

Inter and Intraobserver Variability of Total Skin Thickness Score (Modified Rodnan TSS) in Systemic Sclerosis

PHILIP CLEMENTS, PETER LACHENBRUCH, JAMES SIEBOLD, BARBARA WHITE, STEVEN WEINER, RICHARD MARTIN, ARTHUR WEINSTEIN, MICHAEL WEISMAN, MAUREEN MAYES, DAVID COLLIER, FREDRICK WIGLEY, THOMAS MEDSGER, JR, VIRGINIA STEEN, LARRY MORELAND, MARY DIXON, MARY MASSA, EDWARD LALLY, DEBORAH McCLOSKEY, JOHN VARGA, FRANCES INGENITO, and DANIEL FURST

ABSTRACT. *Objective.* Assessment of the inter and intraobserver variability of the modified Rodnan (m-Rodnan) total skin thickness score by clinical palpation [a commonly used outcome measure in trials of systemic sclerosis (SSc)].

Methods. Skin thickness was assessed by clinical palpation of 17 body areas on a 0 to 3 scale (normal, mild, moderate, severe). The m-Rodnan total skin thickness score was derived by summation of the scores from all 17 body areas. Using the m-Rodnan, 6-7 investigators assessed skin thickness in 5-6 patients with SSc (22 patients and 23 examiners total) at each of 4 sessions for the determination of interobserver variability (accuracy). In addition 21 of the investigators then assessed m-Rodnan in 2-3 patients each (60 patients total) 3 times over a 2-8 week period to quantitate intraobserver variability (reliability).

Results. Interobserver and intraobserver mean \pm within patient standard deviations (SD) for the m-Rodnan were found to be 17.7 ± 4.6 and 20.7 ± 2.45 , respectively.

Conclusion. The m-Rodnan total skin thickness score is at least as reliable for measuring skin thickness in SSc as are the ARA and Ritchie joint tenderness counts for assessing joint disease in rheumatoid arthritis. These data are useful for the determination of sample size and for the definitions of clinically meaningful response. Assessment of skin score is sufficiently reproducible to include as a measure of disease outcome, especially if patients are serially evaluated by the same investigator. (*J Rheumatol* 1995;22:1281-5)

Key Indexing Terms:

SCLERODERMA SKIN SCORE SKIN THICKNESS SYSTEMIC SCLEROSIS

Tightening and thickening of the skin (scleroderma) is a hallmark of systemic sclerosis (SSc)^{1,2}. Assessment of the extent of skin involvement provides the basis for classification of disease into diffuse cutaneous and limited cutaneous syndromes¹⁻³. Extent and severity of skin thickening have

been related to survival, risk of accrual of visceral involvement and overall disease progressivity³⁻⁶.

The role of serial assessment of skin involvement by semi-quantitative skin scoring, employing simple clinical palpation of the skin, remains controversial. While skin scoring

From University of California at Los Angeles, School of Medicine, Department of Medicine, Division of Rheumatology; and School of Public Health, Department of Biostatistics, Los Angeles, CA, USA.

Supported in part by grants from United Scleroderma Foundation, Scleroderma Federation, FDA Orphan Drug Program FD-R-000863, Arthritis Foundation Clinical Science Grant program, GCRC grants M01 RR00827 and RR00865, NIH UCLA Multipurpose Arthritis and Musculoskeletal Diseases Center grant P60 AR-36834.

P.J. Clements, MD, Professor of Medicine, Rheumatology, UCLA School of Medicine; D.H. Collier, MD, Chief of Rheumatology, Denver General Hospital, Associate Professor of Medicine, University of Colorado Health Sciences Center; M. Dixon, RN, BHS, Clinical Research Associate, Agouron Pharmaceuticals Inc, La Jolla, CA; D.E. Furst, MD, Clinical Professor of Medicine, Virginia Mason Research Center, Seattle; F. Ingenito, APRN, Adult Nurse Practitioner, Rheumatic Diseases, University of Connecticut Health Center; P.A. Lachenbruch, PhD, Chief of Biostatistics Branch, Division of Biostatistics and Epidemiology, CBER/FDA, Rockville, MD; E.V. Lally, MD, Director of Rheumatology, Brown University; R.W. Martin, MD, Assistant Professor of Medicine, Michigan State University, Grand Rapids, MI; M. Massa, MD, Associate Professor of Medicine, Dermatology, Loyola University of Chicago; M.D. Mayes, MD, Associate Professor, Rheumatology, Wayne State University;

D.A. McCloskey, RN, Coordinator, Scleroderma Program, Robert Wood Johnson Medical School; T.A. Medsger, Jr, MD, Professor of Medicine, Rheumatology and Clinical Immunology, University of Pittsburgh School of Medicine; L.W. Moreland, MD, Associate Professor of Medicine, Clinical Immunology and Rheumatology, University of Alabama at Birmingham; J.R. Seibold, MD, Chairman, Clinical Pharmacology, Robert Wood Johnson Medical School; V. Steen, MD, Associate Professor of Medicine, Rheumatology and Clinical Immunology, University of Pittsburgh School of Medicine; J. Varga, MD, Associate Professor of Medicine, Rheumatology Research, Thomas Jefferson University; S. Weiner, MD, FACP, Associate Clinical Professor of Medicine, Rheumatology, UCLA School of Medicine; A. Weinstein, MD, Professor of Medicine, Rheumatic Diseases and Immunology, New York Medical College; M.H. Weisman, MD, Professor of Medicine, UCSD School of Medicine; B. White, MD, Associate Professor of Medicine, Rheumatology and Clinical Immunology, University of Maryland School of Medicine; F.M. Wigley, MD, Associate Professor of Medicine, Molecular and Clinical Rheumatology, The Johns Hopkins University School of Medicine.

Address reprint requests to Dr. P. Clements, UCLA Medicine/Rheumatology, 1000 Veteran Ave., Rm. 32-48, Los Angeles, CA 90024-1670.

Submitted May 2, 1994 revision accepted November 4, 1994.

is frequently employed in clinical trials and reported as a measure of outcome^{5,7,8}, there are lingering concerns about whether skin scores are valid independent measures of outcome⁹.

The original Rodnan total skin thickness score assessed cutaneous thickening in 26 body areas utilizing a rating of 0 (normal thickness) to 4 (extreme thickening)^{4,10}. This scale included areas felt to be difficult to assess reliably (toes), areas that are normally relatively thickened (upper back) and overrepresented anatomic areas of relatively small body surface area (right and left breasts individually as opposed to chest and abdomen). Accordingly, a modified Rodnan (m-Rodnan) total skin thickness score has been in widespread use in which 17 anatomic areas (face, anterior chest, abdomen and, right and left separately; the fingers, hands, forearms, upper arms, thighs, lower legs, and feet) are scored by clinical palpation using a 0 to 3 scale (0 is normal and 1–3 reflect mild, moderate, and severe skin thickening)¹¹. Validation of the m-Rodnan method has been reported in abstract form¹².

This study was undertaken by those members of the Scleroderma Clinical Trials Consortium who are participants in the ongoing High-Dose vs Low-Dose D-penicillamine in Early, Diffuse Scleroderma Trial^{9,11} in the United States, which incorporates the m-Rodnan as a measure of outcome. We sought information as to the accuracy (interobserver variability) and reproducibility (intraobserver variability) of the m-Rodnan. Similar data were gathered on several other physical measures commonly used to evaluate scleroderma (maximum oral aperture, hand extension, fist closures)¹³. We further sought to analyze potential sources of variability and to assess whether further simplifications in total skin thickness scoring techniques are warranted.

MATERIALS AND METHODS

Patients. All patients in this study met 1980 American College of Rheumatology preliminary criteria for classification as definite SSc¹⁴. Eighty-two subjects with SSc were evaluated including 75 with diffuse cutaneous scleroderma (skin thickening of any degree proximal to the elbows and knees, often involving the torso) and 7 with limited cutaneous scleroderma (skin thickening of any degree limited to sites distal to the elbows and knees, with or without facial involvement)³.

Skin assessments. All skin scores were performed utilizing the m-Rodnan total skin thickness score technique¹¹. Participants were instructed by one of 2 participants (JRS, DEF) either of whom or both attended each session as described below. Participants were asked to assess skin thickening by clinical palpation as an isolated clinical finding and specifically requested not to consider skin immobility (tethering) as part of their scoring. The 17 anatomic areas were rated as 0 (normal skin thickness), 1+ (mild but definite skin thickening), 2+ (moderate skin thickening) and 3+ (severe skin thickening) and the m-Rodnan derived by summation of the scores from all 17 areas (range 0–51).

Instruction was given in techniques of measurement of maximum oral aperture (maximum vertical interlabial distance in millimeters during 3 sequential efforts of the patient at maximum active oral opening); active hand extension (maximum distance in millimeters from the external most point of the thumb to the external most point of the 5th finger during maximum unsupported active hand extension by the patient); and active fist

closures (shortest distance in millimeters from the tip of the 4th finger to the distal palmar crease during 3 sequential efforts at maximum fist closure). Standardized transparent millimeter rules were provided.

Study design. Interobserver variability. Four separate sessions were held (UCLA, Pittsburgh, UMDNJ twice) each attended by 6 or 7 examiners¹¹. With the exception of one examiner who attended 3 sessions and 3 examiners who each attended 2 sessions, each examiner attended only 1 session. Following didactic instruction in m-Rodnan technique, one or 2 patients were examined as teaching cases in a blinded, randomized manner by the 6 or 7 examiners, after which all examiners discussed their findings, reexamined the patient if necessary and came to a consensus about the grading of each of the body areas examined. Agreement on techniques and definitions was reached on each of the 2 teaching patients and all examiners had the opportunity to reexamine those areas in which disagreement had initially occurred. The remaining 5 or 6 patients were examined separately by each examiner, blinded to the other examiners' results and without consultation. To assure uniformity across the 4 separate scoring sessions at different sites, one or both instructor-examiners (DEF or JES) attended all 4 sessions and communicated the consensus opinion; written notes were kept of definitions which were communicated to each of the separate sessions. A total of 20 patients with diffuse cutaneous (d)SSc and 2 with limited cutaneous (l)SSc were examined by 23 examiners.

Intraobserver variability. Each investigator returned to their clinic where they assessed the m-Rodnan in 3 patients on 3 separate occasions each within a 2 to 8 week period without reference to previous examinations. For the purposes of this exercise, it was assumed that skin scores of individual patients did not change significantly over the 2 to 8 week study period. A total of 55 patients with dSSc and 5 with lSSc were examined by 21 examiners.

Analysis. Definitions. Unless otherwise specified the term mean \pm SD refers to the measurement of examiner, patient and residual variation inclusive, while the term mean \pm within patient SD refers to (a) the measurement of among and within examiner and residual variation inclusive (excluding patient variation, which is assumed to be negligible) in the interobserver study and (b) the measurement of within examiner and residual variation (excluding patient variation, which is assumed to be negligible) in the intraobserver study. Coefficients of variation were calculated as follows: (mean \div SD) multiplied by 100.

Interobserver variability study. The amount of disagreement among examiners assessing patients with a range of skin involvements was analyzed by calculating and pooling the standard deviations (SD) of the m-Rodnan total skin thickness scores and of the other physical measures of skin involvement (maximum oral aperture, active hand extension and active fist closure) recorded on each patient by the 6 or 7 examiners (within patient SD). The mean, within patient SD and variance of all m-Rodnan skin scores and of the other physical measurements on a single patient were calculated. Overall mean values for the m-Rodnan and the other physical measures were calculated by averaging the mean values of all the patients examined. The overall within patient SD was derived by calculating the square root of the sum of the variances divided by the number of patients evaluated. This overall within patient SD provides a quantitative assessment of interobserver disagreement. The results were corrected for multiple observations by the 4 examiners who attended more than one session.

Intraobserver variability study. The mean, within patient SD and variances of all 3 skin thickness score measurements made by an examiner on each patient were calculated. An overall mean skin thickness score for all 3 of the patients with SSc assessed by each examiner was calculated by averaging the mean skin thickness scores for all 3 patients. The overall intraobserver within patient SD for each examiner and for the entire group of 21 examiners were derived by calculating the square roots of the sum of the variances divided by the number of patients evaluated in that study (i.e., for one investigator the within patient SD equals the square root of the sum of the variances of the 3 patients after dividing the sum by 3; for all 21 investigators, it equals the square root of the sum of variances derived from the 21

individual calculations above after dividing the sum by 21). This overall within patient SD provides a quantitative assessment of the intraobserver disagreement.

RESULTS

Primary analyses. Interobserver variability. The analysis of the overall mean m-Rodnan total skin thickness score and the overall within patient SD (17.7 ± 4.6) for the interobserver variability study (Table 1) has been published¹¹. Since it obviates error introduced by among patient variability, the calculation of within patient SD (derived from assessments on a single patient by multiple examiners) provides the best measure of interobserver variability.

During the interobserver variability study, the maximum oral aperture, active hand extension, and active fist closure were also measured. The overall mean \pm within patient SD for each of these variables is shown in Table 1.

Intraobserver variability. The intraobserver mean and within patient SD of the m-Rodnan for each investigator (each of whom examined 2–3 patients) are displayed in Table 2. The overall mean \pm within patient SD was 20.7 ± 2.45 , with a coefficient of variation of 12% (Table 2). Similarly since errors introduced by among patient and among examiner variability are eliminated, the calculation of within patient SD (derived by serial measurements on a single patient by a single examiner) provides the best measure of intraobserver variability.

Exploratory analyses. The potential effects of the training sessions were considered. The mean and within patient SD of the m-Rodnan for the 6 teaching patients (using the original, blinded and uncorrected scores) and the 23 examiners were calculated (20.6 ± 4.0) and were compared to the larger group of 22 patients with SSc examined as part of the blinded, randomized variability study (17.7 ± 4.6 , differences not significant). While the small interobserver variability for both exercises followed a didactic teaching session, these

Table 1. *Interobserver variability of m-Rodnan total skin score and other physical measures of skin involvement in 20 patients with SSc*

	Mean	Within PT SD*	Coefficient of Variation (%)**
Modified Rodnan total skin thickness score	17.7	± 4.6	25
Maximum oral aperture (mm)	48.5	± 3.1	6
Active hand extension (mm)			
Right	163.9	± 6.8	4
Left	168.6	± 6.8	4
Active fist closure (mm)			
Right	32.5	± 5.9	18
Left	32.3	± 6.0	19

* SD = standard deviation.

** Coefficient of variation = $\sigma/\mu \times 100$ or $[(\text{mean} \div \text{SD}) \times 100]$.

Table 2. *Intraobserver variability (mean and within patient SD) of m-Rodnan skin thickness score (one examiner assessing 3 patients on 3 separate occasions)**

Examiner	Mean	Within Patient SD
E1	21.7	1.73
E2	28.6	3.22
E3	22.8	2.77
E4	21.5	3.16
E5	44.1	3.70
E6	24.3	2.43
E7	9.5	2.83
E8	14.9	2.75
E9	18.3	1.97
E10	11.2	1.05
E11	20.8	3.61
E12	18.1	3.13
E13	25.4	1.73
E14	16.1	1.60
E15	19.7	1.60
E16	22.9	3.20
E17	14.0	1.70
E18	24.3	2.21
E19	36.3	0.58
E20	18.7	1.91
E21	8.9	0.88
Mean	20.7	2.45

* 17 body areas assessed, 0–3 scale. SD = standard deviation.

results suggest that the clinical practicum did not improve or change interobserver examination techniques.

The physician participants in this study included senior experience clinicians and relatively junior individuals. We could detect statistically significant differences in the m-Rodnan scores of only 2 junior investigators during the interobserver variability study. By the time these 2 investigators had returned home to complete their intraobserver studies, their variability data were well within the limits of the senior investigators. The results of the two instructors were very close to the mean of all observers.

Other differences were noted in individual physician approaches to use of skin scoring that we were not able to quantify in this study: these include a tendency by some to confuse thickening with tethering; a general inability to discriminate the contribution of local edema to cutaneous thickening; and apparent difficulty in assessing legs and feet in comparison to the upper extremities.

Several *post hoc* analyses were performed to determine whether the m-Rodnan could be shortened to include fewer than 17 sites. While complete examination of the skin is clearly appropriate for patient care, a shortened but efficient skin scoring technique would be useful for studies, as it might speed data collection. It would only be useful, however, if significant information is not lost. To test for this possibility, several analyses were performed to determine whether fewer than 17 sites could be used to predict the full m-Rodnan skin thickness score in the 20 patients with dSSc examined in the interobserver variability study; right limbs (the sum

of the 7 body areas of the right upper and lower limbs); left limbs (analogous to the right above); both upper limbs (the sum of 8 areas on both upper limbs); right upper limb (the sum of 4 body areas on the right upper limb); left upper limb (analogous to the right upper limb); upper body only (the sum of 11 areas: face, chest, abdomen, and both upper arms, forearms, hands and fingers); lower body only (the sum of 6 areas: both thighs, lower legs, and feet); central body only (the sum of 6 areas: both upper arms and thighs, chest and abdomen); and distal body only (the sum of 11 areas: face and both fingers, hands, forearms, lower legs, and feet).

The mean \pm SD m-Rodnan total skin thickness score derived from these alternate methods of summing skin scores (150 patient-examiner interactions) are compared to the full m-Rodnan in Table 3 along with their partial correlations (in which the effect of observers and patients were removed). Not unexpectedly, all the alternative scoring techniques were highly correlated with the m-Rodnan ($p < 0.0001$). The best correlations with the m-Rodnan scores were derived from examination of the limbs (0.92), indicating that limb (upper plus lower) gives nearly as much information as adding truncal data. Since symmetry existed, it does not appear to matter which side of the body was examined. Other reduced scores were less closely correlated with the m-Rodnan, although upper body was more closely correlated than lower body (0.86 vs 0.77). These alternative skin score summing methods deserve to be evaluated in future studies as simpler methods of assessing skin thickness in patients with dSSc.

Skin thickening of body skin areas of the right upper and lower limbs (fingers, hands, forearm, upper arms, thighs, legs, and feet) were individually compared to those of the left side to see if skin thickening was a symmetrical process. With the exception of the dorsum of the feet, the right side was not significantly different from the left: the mean score \pm SD of 0.49 ± 0.81 of the left foot was significantly greater than that of 0.42 ± 0.75 of the right ($p = 0.004$).

Table 3. *Partial correlations* of 9 alternative skin score summing methods with the m-Rodnan total skin score as assessed in 20 patients with diffuse cutaneous scleroderma (all $p < 0.0001$)*

	# Areas	Skin Score	Correlation with m-Rodnan
m-Rodnan	17	17.7 ± 8.4	1.00
Left limbs	7	7.6 ± 3.8	0.92
Right limbs	7	7.6 ± 3.9	0.92
Upper body	11	13.5 ± 6.1	0.86
Distal	11	12.1 ± 5.8	0.86
Sum of both upper limbs	8	10.9 ± 4.9	0.86
Lower body	6	4.3 ± 4.0	0.77
Proximal	6	5.9 ± 4.1	0.75
Right upper body	4	5.9 ± 2.5	0.69
Left upper body	4	5.4 ± 2.4	0.67

* Correlation of variables after the effect of patients and observers have been removed.

DISCUSSION

Our study demonstrates that following a didactic teaching session, performance of the m-Rodnan by rheumatologists is both acceptably accurate and reproducible. The coefficient of variation for interobserver variability (accuracy) of the m-Rodnan is 25%. The coefficient of variation for intraobserver variability (reproducibility) is only 12%. These results are actually considerably better than standard clinical techniques of assessing rheumatoid arthritis (Ritchie indices and American Rheumatism Association joint counts) in which the coefficients of variation are 37 and 43%, respectively¹⁵.

Similar results were noted for other commonly employed physical measures reflective of skin involvement in SSc including active hand extension and maximal oral aperture. Active fist closure had appreciably higher variation, in part attributable to the diversity of hand deformities in the patients studied but also to variability in examiner technique.

The relative precision of these accessible and inexpensive clinical measures suggests that they are appropriate variables to consider in serial assessments of patients with SSc as outcome measures in interventional trials. Our observation of lower intraobserver variability than interobserver variability emphasizes the appropriateness of using the same examiner throughout the course of a clinical trial.

Skin involvement in SSc is locally symptomatic, interferes with joint motion, and impairs both functional and cosmetic status. Nonetheless, its overall importance in assessing SSc outcome remains unclear. The present data suggest that clinical assessment of skin thickening is a valid measure of outcome and that the m-Rodnan total skin thickness score is sufficiently precise to be studied in comparison to other measures of outcome including patient self-assessment, activities of daily living, visceral disease, and survival. We suggest that m-Rodnan continue to be employed in SSc trials and that data be sought that permit relevant comparisons to other outcomes. It is not clear, for example, what level of improvement in skin thickening might be judged "clinically meaningful" either as a surrogate for disease progressivity or as an individual disease manifestation.

It is taught that the skin thickening of SSc is a symmetrical process, a contention supported by Rodnan in his study of forearm punch biopsies: the weights of paired skin biopsy cores taken from the right and left forearms were within 10 mg of each other in 16 of 21 patients with SSc biopsied¹⁰. Although we did not specifically analyze symmetry within individual patients, our study employing clinical assessment of skin thickness supports this conclusion for the group as a whole. While not a major portion of the exploratory data analyses, symmetry of the left and right sides was good with no significant differences between left and right among the 7 areas examined except in the feet, where a 2% mean difference was found (data not shown). While statistically differ-

ent ($p = 0.004$), this degree of difference is not clinically meaningful.

Not surprisingly, we found that the results of all 9 alternate methods of scoring were highly correlated with the results of the m-Rodnan technique. Since the correlation of coefficients for the limbs accounted for 85% of the data from the full m-Rodnan score and all other approaches accounted for only 45–74%, the best of the reduced skin scores probably involves examination of the limbs alone. Our data also suggest that the variability of the m-Rodnan would be minimally compromised by further reducing the number of body areas assessed.

We are concerned, however, that a reduction in the number of areas, and thus a reduction in the range of skin scores possible, could reduce the sensitivity of the measure in serial assessments. For example, the proximal body skin score (Table 3), employing 6 anatomic areas, offers a range of only 0 to 18; but in this group of patients with dSSc it produced a skin score of 5.9 ± 4.1 . An improvement of 33% in a subject with a skin score of 18 is an absolute improvement of 6 but this absolute improvement of 6 is also a complete remission in nearly half the subjects who had skin scores < 5.9 , an unlikely result. In addition, in a previous analysis of m-Rodnan in which variability of high and low scores were compared, the within patient SD was constant, regardless of how high or low the mean scores were¹¹. The reduction in the number of areas could also result in the loss of ability to detect change in a patient, which then could result in a false-negative result about the efficacy of a study treatment. Since this analysis was an exploratory cross sectional exercise, the findings will need to be tested in other trials and longitudinally before any of these scaled down scoring techniques can be accepted as equivalent to the m-Rodnan score.

The present data demonstrate that intraobserver variability for clinical skin scoring in SSc is low and the technique is reproducible and accurate. Clinical researchers in rheumatology perform skin scores with an accuracy and precision comparable to or better than accepted measures of outcome in rheumatoid arthritis (i.e., joint counts). Skin scoring should be included as an outcome measure in clinical trials for SSc as an accessible, noninvasive, inexpensive, and reliable tool for assessing cutaneous disease status and changes over time^{2,16}. Skin scoring can be viewed as a clinical surrogate marker for fibrosis although its value as such will need to be validated by prospective trials linking its behavior to measures of internal organ dysfunction.

ACKNOWLEDGMENT

The authors acknowledge Nancy Allegar, MD, Brian Andrews, MD, and Mary Unanue, MD, for their participation.

REFERENCES

1. Seibold JR: Scleroderma. In: Kelley WN, Harris ED, Ruddy S, Sledge CB, eds. *Textbook of Rheumatology*, 4th ed. Philadelphia: WB Saunders, 1993:1113–43.
2. Medsger TA Jr: Systemic sclerosis (scleroderma). Localized forms of scleroderma and calcinosis. In: McCarty DJ, Koopman WJ, eds. *Arthritis and Allied Conditions*, 12th ed. Philadelphia: Lea and Febiger, 1993:1253–92.
3. LeRoy EC, Black C, Fleischmajer R, et al: Scleroderma (systemic sclerosis): Classification, subsets and pathogenesis. *J Rheumatol* 1988;15:202–5.
4. Steen VD, Medsger TA Jr, Rodnan GP: D-penicillamine therapy in progressive systemic sclerosis (scleroderma). *Ann Intern Med* 1982;97:652–8.
5. Clements PJ, Lachenbruch PA, Ng SC, Simmons M, Sterz M, Furst DE: Skin score. A semiquantitative measure of cutaneous involvement that improves prediction of prognosis in systemic sclerosis. *Arthritis Rheum* 1990;33:1256–63.
6. Steen VD: Factors predicting the development of renal involvement in progressive systemic sclerosis. *Amer J Med* 1984;76:779–86.
7. Jimenez SA, Sigal SH: A 15-year prospective study of treatment of rapidly progressive systemic sclerosis with D-penicillamine. *J Rheumatol* 1991;18:1496–503.
8. Clegg DO, Reading JC, Mayes MD, et al: Comparison of aminobenzoate potassium and placebo in the treatment of scleroderma. *J Rheumatol* 1994;21:105–10.
9. Seibold JR, Furst DE, Clements PJ: Why everything (or nothing) seems to work in the treatment of scleroderma (editorial). *J Rheumatol* 1992;19:673–767.
10. Rodnan GP, Lipinski E, Luksick J: Skin thickness and collagen content in progressive systemic sclerosis (scleroderma) and localized scleroderma. *Arthritis Rheum* 1979;22:130–40.
11. Clements PJ, Lachenbruch PA, Seibold JR, et al: Skin thickness score in systemic sclerosis: An assessment of interobserver variability in 3 independent studies. *J Rheumatol* 1993;20:1892–6.
12. Ramsden MF, Goldsmith CH, Lee P, Sebaldt R, Baer P: Clinical assessment of scleroderma: Observer variation in five methods (abstr). *Arthritis Rheum* 1986;29:S61.
13. Pope J, Bellamy N: Outcome measurement in scleroderma clinical trials. *Semin Arthritis Rheum* 1993;23:22–33.
14. Subcommittee for Scleroderma Criteria of the American Rheumatism Association Diagnostic and Therapeutic Committee: Preliminary criteria for the classification of systemic sclerosis (scleroderma). *Arthritis Rheum* 1980;23:581–90.
15. Eberl DR, Fasching B, Rahlfs V, Schleier I, Wolf R: Repeatability and objectivity of various measurements in rheumatoid arthritis. *Arthritis Rheum* 1976;19:1278–86.
16. Clements P, Lachenbruch P, Furst D, Paulus H: The course of skin involvement in systemic sclerosis over three years in a trial of chlorambucil versus placebo. *Arthritis Rheum* 1993;36:1575–9.