

## Original article

# Development and initial validation of the Localized Scleroderma Skin Damage Index and Physician Global Assessment of disease Damage: a proof-of-concept study

Thaschawee Arkachaisri<sup>1,2</sup>, Soamarat Vilaiyuk<sup>1</sup>, Kathryn S. Torok<sup>1</sup> and Thomas A. Medsger Jr<sup>3</sup>

## Abstract

**Objective.** To develop and assess the psychometric properties of the Localized Scleroderma (LS) Skin Damage Index (LoSDI) and Physician Global Assessment of disease Damage (PGA-D).

**Methods.** Damage was defined as irreversible/persistent changes (>6 months) due to previous active disease/complications of therapy. Eight rheumatologists assessed the importance of 17 variables in formulating the PGA-D/LoSDI. LS patients were evaluated by two rheumatologists using both tools to assess their psychometric properties. LoSDI was calculated by summing three scores for cutaneous features of damage [dermal atrophy (DAT), subcutaneous atrophy (SAT) and dyspigmentation (DP)] measured at 18 anatomic sites. Patient GA of disease severity (PtGA-S), Children's Dermatology Life Quality Index (CDLQI) and PGA-D were recorded at the time of each examination.

**Results.** Thirty LS patients (112 lesions) and nine patient-visit pairs (18 lesions) were included for inter- and intra-rater reliability study. LoSDI and its domains DAT, SAT, DP and PGA-D demonstrated excellent inter- and intra-rater reliability (reliability coefficients 0.86–0.99 and 0.74–0.96, respectively). LoSDI correlated moderately with PGA-D and poorly with PtGA-S and CDLQI. PGA-D correlated moderately with PtGA-S, but poorly with CDLQI.

**Conclusions.** To complete the LS Cutaneous Assessment Tool (LoSCAT), we developed and evaluated the psychometric properties of the LoSDI and PGA-D in addition to the LS Skin Severity Index (LoSSI). These instruments will facilitate evaluation of LS patients for individual patient management and clinical trials. LoSDI and PGA-D demonstrated excellent reliability and high validity. LoSCAT provides an improved understanding of LS natural history. Further study in a larger group of patients is needed to confirm these preliminary findings.

**Key words:** Localized scleroderma, Outcome measures, Damage index, Global assessment.

<sup>1</sup>Department of Pediatrics, Division of Rheumatology, Children's Hospital of Pittsburgh, University of Pittsburgh School of Medicine, Pittsburgh, PA, <sup>2</sup>Rheumatology and Immunology Service, Department of Paediatric Subspecialties, KK Women's and Children's Hospital, Singapore and <sup>3</sup>Department of Internal Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA.

Submitted 2 September 2009; revised version accepted 2 October 2009.

Correspondence to: Thaschawee Arkachaisri, Rheumatology and Immunology Service, KK Women's and Children's Hospital, 100 Bukit Timah Road, Children's Tower, L5, Singapore 229899.  
E-mail: thaschawee.arkachaisri@kkh.com.sg

## Introduction

Localized scleroderma (LS) is a chronic autoimmune disease characterized by fibrotic and atrophic changes after an initial phase of inflammation affecting primarily skin and underlying structures [1]. The incidence of LS is estimated to be 2.7 cases per 100 000 persons at risk per year [2]. Until recently, there have been no clinical outcome measures developed and validated for LS. Without such instruments, assessing clinical changes in routine LS patient care and the development and conduct of clinical trials are impeded or impossible.

We developed and initially validated the first scoring method for measuring LS skin activity during the early inflammatory phase—the LS Skin Severity Index (LoSSI) [3, 4]. The psychometric properties of LoSSI were recently evaluated by the Localized Scleroderma Clinical and Ultrasound Study Group (LOCUS); a modified index (mLoSSI) was recommended [4].

Although not a life-threatening disease, LS, regardless of subtype, almost always results in chronic and/or irreversible changes in the skin or underlying tissue. Extracutaneous complications are uncommon. They include orthopaedic (joint contracture and limb length discrepancy), neurological (seizure, abnormal MRI and headache) and ocular (glaucoma) involvement [5, 6].

In order to accurately appraise the efficacy of therapeutic interventions, it is necessary to have objective measures of both disease activity and disease damage. In this study, we proposed, developed and assessed the psychometric properties (reliability and validity) of the LS Skin Damage Index (LoSDI) and Physician Global Assessment of LS disease Damage (PGA-D).

## Patients and methods

### Patients

LS patients were recruited from the Scleroderma Clinic at Children's Hospital of Pittsburgh of the University of Pittsburgh Medical Center. Diagnosis and classification of LS were made according to the recommendations of Peterson *et al.* [7]. The University of Pittsburgh Institutional Review Board approved the study, and patient informed consent was obtained before entering the study.

### Study design

There were two phases of the study.

- (i) The first phase was to obtain consensus agreement on the cutaneous and extracutaneous variables which should be considered as damage changes of LS and to assess the content validity of each of these items in formulating the PGA-D (both cutaneous and extracutaneous variables) and the LoSDI (cutaneous manifestations only).
- (ii) The second phase involved evaluating the psychometric properties of the LoSDI in our pilot cohort LS patients by two rheumatologists.

### Development of LoSDI and PGA-D

The LOCUS included seven paediatric rheumatologists and one paediatric dermatologist experienced in the care of juvenile LS patients who volunteered their time and resources to a multicentre study devoted to the development and validation of LS outcome measures. A two-part survey was used, which is the first to suggest the variables and the second to rank these variables pertaining to both cutaneous and extracutaneous disease activity (not included in this report) and damage. The survey was sent to the LOCUS members and one other author (T.A.M.). Disease damage was defined as

irreversible changes or persistent changes of the lesion (>6 months) due to previous active disease or complications of therapy. Eight physicians completed the survey by suggesting and then ranking the importance of 17 clinical signs and symptoms and 10 laboratory tests on a scale of 0–4 (0: unimportant; 1: mildly important; 2: moderately important; 3: very important; and 4: extremely important). The second (follow-up) survey was to assess the LOCUS group perception of the cutaneous changes of LS lesions that they would consider as persistent [skin thickness, dermal atrophy (DAT), subcutaneous atrophy (SAT) and dyspigmentation (DP)]. Cutaneous changes were considered to be completely reversible, partially reversible (activity > damage or damage > activity) or completely irreversible. Mean ranks, per cent agreement and item- and scale-level content validity indices (I-CVI and S-CVI) were used to determine the content validity of LoSDI and PGA-D. The physicians participating in these surveys had, on average, 19.9 (range 8.0–40.0, median 18) years in practice and reported that they had an average of 85 LS patient visits per year (range 24.0–144.0, median 108). As a result of both surveys, three skin changes formed the LoSDI (see below).

### Description of the damage index

LoSDI was developed to describe the presence and extent of LS skin damage changes in 18 cutaneous anatomic sites [4]. Three domains were included and graded from 0 to 3 and then summed to obtain the LoSDI, as follows.

- (i) DAT: 0: normal skin; 1: mild skin atrophy, i.e. shiny skin; 2: moderately atrophy, i.e. visible blood vessels or mild 'cliff-drop' sign; and 3: severe skin atrophy, i.e. obvious 'cliff-drop' sign.
- (ii) SAT: 0: normal subcutaneous thickness; 1: flattening or  $\leq 1/3$  fat loss; 2: obvious concave surface or  $1/3$ – $2/3$  fat loss; and 3: severe subcutaneous fat loss (> $2/3$  loss).
- (iii) DP: assessing both hyper- or hypopigmentation, whichever is most prominent: 0: normal skin pigmentation; 1: mild; 2: moderate; and 3: severe DP.

All skin damage changes were compared with the contralateral skin area or ipsilateral skin area if bilateral involvement was present in order to minimize inter-subject variability. If a given anatomic site had more than one lesion, the most severe score obtained from each domain would be used to calculate the LoSDI (range 0–162). For example, a patient had two LS lesions on the abdomen (one anatomic site). Lesion A had DAT1, SAT2 and DP3 (ivory discolouration), and lesion B had DAT3, SAT0 and DP2 (moderately hyperpigmentation). LoSDI for this patient at that visit would be 3 (DAT) + 2 (SAT) + 3 (DP) or 8. This approach eliminates sampling confusion if a given anatomic site contains more than one lesion. The method views multiple small lesions as if there is one large lesion in a given anatomic site.

### Global assessment

PGA-D and patient global assessment of disease severity (PtGA-S) were obtained at each visit. Patients were instructed to consider the past 1 month for their answers. For PGA-D, raters assessed global disease damage (using clinical variables with high content validity) on a 100-mm visual analogue scale with the anchors being 0 (no damage) and 100 (severe damage). For PtGA-S, anchors were 0 (not severe) and 100 (very severe). PtGA-S was completed by patients aged  $\geq 8$  years, otherwise by the accompanying parent [3, 4] blinded to the physician ratings.

### Quality of life

Children's Dermatology Life Quality Index (CDLQI) was completed by patients aged  $\geq 8$  years, otherwise by the accompanying parent [8].

### Psychometric study

Two rheumatologists (T.A. and S.V.) independently assessed the LoSDI and PGA-D for inter-rater reliability evaluation. Records were blinded between patient visits. Intra-rater reliability assessment was obtained by the same examiners re-evaluating the same patients 8–12 weeks after the initial examination. All patients included in this portion of the study had stable disease for at least 3 months. At each visit, patients completed the PtGA-S.

### Statistical analysis

All analyses were performed using statistical programs SPSS v.16 (SPSS, Chicago, IL, USA) and STATA v.8 (Stata, College Station, TX, USA). Mean (s.d.) or median [interquartile range (IQR)] was used to describe data where appropriate. Weighted kappa ( $\kappa_w$ ) and per cent raw agreement or intra-class correlation coefficient (ICC) were used to assess inter-rater reliability. Spearman's correlation coefficient assessed the intra-rater reliability of both raters. Interpretation of agreement followed the recommendation of Landis and Koch [9]: 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 [9].

Content validity concerns the extent to which a specific set of items reflects a content domain [10]. Content validity of LoSDI and PGA-D were evaluated by using mean rank, per cent consensus agreement and I-CVI and S-CVI [11]. I-CVI was calculated by dividing the number of experts ranking each variable being moderately to extremely important by the total number of experts, and S-CVI was obtained by summing I-CVI and dividing by the number of variables [11]. An I-CVI  $\geq 0.78$  (for item level) and S-CVI  $\geq 0.9$  (for scale level) were considered as evidence of excellent content validity [11]. Correlations between outcome measures as calculated by Spearman's correlation were taken as evidence of construct validity, which is concerned with the extent to which a particular measure (in this case, the LoSDI) relates to other measures consistent with theoretically derived hypotheses concerning the concepts (or construct) that are being measured [12]. Spearman's correlation

coefficient  $\geq 0.7$  was defined as strong,  $\geq 0.4$  and  $<0.7$  as moderate and  $<0.4$  as poor. All statistical tests were two sided, and a  $P$ -value  $<0.05$  was considered to be statistically significant.

## Results

### Patients

Thirty patients with average age at LS onset of 8.2 [median (IQR) 7.5 (5–11.3)] years were included in this pilot study. There were 22 females and 8 males with a mean age of 11.8 [median (IQR) 12.5 (8.0–14.3)] years. Five patients had plaque morphea, eight had linear scleroderma, six had mixed morphea/linear scleroderma, four had generalized morphea, four had subcutaneous morphea and three had *en coup de sabre*. Each patient had on average 3.5 lesions [median (IQR) 3.0 (1.8–6.0)]. Median (IQR) of duration of disease was 29.0 (20.5–53.5) months, and that of treatment duration was 16.0 (5.8–29.0) months. Twenty-seven patients were on different treatment regimens (11, MTX; 12, MTX and prednisone; 2, prednisone only; and 2, topical therapies).

### Damage domain frequencies

One hundred and twelve anatomic sites were assessed during 39 patient-visits by two examiners independently. Figure 1 shows the distribution of DAT, SAT and DP. One-fifth of the anatomic sites had no DAT, but 45, 44 and 2% had 1+, 2+ and 3+ DAT, respectively. For SAT, one-third had no SAT and 46% had mild, 14% had moderate and 9% had severe atrophy. Ninety-six per cent of our sample lesion sites had various degree of DP: 49%, 1+; 40%, 2+; and 7%, 3+ DP. As shown in Fig. 2, the mean (s.d.) LoSDI was 11.7 (10.4) [median (IQR) 8.0 (3.0–16.0)].

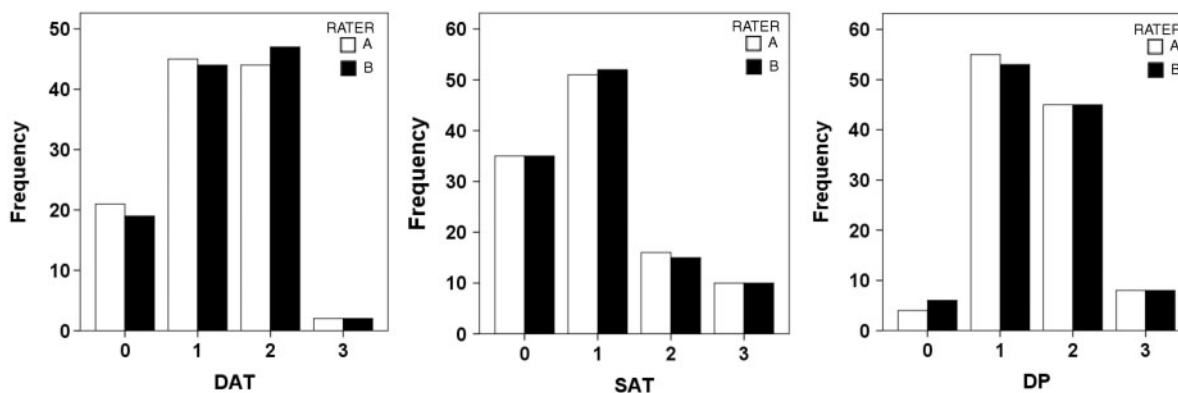
### Reliability

Analyses of both inter- and intra-rater reliability of damage domains and LoSDI are summarized in Table 1.

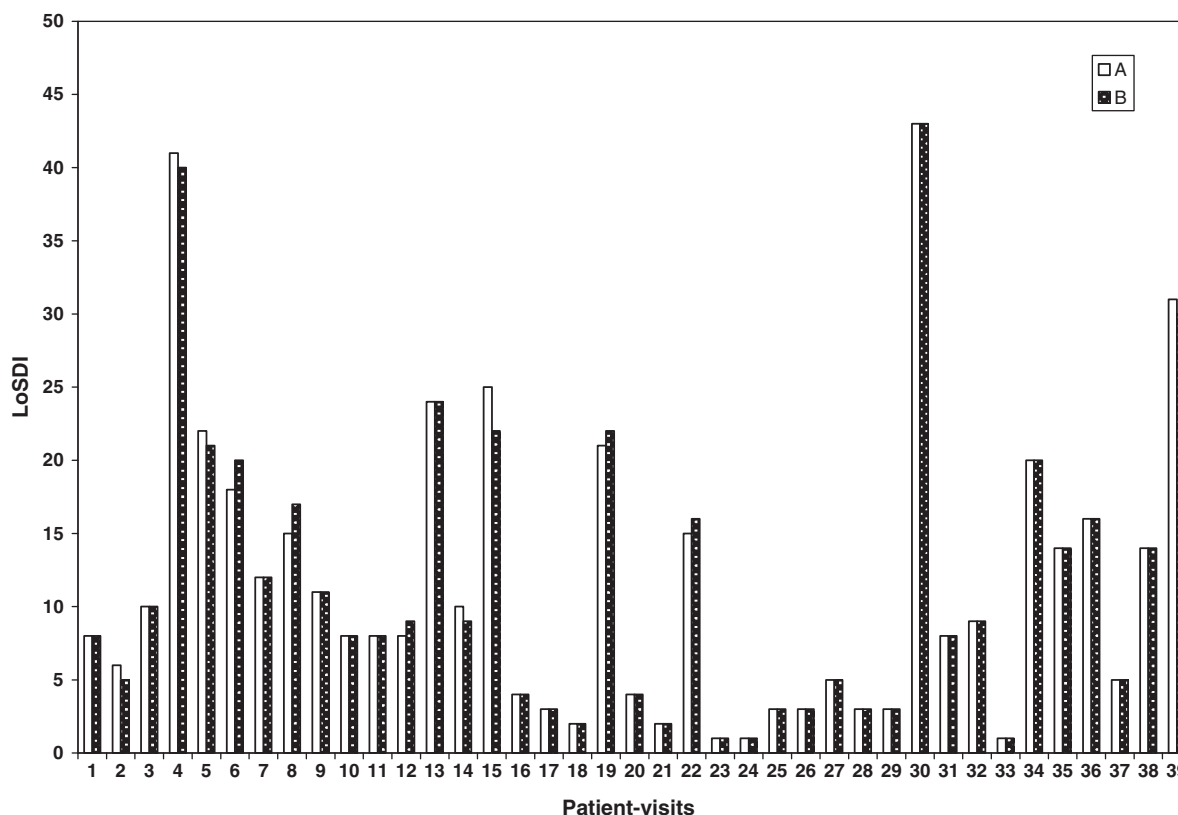
*Inter-rater reliability.* As assessed by  $\kappa_w$  statistic and per cent raw agreement, all domains—DAT, SAT and DP—had almost perfect inter-rater reliability ( $\kappa_w > 0.80$ ;  $>95\%$  agreement). The total damage score (LoSDI) for each patient showed excellent inter-rater reliability by ICC (0.99; 95% CI 0.99, 1.00).

*Intra-rater reliability.* Nine LS patients (six females and three males) with stable disease and complete follow-up skin damage scores (median 13 weeks apart) were included in the intra-rater reliability study. There were nine pair-visits and 35 cutaneous anatomic sites were evaluated and compared. These patients had a median (IQR) disease duration of 30 (23–64) months. Wilcoxon signed rank tests comparing damage scores and LoSDI at two time points did not show statistically significant differences—DAT ( $z = -0.58$ ;  $P = 0.56$ ), SAT ( $z = -0.82$ ;  $P = 0.41$ ), DP ( $z = -1.89$ ;  $P = 0.06$ ) and LoSDI ( $z = -0.41$ ;  $P = 0.69$ ). These results suggest that the damage domains chosen do not change over a 3-month time period.

**Fig. 1** Localized Scleroderma Skin Damage domains distribution. Bars represent the frequencies of each domain score. Distribution recorded by two raters (A and B) from 112 cutaneous surface anatomic sites examined.



**Fig. 2** LoSDI distribution. Bars represent LoSDI scores recorded by two raters (A and B) from 39 patient-visits.



Spearman’s correlation analysis of intra-rater reliability for each domain showed moderate reliability for SAT ( $r_s=0.75$ ) and DP ( $r_s=0.74$ ) domains and excellent reliability for DAT domains ( $r_s=0.91$ ) and LoSDI ( $r_s=0.96$ ) (Table 1).

*PGA-D.* As shown in Table 1, PGA-D had excellent inter- and intra-rater reliability as assessed by both ICC and Spearman’s correlation coefficient. The Wilcoxon signed rank test did not show a statistically significant difference in PGA-D between two time points ( $z=-1.92$ ;  $P=0.06$ ).

**Validity**

*Content validity.* Mean rank, per cent consensus agreement and CVI were used to analyse content validity of the LoSDI and PGA-D. Seventeen clinical variables related to LS disease damage were assessed by LOCUS to formulate the PGA-D clinical determinants. Only DAT, SAT and DP were considered to be moderate to extremely important variables (mean rank 2.5–3.0 and 86–88% consensus agreement) for LS cutaneous damage by this expert panel (Table 2). For LoSDI, all three domains also

had high I-CVI (>0.78) and S-CVI (0.92). From our second survey assessing the perception of how reversible these LS skin changes are, all experts considered skin thickness as either completely reversible (43%) or reflecting activity more than damage. DP were considered to reflect damage > activity (100%) and DAT was viewed as completely irreversible (29%) or as damage > activity (71%). SAT was considered to be completely irreversible (57%) or as damage > activity (43%).

For PGA-D clinical determinants, 13 variables were considered to be moderate to extremely important variables when one is assessing LS PGA-D. The S-CVI for PGA-D was 0.98, suggesting that at both the item and the scale level, using these 13 variables yielded high content validity measurements. As shown in Table 2, there was no consensus agreement on adding the skin thickness both at the centre and at the border of individual lesions, uveitis

or arthritis as part of the PGA-D assessment (mean rank 1.3, <75% agreement within 1 rank).

*Convergent construct validity.* Spearman's correlation coefficient ( $r$ ) assessed the degree of correlation between pairs of outcome measures to demonstrate evidence of convergent construct validity. As shown in Figure 3, LoSDI correlated moderately with PGA-D ( $r=0.58$ ;  $P<0.001$ ). Moderate correlation was also found between both raters' PGA-D and PtGA-S ( $r=0.42$  and  $0.46$ ,  $P=0.02$  and  $0.03$ , respectively). However, poor correlations ( $r<0.4$ ) were found between LoSDI vs ptGA-S, LoSDI vs CDLQI and PGA-D vs CDLQI. The CDLQI correlated moderately with ptGA-S ( $r=0.41$ ;  $P=0.03$ ).

## Discussion

LS is an uncommon and non-fatal autoimmune disease. Its consequences, including the results of previous active inflammation or adverse reactions from therapy affecting cutaneous and underlying tissues, have not been well characterized. Cutaneous and subcutaneous damage can cause both physical and psychological morbidity. Without a reliable instrument to quantify the accumulated morbidity due to LS, the natural history of the disease cannot be accurately described. We have developed and reported the initial pilot psychometric properties of the LoSDI and PGA-D.

The LoSDI is an instrument using simple clinical examination of the skin and subcutaneous tissues to quantify disease damage. It is simple and non-time consuming and thus is feasible for use in both routine clinical care and clinical trials. In order to develop the LoSDI, skin changes

**TABLE 1** The LoSDI and its domain reliability

Domains	Inter-rater reliability ( $\kappa_w^a$ )	Intra-rater reliability ( $r_s^{b,\dagger}$ )
DAT	0.92 (97.92)	0.91
SAT	0.86 (95.54)	0.75
DP	0.90 (97.62)	0.74
LoSDI	0.99 (0.99, 1.00) <sup>c</sup>	0.96
PGA-D	0.90 (0.78, 0.95) <sup>c</sup>	0.89

<sup>a</sup>Weighted  $\kappa$ -statistic (per cent raw agreement). <sup>b</sup>Spearman's correlation coefficient. <sup>c</sup>ICC (95% CI). <sup>†</sup> $P<0.001$  unless specified.

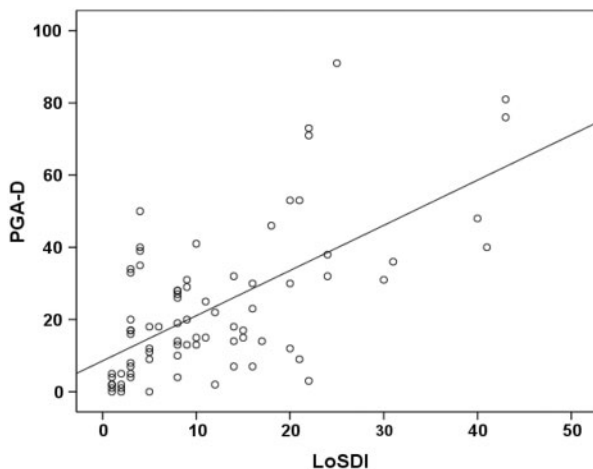
**TABLE 2** Clinical variables used in the physicians' determinations of global assessment of disease damage in localized scleroderma<sup>a</sup>

Clinical variables with consensus ( $\geq 75\%$ agreement) in assessing disease damage	Mean rank (0–4)	Per cent consensus agreement	I-CVI <sup>b</sup>
Very/extremely important			
Facial atrophy	3.9	100	1.00
Skeletal muscle atrophy	3.9	100	1.00
Physical disability	3.9	100	1.00
Joint contracture	3.8	100	1.00
Bone atrophy	3.8	100	1.00
Eye (cataract/glaucoma)	3.6	100	1.00
Limb length discrepancy	3.6	100	1.00
Central nervous system symptoms i.e. seizure	3.6	75	1.00
Abnormal brain MRI	3.1	75	1.00
SAT	3.0	88	1.00
Psychosocial/quality of life impairment	2.9	88	1.00
DP (hyper/hypopigmentation, ivory)	2.8	86	0.88
Moderately important			
DAT	2.5	88	0.88

<sup>a</sup>Clinical variables that were ranked, but with no consensus (<75% consensus agreement and CVI < 0.78) achieved on global assessment of localized scleroderma disease activity, are listed in rank order as follows: skin thickening at the border of a lesion, skin thickening at the centre of a lesion, uveitis and arthritis ( $n=8$ ). <sup>b</sup>I-CVI Item-level Content Validity Index.



**Fig. 3** Scatter plot showing the correlation between the LoSDI and PGA-D. Spearman's correlation coefficient = 0.58 ( $P < 0.001$ ).



thought to represent damage rather than activity were first proposed by two of the authors (T.A. and T.A.M.). These cutaneous damage changes included DAT, SAT and DP (hypopigmentation/ivory discolouration or hyperpigmentation). These physical examination findings were congruent with the recommendations of members of the LOCUS group as very or extremely important in assessing LS cutaneous disease damage. Furthermore, all three of these skin changes were considered as either completely irreversible (DP) or partially reversible, which represented more damage than activity (DAT and SAT). Interestingly, the experts had different opinion regarding the reversible possibility of the latter two findings—the DAT and SAT; hence this semiquantitative assessment instrument will add more understanding to the LS natural history.

Forty per cent of the experts considered skin thickness to be completely reversible and 60% regarded this finding more a reflection of activity than damage. Skin thickness at the margin of a lesion was thought to be secondary to an active inflammatory process—dermal oedema and inflammatory cell infiltration. In contrast, skin thickness at the centre of a lesion was considered to be a product of fibrosis with excessive collagen deposition. In clinical practice, it may not be expected for inexperienced physicians to accurately palpate different parts (margin or centre) of an LS lesion. Even experienced physicians have problems assessing small LS lesions. Despite these limitations, skin thickness was included in the LoSSI [3, 4].

The method used to assess content validity of our instruments (LoSDI and PGA-D) was consensus agreement and CVI at both item and scale level. This quantitative validity evaluation is intended to ascertain that a given scale contains all important variables/domains representing the construct of interest. The former was used by the juvenile dermatomyositis collaborative study group for the development of its physician global assessment [13], and the latter has been used widely in a number of

scale development investigations [11]. Both methods started with a survey of experts, who rated each variable according to its degree of importance, in this case reflecting LS disease damage. Mean rank and per cent consensus agreement were then calculated. The variables that achieved at least moderately important ( $\geq 75\%$ ) agreement were chosen. The CVI adopted the same approach but used the proportion of panelists for calculation. The interpretation of CVI, both at item or scale levels, was based on the number of panelists, as suggested in detail by Polit *et al.* [11, 14]. We used quantitative content validity assessment to demonstrate the content validity of both LoSDI and PGA-D. Both instruments, LoSDI—three skin signs and PGA-D—13 clinical variables, demonstrated excellent content validity. Without assigning which clinical variables to include in the global assessment of LS disease damage, reliability coefficients were shown to be very low (data not shown).

By assessing the degree of DAT, we graded epidermal atrophy (as manifested by the shiny appearance of the lesion) as mild DAT or DAT 1 to make the assessment simple [15]. The DAT2 and -3 scores depend on the degree of dermal depression. A good example of severe DAT is so called 'cliff-drop' appearance, which is a sharply demarcated depressed area of the skin as seen commonly in the atrophic variant of LS (atrophoderma of Pasini and Pierini) [16]. The DAT domain showed excellent inter- and intra-rater reliability ( $\kappa_w$  0.92 and  $r_s$  0.91, respectively). More accurate measurement of DAT could be done only by very high frequency ultrasound (20–50 MHz), which is not readily available in the USA [17]. More studies are needed to standardize the ultrasound technique and to validate and assess its reliability.

SAT, unlike DAT, can be measured by 10–15 MHz ultrasound, as it has a better penetration as compared with the 20–50 MHz ultrasound machine [17, 18]. However, some technical aspects remain problematic, the artefact causes by the ultrasound probe pressuring on the skin and underlying subcutaneous tissue during the scan and the scanning axis and locations of the lesion, for which one has to be certain that the scans obtained from each follow-up study are reliably and accurately comparable. This technical standardization is underway by the LOCUS group. At this time, clinical examination—assessing the degree of SAT by visual examination and palpation—remains the most reliable, feasible and cost-effective method available. Our pilot study demonstrated good to excellent inter- ( $\kappa_w$  0.86) and intra-rater ( $r_s$  0.75) reproducibility.

Unlike DAT and SAT, degree of DP can only be assessed by clinical visual examination. Both post-inflammatory pigmentary changes can occur in the same lesion. In order to avoid redundancy and confusion, we proposed to use the most representative/severe DP of each lesion for DP grading. In this preliminary study, DP yielded good intra-rater ( $r_s$  0.74) and excellent inter-rater reliability ( $\kappa_w$  0.90), comparable with erythema grading in LoSSI [4].



In order to assess the 'extent' of LS skin damage, the same concept used for LoSSI was applied to LoSDI. We proposed to assess the three domains (DAT, SAT and DP) in 18 anatomic sites rather than attempting to estimate the affected surface area. The latter has been difficult to reproduce both within and across physicians, hence making it inaccurate and unreliable [19–22]. Furthermore, some lesions, especially in linear scleroderma, tend to coalesce. Thus, counting the number of the lesions to estimate extent is unsuitable. In a given anatomic site, if more than one lesion is involved, we propose to use the most representative/most severe degree of each domain for damage calculation purposes. This approach will allow us to consider many small lesions in one anatomic site to be a single large lesion, thus minimizing sampling confusion and being more clinically relevant. After all, the most severe atrophic or pigmentation change of lesions causes physical or mental disability.

The members of the LOCUS group agreed on a common set of 13 clinical variables for global assessment of LS disease damage. These variables also demonstrated high CVI at the item level. Formulating these 13 clinical variables as PGA-D assessment yielded high scale level CVI (S-CVI = 0.98). This global assessment was comprehensive, including not only included cutaneous but also extracutaneous manifestations and the psychosocial impact [6]. The PGA-D showed excellent reliability within and across physicians ( $r_s$  0.89 and  $\kappa_w$  0.90, respectively).

We considered the moderate correlations between LoSDI and PGA-D and between PGA-D and PtGA-S as evidence of convergent construct validity. The degree of congruence was comparable with that between LoSSI and physician global assessment of disease activity [4]. The discrepancy between the clinical examination score or LoSDI and global assessment was clearly due to the extracutaneous coverage of PGA-D. As expected, a lower correlation was found between PtGA-S and PGA-D ( $r=0.4$ ) than between PtGA-S and PGA-A ( $r=0.8$ ). This could be, in part, because patients regard disease severity as associated more with activity than with damage [4]. However, consistent with our previous report on the correlation of CDLQI with PGA-A or LoSSI ( $r=0.15$ ,  $0.25$ , respectively), both PGA-D and LoSDI showed poor correlations with CDLQI ( $r=0.20$ ,  $0.18$ , respectively), suggesting that LS may need a disease-specific quality of life instrument [4].

It is our intention to separate the activity/severity (LoSSI) from the damage (LoSDI) index. First, because LoSSI reflects more disease activity, it should guide treatment decisions. LoSDI reflects chronic, more irreversible tissue damage as the result of the disease and/or therapy. Secondly, as pointed out by Albrecht *et al.* [23], during the disease course, LS activity decreases and damage becomes more apparent. Thus, a score containing both activity and damage domains remains stable, whereas the clinical stage may shift completely. LS, unlike atopic dermatitis or psoriasis, often has a smaller cutaneous surface area affected, therefore the range of scores obtained from both LoSSI and LoSDI will rarely be in the

higher ranges. To this end, we propose the LS Cutaneous Assessment Tool (LoSCAT) data sheet collection form (Fig. 4), which contains both LoSSI and LoSDI scored separately.

One limitation of our study, like other pilot reports, is the small number of patients and examiners in a single centre. However, this proof-of-concept report demonstrated that LoSDI is a validated, reliable semiquantitative tool for measuring LS cutaneous disease damage over time. Both LoSDI and PGA-D not only measure the outcomes of therapy, but also provide a better understanding regarding the natural history of the various forms of LS. Larger multicentre study involving more examiners is needed to validate and confirm our initial findings.

#### Rheumatology key messages

- The LoSDI and PGA-D are reliable tools for measuring LS damage changes.
- Both LoSDI and PGA-D are feasible to be used in routine clinical setting and trials.
- LoSCAT provides an improved understanding of LS natural history.

## Acknowledgements

We thank members of the Localized Scleroderma Clinical and Ultrasound Study Group (LOCUS): Suzanne Li, MD, PhD; Marilyn Punaro, MD; Gloria Higgins, MD, PhD; Eglia Rabinovich, MD; Kathleen M. O'Neil, MD; Elena Pope, MD; Ronald Laxer, MD; and Thomas J.A. Lehman, MD, for helping with the survey; 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; and Yan Lin from the Institute for Clinical Research Education, University of Pittsburgh, for assistance with statistical analysis.

*Funding:* Division of Rheumatology, Children's Hospital of Pittsburgh fund.

*Disclosure statement:* The authors have declared no conflicts of interest.

## 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 Arkachaisri T, Pino S. Localized scleroderma severity index and global assessments: a pilot study of outcome instruments. *J Rheumatol* 2008;35:650–7.
- 4 Arkachaisri T, Vilaiyuk S, Li SC *et al.* The Localized Scleroderma Skin Severity Index (LoSSI) and Physician Global Assessment of Disease Activity: a work-in-progress toward the development of localized



- scleroderma outcome measures. *J Rheumatol* 2009; (in press).
- 5 Marzano AV, Menni S, Parodi A *et al*. Localized scleroderma in adults and children. Clinical and laboratory investigations on 239 cases. *Eur J Dermatol* 2003;13: 171–6.
  - 6 Zulian F, Vallongo C, Woo P *et al*. Localized scleroderma in childhood is not just a skin disease. *Arthritis Rheum* 2005;52:2873–81.
  - 7 Peterson LS, Nelson AM, Su WP. Classification of morphea (localized scleroderma). *Mayo Clin Proc* 1995;70: 1068–76.
  - 8 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.
  - 9 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33: 159–74.
  - 10 DeVellis RF. Scale development: theory and applications. In: Bickman L, Rog DJ, eds. Thousand Oaks: Sage Publication, Inc., 2003.
  - 11 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.
  - 12 Carmines EG, Zeller RA. Reliability and validity assessment. In: Lewis-Beck MS, ed. Thousand Oaks: Sage Publications, Inc., 1979.
  - 13 Rider LG, Feldman BM, Perez MD *et al*. Development of validated disease activity and damage indices for the juvenile idiopathic inflammatory myopathies: I. Physician, parent, and patient global assessments. Juvenile Dermatomyositis Disease Activity Collaborative Study Group. *Arthritis Rheum* 1997;40:1976–83.
  - 14 Polit DF, Beck CT. The content validity index: are you sure you know what's being reported? Critique and recommendations. *Res Nurs Health* 2006;29:489–97.
  - 15 Trozak DJ, Tennenhouse DJ, Russell JJ. *Dermatology skills for primary care: an illustrated guide*. Totowa, NJ: Humana Press, 2006.
  - 16 Buechner SA, Ruffi T. Atrophoderma of Pasini and Pierini. Clinical and histopathologic findings and antibodies to *Borrelia burgdorferi* in thirty-four patients. *J Am Acad Dermatol* 1994;30:441–6.
  - 17 Bendeck SE, Jacobe HT. Ultrasound as an outcome measure to assess disease activity in disorders of skin thickening: an example of the use of radiologic techniques to assess skin disease. *Dermatol Ther* 2007; 20:86–92.
  - 18 Li SC, Liebling MS, Haines KA. Ultrasonography is a sensitive tool for monitoring localized scleroderma. *Rheumatology* 2007;46:1316–9.
  - 19 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.
  - 20 Finlay AY. Measurement of disease activity and outcome in atopic dermatitis. *Br J Dermatol* 1996;135:509–15.
  - 21 Gaines E, Werth VP. Development of outcome measures for autoimmune dermatoses. *Arch Dermatol Res* 2008; 300:3–9.
  - 22 Ramsay B, Lawrence CM. Measurement of involved surface area in patients with psoriasis. *Br J Dermatol* 1991;124:565–70.
  - 23 Albrecht J, Taylor L, Berlin JA *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.