargement of existing lesion/new lesion development within 1 month) had no therapy and the other 5 were taking methotrexate (MTX) and prednisone (1) or topical therapies (4). All inactive patients (stable disease without any changes of lesions for > 3 mo) were taking MTX. There was no significant difference in the number of affected cutaneous sites (active 3.14 ± 2.11, inactive 3.45 ± 1.69; p = 0.694).

Table 4. Clinical variables used in the physicians' determinations of global assessment of disease activity in localized scleroderma*. N = 8 physicians.

Clinical Variables with Consensus (≥75% agreement) in Assessing Disease Activity	Mean Rank (0-4)	Consensus Agreement, %	Content Validity Index (κ*)[†]
Very/extremely important			
New lesion within previous month	4.0	100	1.00 (1.00)
Enlargement of an existing lesion within previous month	3.9	100	1.00 (1.00)
Erythema/violaceous color at the border of a lesion	3.8	100	1.00 (1.00)
Uveitis	3.1	75	0.88 (0.88)
Moderately important			
Skin thickening/induration at border of a lesion	2.6	75	0.88 (0.88)
Mildly important/unimportant			
Dermal atrophy	0.8	75	0.25 (0.16)
Subcutaneous atrophy	0.8	75	0.25 (0.16)
Physical disability	0.8	80	0.20 (0.09)
Dyspigmentation (hyper/hypopigmentation)	0.6	75	0.25 (0.16)
Facial atrophy	0.4	100	0.00 (0.00)
Skeletal muscle atrophy	0.4	100	0.00 (0.00)
Psychosocial/quality of life impairment	0.4	100	0.00 (0.00)
Joint contracture	0.3	100	0.00 (0.00)
Cataract/glaucoma	0.0	100	0.00 (0.00)

* Clinical variables that were ranked, but with no consensus (< 75 consensus agreement and content validity index < 0.78) achieved on global assessment of localized scleroderma disease activity, are listed in rank order as follows: skin thickening/induration at the center of a lesion, absolute eosinophil count ≥ 300/mm^3, positive anti-ssDNA antibody, high level of anti-ssDNA antibody, positive antihistone antibody, high level of antihistone antibody, positive anti-Scl-70 antibody, positive antinuclear antibodies (ANA), high ANA level, elevated von Willebrand factor antigen. Clinical variables with no consensus (< 75% consensus agreement and content validity index > 0.78) achieved on global assessment of localized scleroderma disease activity are listed in rank order as follows: elevated erythrocyte sedimentation rate, elevated C-reactive protein, warmth at the border of a lesion, warmth at the center of a lesion, arthritis. [†] Item-level content validity index (K* = modified kappa).

Table 5. Summary of construct validity evaluation between different measures.

Pairs of Measures (n = 27)	Spearman's rho
mLoSSI	
Physician global assessment of disease activity	0.49*
Patient global assessment of disease severity	0.44**
CDLQI	0.25[†]
Physician global assessment of disease activity	
Patient global assessment of disease severity	0.81*
CDLQI	0.15[†]
Patient global assessment of disease severity	
CDLQI	0.21[†]

mLoSSI: modified Localized Scleroderma Skin Severity Index. CDLQI: Children's Dermatology Life Quality Index. * p < 0.001; ** p ≥ 0.001 and < 0.05; [†] p > 0.05.

Responsiveness: As shown in Table 6, the mean changes in mLoSSI and PhysGA-A were significant only in the active LS group. SRM as well as SES were large for both outcome measures in the patients with active LS compared with inactive patients. Thus the mLoSSI and PhysGA-A are capable of detecting change in patients whose clinical status improved after a median period of 3.5 months of therapy. This change was largely attributable to the change in ER (SRM 1.4) as compared to that of ST (SRM 0.4). PtGA-S was found to change significantly after therapy in the active LS group [n = 9, SES 2.16, SRM 1.04 (95% CI 1.57, 0.52)] as compared with the inactive LS group [n = 9, SES 0.57, SRM 0.68 (95% CI 1.27, 0.09)]. As expected, PtGA-S in patients with inactive disease had less PtGA-S change compared with active patients [mean: 6.73 (SD 9.89), p = 0.023, vs 41.92 (SD 40.26), p = 0.008].

DISCUSSION

To date, there is no simple, feasible, or reliable tool to assess LS skin changes. Thermography, ultrasound, and computerized image analysis have limited utility[7-10,24-28]. The applications of these instruments in assessing LS disease activity have been reviewed elsewhere[11].

The LoSSI was developed to measure the extent and intensity of inflammation occurring in the early phase of LS, during which effective therapy may halt disease progression. The LoSSI is simple, easy to use, and brief, making it feasible and suitable for use in a busy clinical setting and in clinical trials. We demonstrate that the LoSSI is a valid, highly reproducible cutaneous assessment tool for LS skin severity that has good sensitivity to change. We also found that PhysGA-A and PtGA-S are reliable and sensitive to change over time. As discussed below, we propose a modi-

Table 6. Responsiveness/sensitivity to change of mLoSSI and PhysGA-A.

Feature	Active, n = 14	Inactive, n = 11
Age, yrs median (IQR)	9.50 (7.75–12.25)	13.00 (9.00–15.00)
Diagnosis*	2 E, 2 SqM, 2L, 3 M, 5 M+L	1 E, 1 SqM, 2 GM, 3 L, 4 M+L
Duration of disease, mo, median (IQR)	19.50 (6.75–45.25)	33.00 (23.00–50.00)
Duration of therapy, mo, median (IQR)	3.50 (2.75–12.75)	19.00 (12.00–29.00)
mLoSSI		
Mean change (SD)	4.07 (3.02)	0.27 (0.65)
p for difference from baseline to followup (Wilcoxon signed-rank test)	< 0.001	0.50
SRM (95% CI)	1.35 (0.82, 1.87)	0.42 (–0.16, 1.01)
SES	0.95	0.33
PhysGA-A		
Mean change (SD)	45.00 (27.65)	0.00 (1.18)
p for difference from baseline to followup (Wilcoxon signed-rank test)	0.043	1.00
SRM (95% CI)	1.63 (0.75, 2.50)	0.00 (–0.59, 0.59)
SES	1.51	0.00

* E: *en coup de sabre*; M: morphea; SqM: subcutaneous morphea, GM: generalized morphea; L: linear sclero-derma. mLoSSI: modified Localized Scleroderma Skin Severity Index. PhysGA-A: physician global assessment of disease activity. SRM: standardized response mean (95% confidence interval) = mean observed change/standard deviation of the difference scores. SRM > 0.8 considered large, > 0.5 moderate, and > 0.2 small. SES: standardized effect size = mean observed change scores/standard deviation of baseline scores. SES > 0.8 considered large, > 0.5 moderate, and > 0.2 small. IQR: interquartile range.

fied version of the LoSSI (mLoSSI) in order to decrease interrater variability and increase the accuracy of this tool.

To increase the sensitivity of the LoSSI to detect extent of disease, we increased the number of cutaneous anatomic sites from 14 in our original report to 18 by separating the hands and fingers from the forearms and the feet and toes from the legs[11]. This change makes it more likely to identify physical and/or emotional disability compared with trunk involvement.

Visual assessment of erythema had moderate interrater variation ($\kappa w = 0.52$) according to Wolkerstorfer, *et al* (score 0–3, 3 examiners, 20 patients with atopic dermatitis)[29]. We found substantial interrater agreement (ICC 0.71) and almost perfect overall intrarater reliability for erythema ($r_s = 0.89$, $p < 0.001$), comparable to that of tender joint count ($\kappa w = 0.71$) in juvenile rheumatoid arthritis (JRA)[30]. In real patients, this domain had excellent sensitivity to change (SRM 1.4) in patients with active LS who responded to therapy.

The standard assessment of the degree of cutaneous induration or thickness in systemic sclerosis (SSc) by palpation is the modified Rodnan method (modified Rodnan skin score; mRSS)[31-33]. In LS, skin thickness represents cellular infiltration and edema during the early phase (disease activity) and excessive collagen deposition (disease damage) in the late phase[33-35]. In healthy children, Foeldvari and Wierk reported that mRSS varied with age and Tanner stage, with healthy children often having mRSS scores that would be considered pathological in adults[36]. To take account of this potential difference in children versus adults, we propose to assess ST in lesional versus unaffected skin (contralateral skin area or nearby ipsilateral area if symmetrical lesions are present). Thus patients serve as their own controls regardless of the Tanner stage and maturity, eliminating inter-subject variability. Our study showed substantial interrater agreement, comparable to that found for the mRSS in patients with SSc[37-39]. Experienced examiners had less interrater variability compared with inexperienced examiners, and there was improvement with training and practice, as also reported for the mRSS[38].

The reliability of the N/E domain can only be assessed in a clinical trial or a longitudinal study. We gave greater weight to this domain (score of 3) as this finding often prompts the physician to intensify therapy. Since the N/E score requires patient or caregiver recall memory, we used a 1-month period to decrease the variability and improve the accuracy of this domain score.

Estimating the extent of disease (surface area involved) in dermatologic conditions remains a challenge for clinical researchers because of poor interrater agreement[40-43]. Some LS lesions have indistinct margins, making the estimation of surface area very difficult. Grouping together multiple small lesions scattered over a given surface area requires extensive training and practice and is time-consuming. In a larger group of examiners, we experienced the same problem. Our surface area scores showed a high interrater variability (ICC 0.35) as reported by others studying atopic dermatitis and psoriasis[40-44]. Less experienced physicians had a higher interrater variability as compared to more experienced physicians[42]. Poor reliability of a domain can interfere with

overall reliability and validity of the outcome measure, as it impairs accuracy and consistency. As suggested by Finlay, recording of area involvement should be based on an assessment of site involvement rather than the virtually impossible task of determining an accurate total percentage involvement[41]. The location of skin affected may be more important than the percentage of involvement[45]. For example, a small linear lesion crossing the elbow causing elbow joint contracture is more disabling than a large lesion on the back where there is less motion and an insignificant cosmetic effect. In our method, the relative weight is skewed toward the exposed area and body zones with higher risk of physical disability, i.e., joint contractures, localized growth failure (66% limbs, 23% trunk, and 11% face/scalp/neck), similar to that used in psoriasis[42].

We propose to omit SA from LoSSI, which does not significantly reduce the overall inter- and intrarater reliability [LoSSI with SA vs LoSSI without SA (mLoSSI): ICC 0.80 vs 0.70 and $r_s = 0.81$ vs $r_s = 0.77$, respectively). Interrater variability of the mLoSSI in our study is comparable to that of the Cutaneous Lupus Erythematosus Disease Area and Severity Index (CLASI; ICC 0.86 for the activity score component), the "Six Area, Six Sign Atopic Dermatitis" severity score (SASSAD; ICC 0.70), and the JRA active joint count (ICC 0.69); and is better than the Cutaneous Assessment Tool (CAT) of juvenile idiopathic inflammatory myopathy (ICC 0.60 for activity score component) and the Dermatomyositis Skin Severity Index (DSSI; ICC 0.44)[30,46-48].

mLoSSI had excellent content validity as evidenced by high content validity index and high experts' consensus agreement at both the item and scale level. Five clinical variables (Table 4) and 2 laboratory variables (elevated ESR or CRP) were included in the assessment of PhysGA-A. Elevated ESR or CRP had less than three-quarters of consensus agreement but had excellent content validity index. From group discussion, both acute-phase reactants are nonspecific but they clearly reflect inflammation. When used for PhysGA-A assessment, one will need to be certain that these indices are not elevated secondary to other illnesses, especially infection. Arthritis is rare in LS, and in some patients could be a feature of other connective tissue diseases overlapping with LS. Therefore, arthritis may not truly reflect LS disease activity[49,50] and is not included in PhysGA-A assessment. Uveitis is the only noncutaneous clinical variable included as one of the PhysGA-A as it represents one of the common ocular involvements found in patients with LS, especially the *en coup de sabre* form. Uveitis represents a systemic autoimmune process and tends to appear early in the course of the disease[51]. Once detected, it always prompts physicians to modify the therapy, thus it is reasonable to be included as one of the PhysGA-A determinants. Moderate correlations between the mLoSSI and PhysGA-A, and the mLoSSI and PtGA-S provide evidence of convergent construct validity of the mLoSSI

instrument. A similar level of correlation between the clinical skin score (SCORAD) and PtGA-S was also reported in patients with atopic dermatitis[45]. Interestingly, we found excellent correlation between PhysGA-A and PtGA-S ($r_s = 0.81$, p < 0.001), higher than that we previously reported[11]. This could be the result of the removal of SA from the PhysGA-A determinants. It confirms our hypothesis that SA may not be important to patients' perception of disease severity.

Over a median period of 3.5 months, significant mean changes of mLoSSI and large SRM/SES (> 0.8) were found in the active LS compared with the inactive LS group. The same result was found for both PhysGA-A and PtGA-S. As in our pilot study[11], ER (SRM = 1.4) contributed strongly and ST (SRM = 0.4) contributed moderately to the overall mLoSSI sensitivity to change. The contribution of ST may be larger if a longer period of followup is used. As predicted, SA showed poor sensitivity to change (SRM = 0.3), confirming that SA is not a sensitive domain for discriminating LS activity.

The generic QOL measurement tool CDLQI was found to have excellent intrarater reliability (ICC 0.80) in patients with LS. The mean CDLQI was 3.8 (SD 2.6), higher than that of normal healthy children [0.4 (SD 0.7)], but lower than that in children with scabies [9.5 (SD 10.5)], atopic dermatitis [7.7 (SD 5.6)], psoriasis [5.4 (SD 5.0)], and acne [5.7 (SD 4.4)][14]. These preliminary CDLQI results suggest that LS has a relatively small effect on a patient's life quality[52] or that a LS-specific QOL tool may be needed. CDLQI correlated poorly with the mLoSSI, which measures different aspects of disease[53,54].

Our study is limited by the lack of a true "gold standard," the rarity of intensely inflamed lesions, and the possibility of recall bias. Proper weighting of ER and N/E, both indicative of active disease, should be reevaluated in the future. The reliability and sensitivity of the N/E domain can only be judged in a longitudinal study.

We developed and modified (mLoSSI) the first LS cutaneous assessment tool in order to facilitate clinical trial design and conduct and to monitor LS in the practice setting. This clinical scoring system fulfills the criteria for dermatologic outcome measures proposed by Finlay[41]. QOL did not correlate well with clinical skin scores or GA, but all are important and must be individually evaluated. The integrative assessment of these 3 areas represents a complementary approach that should help improve the holistic care of patients with LS. The mLoSSI, PhysGA-A, and PtGA-S have moderate to excellent reliability. mLoSSI is a valid tool and showed high sensitivity to clinically meaningful change after effective therapy. It is likely to enhance our understanding of the natural history of LS. This instrument should be included in a future core set of outcome measures for LS clinical trials. The next step in development of the LS cutaneous assessment tool should be to determine the mLoSSI

minimal clinically important difference, and to further validate the tools, i.e., mLoSSI and PhysGA-A, in a larger group of patients, which will require a multicenter prospective longitudinal study.

ACKNOWLEDGMENT

REFERENCES

1. Torres JE, Sanchez JL. Histopathologic differentiation between localized and systemic scleroderma. Am J Dermatopathol 1998;20:242-5.
2. Peterson LS, Nelson AM, Su WP, Mason T, O'Fallon WM, Gabriel SE. The epidemiology of morphea (localized scleroderma) in Olmsted County 1960-1993. J Rheumatol 1997;24:73-80.
3. Peterson LS, Nelson AM, Su WP. Classification of morphea (localized scleroderma). Mayo Clin Proc 1995;70:1068-76.
4. 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.
5. Zulian F, Athreya BH, Laxer R, Nelson AM, Feitosa de Oliveira SK, Punaro MG, et al. Juvenile localized scleroderma: clinical and epidemiological features in 750 children. An international study. Rheumatology 2006;45:614-20.
6. Birdi N, Shore A, Rush P, Laxer RM, Silverman ED, Krafchik B. Childhood linear scleroderma: a possible role of thermography for evaluation. J Rheumatol 1992;19:968-73.
7. Hoffmann K, El-Gammal S, Gerbaulet U, Schatz H, Altmeyer P. Examination of circumscribed scleroderma using 20-MHz B-scan ultrasound. In: Altmeyer P, El-Gammal S, Hoffmann K, editors. Ultrasound in dermatology. Berlin, Heidelberg: Springer-Verlag; 1992:231-43.
8. Li SC, Liebling MS, Haines KA. Ultrasonography is a sensitive tool for monitoring localized scleroderma. Rheumatology 2007;46:1316-9.
9. Martini G, Murray KJ, Howell KJ, Harper J, Atherton D, Woo P, et al. Juvenile-onset localized scleroderma activity detection by infrared thermography. Rheumatology 2002;41:1178-82.
10. Weibel L, Howell KJ, Visentin MT, Rudiger A, Denton CP, Zulian F, et al. Laser Doppler flowmetry for assessing localized scleroderma in children. Arthritis Rheum 2007;56:3489-95.
11. Arkachaisri T, Pino S. Localized scleroderma severity index and global assessments: a pilot study of outcome instruments. J Rheumatol 2008;35:650-7.
12. Finlay AY. Quality of life measurement in dermatology: a practical guide. Br J Dermatol 1997;136:305-14.
13. Holme SA, Man I, Sharpe JL, Dykes PJ, Lewis-Jones MS, Finlay AY. The Children's Dermatology Life Quality Index: validation of the cartoon version. Br J Dermatol 2003;148:285-90.
14. Lewis-Jones MS, Finlay AY. The Children's Dermatology Life Quality Index (CDLQI): initial validation and practical use. Br J Dermatol 1995;132:942-9.
15. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.
16. Lin LI. A concordance correlation coefficient to evaluate reproducibility. Biometrics 1989;45:255-68.
17. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420-8.
18. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Res Nurs Health 2007;30:459-67.
19. Carmines EG, Zeller RA. Reliability and validity assessment. Lewis-Beck MS, editor. Thousand Oaks, CA: Sage Publications, Inc.; 1979.
20. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. 3rd ed. New York: Oxford University Press; 2003.
21. Beaton DE, Hogg-Johnson S, Bombardier C. Evaluating changes in health status: reliability and responsiveness of five generic health status measures in workers with musculoskeletal disorders. J Clin Epidemiol 1997;50:79-93.
22. Katz JN, Larson MG, Phillips CB, Fossel AH, Liang MH. Comparative measurement sensitivity of short and longer health status instruments. Med Care 1992;30:917-25.
23. Liang MH, Larson MG, Cullen KE, Schwartz JA. Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. Arthritis Rheum 1985;28:542-7.
24. Cosnes A, Anglade MC, Revuz J, Radier C. Thirteen-megahertz ultrasound probe: its role in diagnosing localized scleroderma. Br J Dermatol 2003;148:724-9.
25. Seidenari S, Conti A, Pepe P, Giannetti A. Quantitative description of echographic images of morphea plaques as assessed by computerized image analysis on 20 MHz B-scan recordings. Acta Derm Venereol 1995;75:442-5.
26. Serup J. Ten years' experience with high-frequency ultrasound examination of the skin: development and refinement of technique and equipment. In: Altmeyer P, El-Gammal S, Hoffmann K, editors. Ultrasound in dermatology. Berlin, Heidelberg: Springer-Verlag; 1992:41-54.
27. Szymanska E, Nowicki A, Mlosek K, Litniewski J, Lewandowski M, Secomski W, et al. Skin imaging with high frequency ultrasound — preliminary results. Eur J Ultrasound 2000;12:9-16.
28. 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.
29. Wolkerstorfer A, de Waard van der Spek FB, Glazenburg EJ, Mulder PG, Oranje AP. Scoring the severity of atopic dermatitis: three item severity score as a rough system for daily practice and as a pre-screening tool for studies. Acta Derm Venereol 1999;79:356-9.
30. Guzman J, Burgos-Vargas R, Duarte-Salazar C, Gomez-Mora P. Reliability of the articular examination in children with juvenile rheumatoid arthritis: interobserver agreement and sources of disagreement. J Rheumatol 1995;22:2331-6.
31. Clements P, Lachenbruch P, Siebold J, White B, Weiner S, Martin R, et al. Inter and intraobserver variability of total skin thickness score (modified Rodnan TSS) in systemic sclerosis. J Rheumatol 1995;22:1281-5.
32. Clements PJ, Lachenbruch PA, Seibold JR, Zee B, Steen VD, Brennan P, et al. Skin thickness score in systemic sclerosis: an assessment of interobserver variability in 3 independent studies. J Rheumatol 1993;20:1892-6.
33. Rodnan GP, Lipinski E, Luksick J. Skin thickness and collagen content in progressive systemic sclerosis and localized scleroderma. Arthritis Rheum 1979;22:130-40.
34. Furst DE, Clements PJ, Steen VD, Medsger TA Jr, Masi AT, D'Angelo WA, et al. The modified Rodnan skin score is an accurate reflection of skin biopsy thickness in systemic sclerosis. J Rheumatol 1998;25:84-8.

35. Xie Y, Zhang X, Wakasugi S, Makino T, Inoue Y, Ihn H. Immunohistochemical characterization of the cellular infiltrate in localized scleroderma. Int J Dermatol 2008;47:438-42.

36. Foeldvari I, Wierk A. Healthy children have a significantly increased skin score assessed with the modified Rodnan skin score. Rheumatology 2006;45:76-8.

37. Brennan P, Silman A, Black C, Bernstein R, Coppock J, Maddison P, et al. Reliability of skin involvement measures in scleroderma. The UK Scleroderma Study Group. Br J Rheumatol 1992;31:457-60.

38. Czirjak L, Nagy Z, Aringer M, Riemekasten G, Matucci-Cerinic M, Furst DE. The EUSTAR model for teaching and implementing the modified Rodnan skin score in systemic sclerosis. Ann Rheum Dis 2007;66:966-9. Epub 2007 Jan 18.

39. Silman A, Harrison M, Brennan P. Is it possible to reduce observer variability in skin score assessment of scleroderma? The ad hoc International Group on the Assessment of Disease Outcome in Scleroderma. J Rheumatol 1995;22:1277-80.

40. Charman CR, Venn AJ, Williams HC. Measurement of body surface area involvement in atopic eczema: an impossible task? Br J Dermatol 1999;140:109-11.

41. Finlay AY. Measurement of disease activity and outcome in atopic dermatitis. Br J Dermatol 1996;135:509-15.

42. Ramsay B, Lawrence CM. Measurement of involved surface area in patients with psoriasis. Br J Dermatol 1991;124:565-70.

43. Tiling-Grosse S, Rees J. Assessment of area of involvement in skin disease: a study using schematic figure outlines. Br J Dermatol 1993;128:69-74.

44. Ashcroft DM, Wan Po AL, Williams HC, Griffiths CE. Clinical measures of disease severity and outcome in psoriasis: a critical appraisal of their quality. Br J Dermatol 1999;141:185-91.

45. Charman CR, Venn AJ, Williams H. Measuring atopic eczema severity visually: which variables are most important to patients? Arch Dermatol 2005;141:1146-51.

46. Albrecht J, Taylor L, Berlin JA, Dulay S, Ang G, Fakharzadeh S, et al. The CLASI (Cutaneous Lupus Erythematosus Disease Area and Severity Index): an outcome instrument for cutaneous lupus erythematosus. J Invest Dermatol 2005;125:889-94.

47. Charman CR, Venn AJ, Williams HC. Reliability testing of the Six Area, Six Sign Atopic Dermatitis severity score. Br J Dermatol 2002;146:1057-60.

48. Klein RQ, Bangert CA, Costner M, Connolly MK, Tanikawa A, Okawa J, et al. Comparison of the reliability and validity of outcome instruments for cutaneous dermatomyositis. Br J Dermatol 2008;159:887-94. Epub 2008 Jul 4.

49. Lee CW, Kwon CW, Yoo DH, Kim SY. Linear scleroderma occurring in a patient with systemic lupus erythematosus — short report. J Korean Med Sci 1994;9:197-9.

50. Robertson LP, Davies MG, Hickling P. Generalized morphoea in a patient with Felty's syndrome. J Eur Acad Dermatol Venereol 2000;14:191-3.

51. Zannin ME, Martini G, Athreya BH, Russo R, Higgins GC, Vittadello F, et al. Ocular involvement in children with localized scleroderma: a multicenter study. Br J Ophthalmol 2007;91:1311-4. Epub 2007 May 2

52. Hongbo Y, Thomas CL, Harrison MA, Salek MS, Finlay AY. Translating the science of quality of life into practice: What do dermatology life quality index scores mean? J Invest Dermatol 2005;125:659-64.

53. Chuh AA, Chan HH. Effect on quality of life in patients with pityriasis rosea: is it associated with rash severity? Int J Dermatol 2005;44:372-7.

54. Hon KL, Kam WY, Lam MC, Leung TF, Ng PC. CDLQI, SCORAD and NESS: are they correlated? Qual Life Res 2006;15:1551-8.