

The CLASI (Cutaneous Lupus Erythematosus Disease Area and Severity Index): An Outcome Instrument for Cutaneous Lupus Erythematosus

Joerg Albrecht,* Lynne Taylor,† Jesse A. Berlin,† Samuel Dulay,* Gina Ang,§ Steven Fakharzadeh,*¶ Jonathan Kantor,* Ellen Kim,* Giuseppe Militello,* Karen McGinnis,* Stephen Richardson,* James Treat,* Carmela Vittorio,* Abby Van Voorhees,* and Victoria P. Werth*¶

*Department of Dermatology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA; †Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, Pennsylvania, USA; §Department of Dermatology, Jefferson Medical College, Philadelphia, Pennsylvania, USA;

¶VA Hospital Philadelphia, Philadelphia, Pennsylvania, USA

We developed and validated a measurement instrument (CLASI—Cutaneous Lupus Erythematosus Disease Area and Severity Index) for lupus erythematosus that could be used in clinical trials. The instrument has separate scores for damage and activity. A group of seven American Dermato-Rheumatologists and the “American College of Rheumatology Response Criteria Committee on SLE (systemic lupus erythematosus)” assessed content validity. After a preliminary session, we conducted standardized interviews with the raters and made slight changes to the instrument. The final instrument was evaluated by five dermatologists and six residents who scored nine patients to estimate inter- and intra-rater reliability in two sessions. Consultation with experts has established content validity of the instrument. Reliability studies demonstrated an intra-class correlation coefficient (ICC) for inter-rater reliability of 0.86 for the activity score (95% confidence interval (CI) = 0.73–0.99) and of 0.92 for the damage score (95% CI = 0.85–1.00). The Spearman’s ρ (Sp) for intra-rater reliability for the activity score was 0.96 (95% CI = 0.89 to 1.00) and for the damage score Sp was 0.99 (95% CI = 0.97–1.00). Clinical responsiveness needs to be evaluated in a prospective clinical trial, which is ongoing.

Key words: clinical trial/cutaneous lupus erythematosus/discoid lupus erythematosus/outcome instrument/sub-acute lupus erythematosus

J Invest Dermatol 125:889–894, 2005

Characteristic inflammatory skin changes are seen in a large majority of individuals with systemic lupus erythematosus (SLE); however, the cutaneous manifestations of lupus erythematosus (LE) are among the least systematically studied aspects of these autoimmune illnesses. This has resulted in part from the lack of validated tools to determine the impact of therapy on the activity of the cutaneous manifestations of LE.

Disease Activity Measurements for Skin Disease

During the last three decades there have been numerous attempts to develop instruments to measure skin disease. As a good approximation of the available scores, the Medical Algorithm Project lists 108 medical algorithms in dermatology (www.medal.com, accessed 09/15/2004), the majority of which are scores to assess clinical disease. Unfortunately, most of the developed scores are disease-spe-

Abbreviations: CLASI, Cutaneous Lupus Erythematosus Disease Area and Severity Index; CLE, cutaneous lupus erythematosus; DLE, discoid lupus erythematosus; ICC, intra-class correlation coefficient; LE, lupus erythematosus; SLE, systemic lupus erythematosus

The study was conducted at the Department of Dermatology, University of Pennsylvania.

cific. There are some general scores, like the Dermatology Index of Disease Severity (DIDS) (Faust *et al*, 1997), but they are too crude in their assessment of body surface area to be useful for diseases like acne or cutaneous lupus erythematosus (CLE), which only affect relatively small areas of the skin (Williams, 1997).

In a recent retrospective study, Parodi *et al* (2000) noted that 60 measures of SLE were available, but only three of these appeared useful for dermatologists. Based on their experience with 176 patients with CLE, Parodi *et al* (2000) found even these scores inadequate for dermatologists and called for revision.

In order to facilitate future clinical trials, we decided to develop an outcome instrument for CLE.

Results

Literature review A more extensive review of the literature than is possible within the format of this paper will be published by the American College of Rheumatology Subcommittee on Response Criteria for CSLE.¹ Other important

¹Mosca M, Lockshin M, Schneider M, *et al*: ACR Response criteria for cutaneous lupus erythematosus in clinical trials, submitted.

considerations for the design of the CLASI (Cutaneous Lupus Erythematosus Disease Area and Severity Index) were ease of administration and non-reliance on invasive test.

Separate measurements of disease activity and damage The differentiation between activity and damage within the CLASI is unusual for dermatological scores, but established for scores of SLE. This separation leads to two scores for each patient. The alternative summary of the results in one total score would lead to implausibly stable results for scarring forms of lupus, as activity decreases and damage becomes apparent. As a result, a summary score may remain stable, whereas the clinical picture shifts completely. In addition, because either current activity or damage may have an important impact on the patient's quality of life and self-esteem, separate scores are calculated.

Area The CLASI describes the extent of disease in terms of the intensity of involvement of anatomical areas but does not record the percentage of body surface area or the number of lesions. Other scores like Psoriasis Area and Severity Index (PASI) or SCORAD (SCORing Atopic Dermatitis) depend heavily on assessment of the skin area, but this assessment has repeatedly been shown to be hard to reproduce (Tiling-Grosse and Rees, 1993; Charman *et al*, 1999; Charman and Williams, 2000). Because CLE, like guttate psoriasis, involves only small proportions of the skin, area assessment is even more difficult than in atopic dermatitis or plaque psoriasis (Ramsay and Lawrence, 1991). Lesion counting, commonly used for acne, is not a suitable alternative, because the reliability of the lesion counting is not necessarily better (Lucky *et al*, 1996) and the lesions of CLE have widely varying size and may be confluent. Consequently, improvement may lead to divisions of large lesions into smaller lesions, which would lead to a paradoxical increase of the score.

Erythema Our testing of the CLASI showed that activity scores were based largely on the extent of erythema. Erythema is a prominent, easily recognized, and clear-cut symptom of disease activity. It can be assessed reliably even on black skin. The clinical assessment of erythema has been compared to measurement of erythema by instruments, e.g., Doppler flowmeter, and visual assessment has been found to correspond well to the results with these "objective techniques" (Lahti *et al*, 1993; Quinn *et al*, 1993).

Associated symptoms Self reporting of symptoms is common in rheumatic diseases (Rider *et al*, 1997). For example, fatigue has been shown to be a critical factor in determining quality of life among SLE patients (Hochbert *et al*, 1990; Hanly, 1997). We did not integrate these associated symptoms into the score, because their relationship to the physician-assessed symptoms is unclear and we think they are best reported separately. Separation of reported and assessed symptoms is also used for other scores, e.g., the SCORAD (1993; Charman and Williams, 2000) and the Leicester score for atopic dermatitis (Berth-Jones *et al*, 1995). Results of the scores in clinical trials should be presented in terms of percentage change, as the most appropriate description (Farrar *et al*, 2003). Formal

Table I. CLASI scores of participating patients

	SLE type	Raters (N)	Activity score	Damage score
			Mean/median	Mean/median
1	SCLE	9	13.3/13.5	13.8/15.5
2	DLE	9	10.1/10	15.6/15.5
3	DLE	9	36.6/39	40.4/40.5
4	DLE	9	17.4/17.5	19.5/19
5	SCLE	9	13.5/13.5	1.9/2
6	SCLE	8	13/13.5	11.8/10.5
7	SCLE	8	12/11	0.7/0
8	DLE	8	18.2/17.5	11.8/12.5
9	DLE	8	40/37.5	41.6/42

CLASI, Cutaneous Lupus Erythematosus Disease Area and Severity Index; SLE, systemic lupus erythematosus; SCLE, subacute lupus erythematosus; DLE, discoid lupus erythematosus.

assessment of the associated symptoms is currently being conducted within a study to evaluate clinical responsiveness of the instrument.

Content validity Content validity and face validity were confirmed according to the above protocol. Where appropriate, changes were made in accordance with the suggestions.

Patients The patients were volunteers from the outpatient clinic of the department of dermatology of the University of Pennsylvania, which is a tertiary care center. The nine patients represent a full range of activity and damage, based initially on clinical impression, and subsequently confirmed by calculation of activity and damage scores. Their mean activity score ranged from 10.1 to 40 and the mean damage score ranged from 0.7 to 41.6 (Table I).

Inter-rater agreement Inter-rater reliability was high (Table II). The correlation coefficients between pairs of raters ranged from $r_p = 0.82$ to $r_p = 0.99$, indicating that the relative ordering of the patients based on the raters' scores are consistent between raters. In addition to high consistency among raters, the agreement among the raters' scores for a given patient is also high. The intra-class correlation coefficient (ICC) was $r = 0.86$ (95% confidence interval

Table II. Results of reliability testing

	Rater (n)	Patients (n)	
Inter-rater reliability ICC			
Activity	11	9	0.86 (95% CI = 0.73–0.99)
Damage	11	9	0.92 (95% CI = 0.85–1.00)
Intra-rater reliability Spearman's ρ			
Activity	8	4	0.96 (95% CI = 0.89–1.00)
Damage	8	4	0.99 (95% CI = 0.97–1.00)

ICC, Intraclass Correlation Coefficient; CI, confidence interval.

(CI) = 0.73–0.99) for the activity scale and $r = 0.92$ (95% CI = 0.85–1.00) for the damage scale. In addition, the ANOVA results do not show significant differences among the raters' overall scores (the activity F test = 0.32, $p = 0.97$; the damage score F test = 0.32, $p = 0.97$), whereas there were significant differences in overall scores among the patients (the activity score F test = 52.01, $p < 0.0001$; the damage score F test = 105.64, $p < 0.0001$).

Intra-rater reliability Both scales demonstrated high intra-rater reliability (Table II). The differences between first and second rating for the activity subscale ranged from 1 to 4 points of 13 to 15 point total score, but the mean difference between first and second rating was only 2.0 points on average. Slightly higher ratings were given on the second rating by 90% of the raters. Although the difference between the first and second rating in the activity scale was statistically significant ($t = -2.57$; $p < 0.03$), this difference is clinically meaningless. The differences between ratings for the damage scores were even smaller and ranged from 0 to 2 points, with a mean of 0. Four of the eight pairs of ratings were identical, two raters increased by 1 or 2 points, and two raters decreased by 1 or 2 points. The difference between ratings in the damage score did not reach statistical significance ($t = 0.00$; $p = 1.0$). Both scales also demonstrated high consistency throughout regarding the relative ordering of the patients at first and second rating. There were no outliers. These results were reflected in an Spearman's ρ (Sp) of $r_p = 0.96$ for activity (95% CI = 0.89–1.00) and an Sp of $r_p = 0.99$ (95% CI = 0.97–1.00) for the damage scale. The intercept and slope of the regression equation relating 1st scores to second scores show small differences for both subscales (i.e., intercepts of 2 for activity and 0 for damage), and a slope close to 1, ($b = 0.73$ for activity and $b = 0.89$ for damage) for predicting the second rating based on the result of the 1st rating.

Time The average time needed to conduct the patient assessment with the CLASI ranged from less than 1 min to 11 min with an average duration of 5.25 min and a median of 5 min; we did not find any significant variation over time as the experience with the instrument increased.

Discussion

As a first effort toward validating an outcome instrument for CLE, we assessed the CLASI for content validity, inter-rater validity, intra-rater validity, and practical applicability. The score performed excellently in all these aspects, which is especially relevant for multicenter studies that form the backbone of any therapeutic evaluation for this rare disease. It is also notable that the CLASI has not been developed within the framework of a particular clinical trial, because it has been suggested that instruments developed for a particular trial might bias the trial in favor of the treatment. This kind of effect has been demonstrated for schizophrenia (Charman and Williams, 2000).

The CLASI has been designed as one single instrument for at least three clinical entities that constitute CLE, i.e., discoid lupus erythematosus (DLE), subacute lupus erythematosus (SCLE), and SLE. We think that the limitation to one instrument for CLE is necessary and appropriate. There

are a number of CLE cases that may share characteristics of two or even all three groups of CLE. Although pure patient populations are always desirable, they do not reflect clinical reality. In an orphan disease like LE, an instrument that improves the power of clinical research is not helpful if it restricts the population to subgroups that cannot realistically be recruited for even medium-sized clinical trials.

CLE affects primarily photosensitive areas like the face, neck, and hands. Skin symptoms in these uncovered and thus visible areas are more worrisome to patients (Baughman and Sobel, 1970; Nichol *et al*, 1996; Williams, 1997; Krueger *et al*, 2001). To increase the score's responsiveness we have increased the weight assigned to these symptoms, relative to other, less frequently involved parts of the body. This may skew the score slightly away from a purely anatomical reflection of the skin surface. We think this choice is clinically appropriate, however, and may actually be more meaningful to both patient and clinician because it takes the non-linear association between score and severity into account (Ashcroft *et al*, 1999). The table below compares the percentage of the score derived from different areas of the body between the CLASI, the PASI, and the rule of 9's (Table III). The rule of 9's, a method of estimating surface area based on attributing a given percentage to each body part, is used to estimate involved body surface area, e.g., in psoriasis (Ramsay and Lawrence, 1991).

Finlay (1996) defined desirable criteria for an instrument to measure disease activity in atopic dermatitis, which are general enough to be helpful to assess the CLASI:

- (1) The method should be simple enough to use in a busy clinical setting.
- (2) The method should clearly separate scores derived from the observer and from the patient.
- (3) The signs chosen to be recorded should be amenable to change and should be unambiguous in their meaning and proven to be so. If the presence of two signs is highly correlated only one needs to be recorded.
- (4) Recording of area of involvement should be based on an assessment of the site of involvement rather than the virtually impossible task of determining an accurate total percentage involvement.
- (5) Validity testing including repeatability testing by the same and different observers must be carried out.

Table III. Relative weight of body area for CLASI, PAST and rule of q

Area	CLASI (activity and damage) (%)	PASI (%)	Rule of 9 (%)
Head	33	10	9
Chest	14	5	9
Abdomen	7	5	9
Back	14	10	18
Legs	14	40	36
Arms/hands	14	30	18
Mucous membrane	1		

CLASI, Cutaneous Lupus Erythematosus Disease Area and Severity Index; PASI, Psoriasis Area and Severity Index.

Select the score in each anatomical location that describes the most severely affected cutaneous lupus-associated lesion

E x t e n s i o n	activity		damage		Anatomical Location
	Erythema	Scale/ Hypertrophy	Dyspigmentation	Scarring/ Atrophy/ Panniculitis	
	0-absent 1-pink; faint erythema 2- red; 3-dark red; purple/violaceous/ crusted/ hemorrhagic	0-absent; 1-scale 2- verrucous/ hypertrophic	0-absent, 1-dyspigmentation	0 ... absent 1 ... scarring 2 ... severely atrophic scarring or panniculitis	
				See below	Scalp
					Ears
					Nose (incl. malar area)
					Rest of the face
					V-area neck (frontal)
					Post. Neck &/or shoulders
					Chest
					Abdomen
					Back, buttocks
					Arms
					Hands
					Legs
					Feet

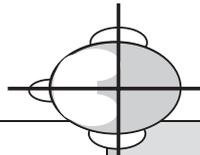
Mucous membrane

Mucous membrane lesions (examine if patient confirms involvement)
0-absent; 1-lesion or ulceration

Dyspigmentation

Report duration of dyspigmentation after active lesions have resolved (verbal report by patient ... tick appropriate box)
<input type="checkbox"/> Dyspigmentation usually lasts less than 12 months (dyspigmentation score above remains)
<input type="checkbox"/> Dyspigmentation usually lasts at least 12 months (dyspigmentation score is doubled)

Alopecia



Recent Hair loss (within the last 30 days/as reported by patient)		NB: if scarring and non-scarring aspects seem to coexist in one lesion, please score both
1-Yes 0-No		

Divide the scalp into four quadrants as shown. The dividing line between right and left is the midline. The dividing line between frontal and occipital is the line connecting the highest points of the ear lobe. A quadrant is considered affected if there is a lesion within the quadrant.

Alopecia (clinically not obviously scarred)	Scarring of the scalp (judged clinically)
0-absent 1-diffuse; non-inflammatory 2-focal or patchy in one quadrant; 3-focal or patchy in more than one quadrant	0- absent 3- in one quadrant 4- two quadrants 5- three quadrants 6- affects the whole skull

Total Activity Score
(For the activity score please add up the scores of the left side i.e. for Erythema, Scale/Hypertrophy, Mucous membrane involvement and Alopecia)

Total Damage Score
(For the damage score, please add up the scores of the right side, i.e. for Dyspigmentation, Scarring/Atrophy/Panniculitis and Scarring of the Scalp)

Figure 1
Cutaneous LE Disease Area and Severity Index (CLASI)

We believe that the CLASI fulfills the above criteria and is thus an appropriate instrument to assess the activity and damage caused by CLE. We hope, therefore, that it will be used in clinical trials of CLE. Preliminary findings indicate that the instrument has sufficient clinical responsiveness to be useful for clinical trials and formal assessment of these properties in clinical trials is forthcoming.

Materials and Methods

Literature review The design of the CLASI and its characteristics are based on a review of the literature on LE and outcome instruments used in dermatology. The most significant results of this review are presented in the results section.

Scoring of the CLASI (Fig 1) The CLASI consists of two scores, the first summarizes the activity of the disease, whereas the second is a measure of the damage done by the disease. Activity is scored on the basis of erythema, scale/hyperkeratosis, mucous membrane involvement, acute hair loss, and non-scarring alopecia. Damage is scored in terms of dyspigmentation and scarring, including scarring alopecia. Patients are asked whether dyspigmentation due to CLE lesions usually remains visible for more than 12 mo, which is taken to be permanent. If so, the dyspigmentation score is doubled. The scores are calculated by simple addition based on the extent of the symptoms. The CLASI is designed as a table where the rows denote anatomical areas, whereas the columns score major clinical symptoms. The extent of involvement for each of the skin symptoms is documented according to specific anatomic areas that are scored according to the worst affected lesion within that area for each symptom.

Associated symptoms Itch, pain and fatigue are recorded separately on visual 1–10 analog scales by the patients.

Assessment of content validity The instrument was assessed in informal interviews and email exchanges, by a group of 7 US Dermato-Rheumatologists, all of whom have considerable expertise with CLE and gave valuable feedback. We also solicited comments from the "American College of Rheumatology Response Criteria Committee on SLE", which consists of Rheumatologists and Dermatologists, in a session at Schloss Mickeln, Duesseldorf, Germany, in September of 2004. Additionally, we conducted a series of extensive standardized interviews with the raters to the point of redundancy after the assessment of the previous version of the score, in order to establish areas for improvement, mostly concerning the wording of the instrument.

Assessment of inter- and intra-rater reliability Inter-rater reliability was assessed by a group of 11 physicians (five board-certified dermatologist and six second- or third-year resident physicians) who scored a group of nine patients according to a predefined protocol in two sessions. In the first session, five patients were scored by nine physicians and in the second session, four patients were scored by eight physicians. Of the total of 11 scoring physicians, six scored all nine of the patients, three scored five patients, and two scored four patients. All physicians recorded the time they spent in each patient's room.

To assess intra-rater reliability, eight of 11 physicians (five board-certified dermatologist and three second- or third-year resident physicians) scored four of the nine patients. These patients also represented a wide range of activity levels (scores ranged from a mean of 11.9 to a mean of 17.8) and damage levels (between a mean of 0.8 and a mean of 11.9). In order to minimize recall, the physicians were not told at the beginning of the session that one additional patient had to be scored a second time at the end of the session to assess the intra-rater reliability.

Patients Nine patients were scored for this study. All had histologically confirmed CLE. Five of the patients had DLE and four

SCLE; two of the patients, one with SCLE and one with DLE, had systemic involvement.

Statistical methods/hypothesis We assessed the inter- and intra-rater reliability of the CLASI based on the two directly calculated subscale scores for activity and damage.

To assess inter-rater reliability we used the ICC, Pearson correlations, and ANOVA. The ICC was calculated using STATA (Version 8). An ICC of 0.5–0.7 was considered minimally acceptable, whereas an ICC above 0.81 is considered to be almost perfect (Landis and Koch, 1977). We had originally planned to use six raters with eleven patients per rater, which would have given us adequate power (80%) to detect an intra-class correlation based on a null hypothesis of an ICC of 0.5 and an alternative hypothesis of an ICC of 0.8 using an *F* test with a significance level of 0.05. Because fewer patients than expected could be recruited at any one time and the CLASI performed better than expected during the preliminary and first session, the final numbers of participants were a compromise to what had been originally planned. An ANOVA was used to test for general "rater effects".

To describe the magnitude and direction of the change in the patients' scores from first rating to 2nd rating of the same patients at the same session, summary statistics were calculated. To assess intra-rater reliability, we used correlations (Pearson and Sp), regressions, and *t* tests. Although we recognize the inherent weakness of the simple linear correlation as a measure of agreement, we report this in recognition of its common use as a measure of intra- and inter-rater reliability (Walter *et al*, 1998).

Ethics The protocol for the study was approved by the institutional review board of the University of Pennsylvania Medical School and is in accordance with the Declaration of Helsinki in its current form. All patients gave written consent before inclusion in the study.

- (1) Content validity committee: Richard D. Sontheimer, MD; Irwin M. Braverman, MD; Jeffrey P. Callen, MD; Melissa I. Costner, MD; Jan Dutz, MD; David Fivenson, MD; Joseph L. Jorizzo, MD; Lela A. Lee, MD; Daniel P. McCauliffe, MD.
- (2) The "American College of Rheumatology Ad Hoc Subcommittee on Response criteria for Cutaneous SLE": Marta Mosca, MD (Co-Chair), Michael Lockshin, MD, MPH (Co-Chair), Matthias Schneider (Co-Chair), Matthew H. Liang, MD, MPH (Co-Chair), Joerg Albrecht, MD, Martin Aringer, MD, Stefano Bombardieri, MD, Jill Buyon, MD, Ricard Cervera, MD, PhD, Paola de Pablo, MD, MPH, Barry Fessler, MD, Rebecca Fischer-Betz, MD, Dafna Gladman, MD, Nicolas Hunzelman, MD, Josef Kalden, MD, Thomas Krieg, MD, Annegret Kuhn, MD, PhD, Lela Lee, MD, David Norris, MD, Jutta Richter, MD, Animesh Sinha, MD, Josef Smolen, MD, Christof Specker, MD, Victoria P. Werth, MD.
- (3) This research was supported by NIH K24-AR 02207 (V. P. W.) and a research training grant in dermatology NIH 2T32-AR-007465 (J. A.).
- (4) We would like to thank the patients who volunteered their time and made this effort possible.

DOI: 10.1111/j.0022-202X.2005.23889.x

Manuscript received December 8, 2004; revised April 19, 2005; accepted for publication May 14, 2005

Address correspondence to: Victoria P. Werth, MD, Professor of Dermatology and Medicine, Department of Dermatology, 2, Rhodes Pavillion, 3600 Spruce Street, Philadelphia, PA 19119, USA. Email: Werth@mail.med.upenn.edu

References

- 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* 141:185–191, 1999

- Baughman RD, Sobel R: Psoriasis. A measure of severity. *Arch Dermatol* 101:390-395, 1970
- Berth-Jones J, Thompson J, Graham-Brown RA: Evening primrose oil and atopic eczema. *Lancet* 345:520, 1995
- Charman C, Williams H: Outcome measures of disease severity in atopic eczema. *Arch Dermatol* 136:763-769, 2000
- Charman CR, Venn AJ, Williams HC: Measurement of body surface area involvement in atopic eczema: An impossible task? *Br J Dermatol* 140:109-111, 1999
- Farrar JT, Berlin JA, Strom BL: Clinically important changes in acute pain outcome measures: A validation study. *J Pain Symptom Manage* 25:406-411, 2003
- Faust HB, Gonin R, Chuang TY, Lewis CW, Melfi CA, Farmer ER: Reliability testing of the dermatology index of disease severity (DIDS). An index for staging the severity of cutaneous inflammatory disease. *Arch Dermatol* 133:1443-1448, 1997
- Finlay AY: Measurement of disease activity and outcome in atopic dermatitis. *Br J Dermatol* 135:509-515, 1996
- Hanly JG: Disease activity, cumulative damage and quality of life in systematic lupus erythematosus: Results of a cross-sectional study. *Lupus* 6:243-247, 1997
- Hochbert MC, Engel EW, Pruitt A, *et al*: Correlations of disease activity with physical disability in SLE. *Arthritis Rheum* 222:S33, 1990 abstract
- Krueger G, Koo J, Menter A, Stern R, Rostad T: The impact of psoriasis on quality of life results of a 1998 National Psoriasis Foundation Patient-Member-ship Survey. *Arch Dermatol* 137:280-284, 2001
- Lahti A, Kopola H, Harila A, Myllyla R, Hannuksela M: Assessment of skin erythema by eye, laser Doppler flowmeter, spectroradiometer, two-channel erythema meter and Minolta chroma meter. *Arch Dermatol Res* 285:278-282, 1993
- Landis JR, Koch GG: The measurement of observer agreement for categorical data. *Biometrics* 33:159-174, 1977
- Liang MH, Socher SA, Larson MG, Schur PH: Reliability and validity of six systems for the clinical assessment of disease activity in systemic lupus erythematosus. *Arthritis Rheum* 32:1107-1118, 1989
- Lucky AW, Barber BL, Girman CJ, Williams J, Ratterman J, Waldstreicher J: A multirater validation study to assess the reliability of acne lesion counting. *J Am Acad Dermatol* 35:559-565, 1996
- Nichol MB, Margolies JE, Lippa E: The application of multiple quality of life instruments in individuals with mild-to-moderate psoriasis. *Pharmacoeconomics* 6:644-653, 1996
- Parodi A, Massone C, Cacciapuoti M, Aragone MG, Bondavalli P, Cattarini G, Rebora A: Measuring the activity of the disease in patients with cutaneous lupus erythematosus. *Br J Dermatol* 142:457-460, 2000
- Quinn AG, McLelland J, Essex T, Farr PM: Quantification of contact allergic inflammation: A comparison of existing methods with a scanning laser Doppler velocimeter. *Acta Derm Venereol* 73:21-25, 1993
- Ramsay B, Lawrence CM: Measurement of involved surface area in patients with psoriasis. *Br J Dermatol* 124:565-570, 1991
- 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* 40:1976-1983, 1997
- SCORAD: Severity scoring of atopic dermatitis: The SCORAD index. Consensus Report of the European Task Force on Atopic Dermatitis. *Dermatology* 186:23-31, 1993
- Streiner DL, Norman GR: *Health Measurement Scales, A Practical Guide to their Development and Use*, 2nd edn. Oxford: Oxford University Press, 1995
- Tiling-Grosse S, Rees J: Assessment of area of involvement in skin disease: A study using schematic figure outlines. *Br J Dermatol* 128:69-74, 1993
- Walter SD, Eliasziw M, Donner A: Sample size and optimal designs for reliability studies. *Stat Med* 17:101-110, 1998
- Williams HC: Is a simple generic index of dermatologic disease severity an attainable goal? *Arch Dermatol* 133:1451-1452, 1997