The Localized Scleroderma Skin Severity Index and Physician Global Assessment of Disease Activity: A Work in Progress 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. RABINOIVCH, MARGALIT ROSENKRANZ, DANIEL A. KIETZ, PAUL ROSEN, STEVEN J. SPALDING, TERESA R. HENN ON, 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/validation 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 SKIN SCORES GLOBAL ASSESSMENT OUTCOME MEASURE QUALITY OF LIFE

From the Division of Rheumatology, Children’s Hospital of Pittsburgh and University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania; Hackensack University Medical Center, Hackensack, New Jersey; Oklahoma University Health Science Center, Oklahoma City, Oklahoma, USA; Hospital for Sick Children and University of Toronto, Toronto, Ontario, Canada; Nationwide Children’s Hospital and Ohio State University, Columbus, Ohio; Texas Scottish Rite Hospital for Children and University of Texas Southwestern Medical Center, Dallas, Texas; Duke University Medical Center, Durham, North Carolina; and Cleveland Clinic, Cleveland, Ohio, USA.

Supported by divisional funds.

T. Arkachaisri, MD, Assistant Professor (currently Associate Professor, Duke-NUS Graduate Medical School, Head and Senior Consultant, Rheumatology and Immunology Service, KK Women’s and Children’s Hospital, Singapore); S. Vilaiyuk, MD, Clinical Fellow; M. Rosenkranz, MD, Assistant Professor; D.A. Kietz, MD, Assistant Professor; P. Rosen, MD, MPH, Assistant Professor; T.R. Hennon, MD, Clinical Fellow; K.S. Torok, MD, Clinical Fellow; E. Cassidy, MD, Clinical Fellow, Children’s Hospital of Pittsburgh; T.A. Medsger Jr, MD, Professor, University of Pittsburgh; S. Li, MD, PhD, Assistant Professor, Joseph M. Sanzari Children’s Hospital and Hackensack University Medical Center; K.M. O’Neil, MD, Associate Professor, Oklahoma University Health Science Center; E. Pope, MD, MSc, Associate Professor, Hospital for Sick Children and University of Toronto; G.C. Higgins, MD, PhD, Associate Professor, Nationwide Children’s Hospital and Ohio State University; M. Punaro, MD, Professor, Texas Scottish Rite Hospital for Children and University of Texas Southwestern Medical Center; E.C. Rabinovich, MD, Associate Professor, Duke University Medical Center; S.J. Spalding, MD, Assistant Professor, Cleveland Clinic.

Address correspondence to Dr. T. Arkachaisri, Rheumatology and Immunology Service, KK Women’s and Children’s Hospital, 100 bukit Timah Road Singapore 229899.

E-mail: thaschawee.arkachaisri@kkh.com.sg

Accepted for publication July 10, 2009.
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. The incidence of LS is estimated to be 2.7 cases per 100,000 persons at risk/year. 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. 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). 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. 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). 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. 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 (PGA-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** involved 15 patients with LS and 2 examiners. Two 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 PhysGA-S and CDLQI were completed by patients. There were 25/27 patients with LS 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.

**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. 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 ± 11.0 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 rheumatologist/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 PhysGA-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.

**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) of the disease in 18 cutaneous anatomic sites (head, neck, chest, abdomen, upper back, lower back, right and left — arms, forearms, hands/fingers, buttocks/thighs, legs and feet). Each anatomic site was scored from 0 to 3, where 0 = normal or postinflammatory hyper/hypopigmentation, 1 =
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 diseases12–14. 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?”.

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

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

Arkachaisri, et al: Scleroderma outcome measures

3

Downloaded on June 8, 2022 from www.jrheum.org
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 0.79–0.99 (data not shown), as did the mLoSSI (ICC 0.78–0.99). Overall interrater reliability between clinical skin scores obtained from the 2 sessions was moderate for SA (rs = 0.51, p < 0.001), excellent for ER (rs = 0.89, p < 0.001), substantial for ST (rs = 0.71, p < 0.001), and excellent for LoSSI (rs = 0.81, p < 0.001) and mLoSSI (rs = 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), 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 (rs = 0.51, p < 0.001), excellent for ER (rs = 0.89, p < 0.001), substantial for ST (rs = 0.71, p < 0.001), and excellent for LoSSI (rs = 0.81, p < 0.001) and mLoSSI (rs = 0.77, p < 0.001).
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.
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 \( \kappa \)* 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 2. Interrater reliability* of Localized Scleroderma Skin Severity Index (LoSSI) and its domains.**

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

* 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.**

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

* 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.

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

* 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/mm3, 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 (κ* = modified kappa).

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

<table>
<thead>
<tr>
<th>Pairs of Measures (n = 27)</th>
<th>Spearman’s rho</th>
</tr>
</thead>
<tbody>
<tr>
<td>mLoSSI</td>
<td></td>
</tr>
<tr>
<td>Physician global assessment of disease activity</td>
<td>0.49*</td>
</tr>
<tr>
<td>Patient global assessment of disease severity</td>
<td>0.44**</td>
</tr>
<tr>
<td>CDLQI</td>
<td>0.25†</td>
</tr>
<tr>
<td>Physician global assessment of disease activity</td>
<td></td>
</tr>
<tr>
<td>Patient global assessment of disease severity</td>
<td>0.81*</td>
</tr>
<tr>
<td>CDLQI</td>
<td>0.15†</td>
</tr>
<tr>
<td>Patient global assessment of disease severity</td>
<td></td>
</tr>
<tr>
<td>CDLQI</td>
<td>0.21†</td>
</tr>
</tbody>
</table>

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.

DISCUSSION

To date, there is no simple, feasible, or reliable tool to assess LS skin changes. Thermography, ultrasound, and computerized image analysis have limited utility7-10,24-28. The applications of these instruments in assessing LS disease activity have been reviewed elsewhere11.

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-

Arkachaisri, et al: Scleroderma outcome measures

www.jrheum.orgDownloaded on June 8, 2022 from www.jrheum.org
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\textsuperscript{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)\textsuperscript{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)\textsuperscript{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)\textsuperscript{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\textsuperscript{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\textsuperscript{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\textsuperscript{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\textsuperscript{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\textsuperscript{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\textsuperscript{40-44}. Less experienced physicians had a higher interrater variability as compared to more experienced physicians\textsuperscript{42}. Poor reliability of a domain can interfere with

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

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

* E: en coup de sabre; M: morphea; SqM: subcutaneous morphea, GM: generalized morphea; L: linear scleroderma. 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 $= \text{mean observed change scores/standard deviation of baseline scores}$. SES $> 0.8$ considered large, $> 0.5$ moderate, and $> 0.2$ small. IQR: interquartile range.
overall reliability and validity of the outcome measure, as it
imparts accuracy and consistency. As suggested by Finlay,
recording of area involvement should be based on an assess-
ment of site involvement rather than the virtually impossible
task of determining an accurate total percentage involve-
ment. The location of skin affected may be more important
than the percentage of involvement. For example, a small
linear lesion crossing the elbow causing elbow joint con-
tacture 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 physi-
cal disability, i.e., joint contractures, localized growth fail-
ure (66% limbs, 23% trunk, and 11% face/scalp/neck), simi-
lar to that used in psoriasis.

We propose to omit SA from LoSSI, which does not sig-
ificantly reduce the overall inter- and intrarater reliability
[LoSSI with SA vs LoSSI without SA (mLoSSI): ICC 0.80
vs 0.70 and r = 0.81 vs r = 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 com-
ponent), 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 con-
sensus agreement but had excellent content validity index.
From group discussion, both acute-phase reactants are non-
specific 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 dis-
eases overlapping with LS. Therefore, arthritis may not truly
reflect LS disease activity and is not included in
PhysGA-A assessment. Uveitis is the only noncutaneous
clinical variable included as one of the PhysGA-A as it rep-
resents 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. Once detect-
ed, 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 evi-
dence 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. Interestingly, we found
excellent correlation between PhysGA-A and PtGA-S (r =
0.81, p < 0.001), higher than that we previously reported.
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, 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 predict-
ed, SA showed poor sensitivity to change (SRM = 0.3), con-
firming 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 der-
matitis [7.7 (SD 5.6)], psoriasis [5.4 (SD 5.0)], and acne [5.7
(4.4)]14. These preliminary CDLQI results suggest that
LS has a relatively small effect on a patient’s life quality or
that a LS-specific QOL tool may be needed. CDLQI corre-
lated poorly with the mLoSSI, which measures different
aspects of disease.

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 cuta-
aneous 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 dermato-
logic outcome measures proposed by Finlay. QOL did not
correlate well with clinical skin scores or GA, but all are
important and must be individually evaluated. The integra-
tive assessment of these 3 areas represents a complemen-
tary 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 under-
standing 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 cuta-
aneous assessment tool should be to determine the mLoSSI

Arkachaisri, et al: Scleroderma outcome measures
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
We extend thanks to patients and their families who kindly gave a day of their time to participate in this project; Prof. Andrew Y. Finlay and Dr. M.S. Lewis-Jones from the Department of Dermatology, Wales College of Medicine, Cardiff University, Wales, UK, for their permission to use CDLQI in our study; Dr. Ronald Laxer for helping with the survey; Yan Lin, PhD, Institute for Clinical Research Education, University of Pittsburgh, Pittsburgh, Pennsylvania, USA, for assistance with statistical analysis; and Carolyn Confer, Marsha and Chase Ebaugh, Barbara Persichetti, and Marilyn R. Cieri for their assistance with the project.

REFERENCES
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.