ARTHRITIS & RHEUMATISM
Vol. 43, No. 11, November 2000, pp 2445–2454

SKIN THICKNESS SCORE AS A PREDICTOR AND CORRELATE OF OUTCOME IN SYSTEMIC SCLEROSIS

High-Dose Versus Low-Dose Penicillamine Trial

PHILIP J. CLEMENTS, ERIC L. HURWITZ, WENG KEE WONG, JAMES R. SEIBOLD, MAUREEN MAYES, BARBARA WHITE, FREDRICK WIGLEY, MICHAEL WEISMAN, WALTER BARR, LARRY MORELAND, THOMAS A. MEDSGER, JR., VIRGINIA D. STEEN, RICHARD W. MARTIN, DAVID COLLIER, ARTHUR WEINSTEIN, EDWARD LALLY, JOHN VARGA, STEVEN R. WEINER, BRIAN ANDREWS, MICHA ABELES, and DANIEL E. FURST

Objective. **To study the clinical implications of a skin thickness score ≥20 at first visit and of softening of sclerodermatous skin in a cohort of systemic sclerosis (SSc) patients with diffuse cutaneous scleroderma.**

Methods. **Skin and visceral involvement were assessed in 134 SSc patients with diffuse scleroderma (mean ± SD duration of SSc 10 ± 4 months) as they entered a multicenter drug trial and again at 2 years of followup. Advent of mortality and scleroderma renal crisis (SRC) were assessed during a followup of 4.0 ± 1.1 years (mean ± SD). Logistic and linear regression were used to examine the relationship of baseline skin score to morbidity, mortality, and visceral involvement and the relationship of changes in skin score to changes in physical examination, laboratory, and functional variables over 2 years.**

Results. **A baseline skin score ≥20 was associated with heart involvement at baseline (odds ratio [OR] 3.10, 95% confidence interval [95% CI] 1.25–7.70) and was predictive of mortality (OR 3.59, 95% CI 1.23–10.55) and SRC (OR 10.00, 95% CI 2.21–45.91) over 4 years. Multivariate linear regression demonstrated that a model with skin score at baseline ($P = 0.0078$) and changes in large joint contractures ($P = 0.0072$), tender joint counts ($P = 0.0119$), handspread ($P = 0.0242$), and Health Assessment Questionnaire disability index (HAQ-DI) ($P = 0.0244$) explained the change in skin score over 2 years ($R^2 = 0.567$). Multivariate logistic regression demonstrated that the investigator's global assessment of improvement was best explained by a model with skin score and HAQ-DI ($R^2 = 0.455$).**

Conclusion. **A baseline skin score ≥20 was associated with heart involvement at baseline and predicted mortality and SRC over the subsequent 4 years. Improvement in skin score in these patients with diffuse cutaneous scleroderma was associated with improvement in hand function, inflammatory indices, joint contractures, arthritis signs, overall functional ability,**

Supported in part by grants from the Scleroderma Federation, the United Scleroderma Foundation, the FDA Orphan Drug Program, and the Arthritis Foundation, CRC grants M01RR-00865 and M01RR-00827, and bequests from the estates of Winifred Krause, Morris Goldsmith, and Takako Ito.

Philip J. Clements, MD, MPH, Steven R. Weiner, MD: University of California, Los Angeles, School of Medicine; Eric L. Hurwitz, PhD, Weng Kee Wong, PhD: University of California, Los Angeles, School of Public Health; James R. Seibold, MD: University of Medicine and Dentistry of New Jersey–Robert Wood Johnson Medical School, New Brunswick; Maureen Mayes, MD, MPH: Wayne State University, Hutzel Hospital, Detroit, Michigan; Barbara White, MD: University of Maryland School of Medicine, Baltimore; Fredrick Wigley, MD: The Johns Hopkins University, Baltimore, Maryland; Michael Weisman, MD: Cedars–Sinai Medical Center, Los Angeles, California; Walter Barr, MD: Northwestern University Medical School, Chicago, Illinois; Larry Moreland, MD: University of Alabama at Birmingham; Thomas A. Medsger, Jr., MD: University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania; Virginia D. Steen, MD: Georgetown University Medical Center, Washington, DC; Richard W. Martin, MD: Michigan State University College of Human Medicine, Grand Rapids; David Collier, MD: University of Colorado Health Sciences Center, Denver; Arthur Weinstein, MD: Georgetown University Medical Center, Washington, DC; Edward Lally, MD: Brown University School of Medicine, Roger Williams General Hospital, Providence, Rhode Island; John Varga, MD: University of Illinois, Chicago; Brian Andrews, MD, PhD: University of California, Irvine; Micha Abeles, MD: University of Connecticut Health Center, Farmington; Daniel E. Furst, MD: Virginia Mason Research Center, Seattle, Washington.

Address reprint requests to Philip J. Clements, MD, MPH, University of California, Los Angeles, Medicine—Rheumatology, 32-59 Rehabilitation Center, Box 951670, Los Angeles, CA 90095-1670.

Submitted for publication April 11, 2000; accepted in revised form July 17, 2000.

and the examining investigator's global assessment of improvement.

Tightening and thickening of the skin (scleroderma) is a hallmark feature of systemic sclerosis (SSc). The extent of scleroderma is the major criterion for classification of SSc into 2 principal subgroups (diffuse and limited disease) (1). The clinical impact of skin thickening is incompletely understood. The overall extent of skin involvement has been shown to influence patient function (2–4), range of motion (2,4,5), and survival (6,7). Other data suggest that skin thickening serves as an effective clinical surrogate for assessment of future progression of the disease; greater extent and severity of skin thickening are predictive of renal and cardiac involvement, decreased survival, and increased disability (4,6–8).

Veteran clinical observers of SSc recognize that the clinical behavior of skin thickening in the individual patient with diffuse cutaneous involvement is unpredictable, but also recognize that trends in groups of patients are poorly quantified. The consensus is that skin involvement worsens in the early stages of the illness and is followed by a period of relative lack of change (plateauing), and then by spontaneous improvement during the later stages of the disease (9–11). These observations parallel the sequential pathologic changes that occur, such as edema and inflammation, fibrosis and induration, and atrophy.

Little is known of the prognostic significance of the disease phase involving skin softening. Steen recently noted that the 5- and 10-year survival among SSc patients with diffuse scleroderma whose skin scores declined over a 2-year study period were significantly better than the rates of survival in a similar group whose skin involvement did not improve (12). Seibold et al noted, in a 6-month placebo-controlled trial of the antifibrotic therapy recombinant human relaxin, that softening of skin was paralleled by improved patient function and finger motion (13). The absence of therapies proven effective against skin thickening has impaired our ability to quantify the clinical meaningfulness of a reduction in skin thickness.

The present study explores the clinical implications of softening of sclerodermatous skin by demonstrating that skin softening is associated with improvements in multiple physical, functional, and/or other outcomes. Changes in skin thickness scores over a 2-year period were related to the course of other SSc-associated parameters in a cohort of 134 SSc patients who were being followed up as part of a therapeutic trial. Herein we present data that suggest that improvement in skin thickening is associated with improvement in overall functional capacity, hand function, signs of systemic inflammation, and other aspects of SSc cutaneous and musculoskeletal involvement (i.e., large joint contractures, oral aperture, and joint tenderness and swelling).

PATIENTS AND METHODS

Patients. To be eligible for the study, patients had to meet the American College of Rheumatology (formerly, the American Rheumatism Association) criteria for the classification of SSc (14), have diffuse cutaneous scleroderma (having only limited cutaneous scleroderma was an exclusion criterion), and have a duration of SSc $\leq$18 months from the onset of the first manifestation other than Raynaud's phenomenon (15). All patients signed a consent form approved by the Institutional Review Board at each center. The details and results of the drug trial have been published elsewhere (16).

Patients were excluded for the following reasons: skin thickening distal to, but not proximal to, the elbows and knees (limited cutaneous scleroderma), age $<$18 years or $>$75 years, pregnancy, presence of another well-defined rheumatic disease or of only a localized form of scleroderma, severe organ involvement (defined previously [16]), another chronic debilitating illness (e.g., cancer), history of chronic blood dyscrasia, occurrence of scleroderma renal crisis (SRC) within the preceding 2 months, or intractable congestive heart failure. If patients required treatment with corticosteroids, the dose had to be stable at $\leq$10 mg prednisone (or equivalent) per day for at least 1 month prior to entry. Treatment with D-penicillamine, azathioprine, cyclophosphamide, methotrexate, chlorambucil, potassium aminobenzoate, colchicine, or captopril had to have been discontinued for $>$1 month prior to study entry.

Diagnostic tests. At baseline and every 6 months thereafter for up to 2 years, skin thickness was quantified using the modified Rodnan skin thickness scoring technique (17), in which skin thickness was assessed clinically in each of 17 body surface areas on a 0–3 scale: 0 = normal, 1 = mild thickness, 2 = moderate thickness, 3 = severe thickness (maximum score of 51). The reliability characteristics of this measure have been addressed previously (17). Active handspread was measured from the outermost aspect of the fifth digit to the outermost aspect of the first digit during maximal handspread. Fist closure was recorded as the distance from the distal end of the ring finger to the distal palmar crease during maximal fist closure (full closure was scored as 0). The vertical interlabial distance was measured with the patient maximally opening the mouth (oral aperture).

Organ system function was assessed at baseline and yearly thereafter for up to 2 years, as follows: pulmonary function tests (diffusing capacity for carbon monoxide [DLCO] and forced vital capacity [FVC], recorded in ml as well as % predicted), a posteroanterior chest radiograph (examined by a radiologist for cardiomegaly and the presence and degree of interstitial fibrosis; the cardiothoracic ratio was measured by investigators), 24-hour urinalysis for creatinine clearance and for protein content, measurement of serum creatine kinase (CK; recorded as the percentage of the upper limit of normal),

manual muscle testing, determination of joint tenderness and swelling not considered by the investigator to be due to another etiologic process (each assessed on a 0–1 scale in the elbows, wrists, metacarpophalangeal joints, and knees [8 joints total], with the number of tender or swollen joints summed to give a joint tenderness count and joint swelling count), presence and number of contractures of the elbows, wrists, and knees bilaterally, presence of palpable tendon friction rubs of the elbows, wrists, hands, knees, ankles, and other areas, measurement of serum creatinine, and routine urinalysis. Patients also self-assessed their functional capacity by completing the 20-item disability index of the Health Assessment Questionnaire (HAQ-DI). In this instrument (2–4), the 20 items (each assessed on a 0–3 scale) were divided into 8 domains; the highest scores in each of the 8 domains were summed and divided by 8 to derive the HAQ-DI (score range of 0–3).

Assessment of organ system involvement. Lung involvement was considered present if the DLCO was ≤70% predicted, FVC was ≤75% predicted, or definite interstitial changes were noted on the chest radiograph. Renal involvement was defined by a serum creatinine level that was higher than the upper limit of normal, a 24-hour endogenous creatinine clearance of <70 ml/minute (corrected to 1.73 mm^2 body surface area), or the occurrence of SRC. SRC was determined to be present by the patient's physician-examiner, whose assessment was based largely on the recognition of acute renal failure (serum creatinine level ≥2.0 mg/dl) and/or malignant hypertension occurring with blood pressure ≥160/110 mm Hg on at least 2 occasions at least 12 hours apart, accompanied by either persistent abnormalities in the urine (proteinuria, hematuria [unrelated to menses], or casts) or evidence of microangiopathic hemolytic anemia. Muscle involvement was defined by a CK level that was ≥200% of the upper limit of normal or by muscle strength that was ≤4 of 5 (on a 5-point Likert scale) in the proximal muscles (shoulder girdle and hip girdle). Joint involvement was considered present if the joint tenderness count was ≥1 tender joint. Heart involvement was defined by a history of, or presence of, congestive heart failure, cardiac arrhythmia requiring medication, pericarditis, moderate-to-large pericardial effusion, or cardiomegaly, or by a cardiothoracic ratio of >0.5 on the chest radiograph.

Change scores. During the 2-year followup of the 68 patients who completed the trial, many of the disease parameters changed (improved or worsened). Change scores were calculated for these parameters by determining the difference between the parameter value at 2 years and the value at baseline. Linear regression was used in the analysis of change scores that were continuous or ordinal and were normally distributed, while logistic regression was used in the analysis of binary outcome variables. For calculating logistic regression, change scores were dichotomized as "improved" or "not improved" ("no change" was always considered "not improved") based on whether the change was scored above or below 0 (no change).

Outcomes. Five outcomes of interest were studied. *Change in skin score*, a continuous outcome variable, was calculated as the difference in the value of the skin score at 2 years compared with the value at baseline. Four dichotomous outcomes were also examined. The occurrence of *mortality* and *SRC* were assessed for a mean ± SD of 4.0 ± 1.1 years after study entry. Response in skin thickening (*skin response*) was defined as a ≥25% decrease in skin score at 2 years compared with the value at entry. Finally, at 2 years, the physician-examiner made a *global assessment* of each SSc patient's condition as being improved, worsened, or unchanged compared with the status at entry (using a 7-point Likert scale that ranged from 3 grades of improvement [mild, moderate, marked] to no change to 3 grades of worsening [mild, moderate, marked]). Because this ordinal outcome was not normally distributed (the data were heavily skewed toward improvement), the global assessment of the patient's course over 2 years was also dichotomized as "improved" if the investigator assessed the patient's course as any grade of improvement, and "not improved" if the patient's course was assessed as "no change" or as any grade of worsening.

For prediction of 2 of the outcomes, mortality and SRC, baseline and followup data were available on 133 of the patients. Change score data on the 68 patients who completed 2 years of study medication were available for predicting the correlations between explanatory variables and the change in skin score, skin response, and global assessment; followup data were not routinely collected on the patients who withdrew before completing 2 years, and therefore analysis of these 3 outcomes could not be performed in the group of 66 patients who withdrew. There were too few deaths (n = 1) or episodes of SRC (n = 3) to allow analysis of these events in the subset who completed 2 years.

Statistical analysis. The data were entered and maintained in Medlog software (Information Analysis Corporation, Incline Village, NV), but all of the statistical analyses were performed using Stata (College Station, TX) and SAS (Chicago, IL) software. All continuous data are displayed as the mean ± SD, unless otherwise specified. Group means were compared using Student's 2-tailed, unpaired *t*-test. Clinical subsets were developed based on dichotomous variables (divisions occurring naturally or when continuous variables were made dichotomous by division above or below a cutoff point, usually the median). Since the data were frequently not normally distributed, Spearman's (rather than Pearson's) correlation coefficients were calculated. Spearman's correlation coefficients were calculated for mortality, SRC, skin response, and global improvement as binary outcome variables, and for change in skin score as a continuous outcome variable, while other variables served as independent or explanatory variables (continuous where available, dichotomous otherwise).

Logistic regression was used in the analysis in situations where the outcome was binary or, if continuous or ordinal, not normally distributed. Crude odds ratios (OR) and 95% confidence intervals (95% CI) were estimated using logistic regression, with the outcome variable and the explanatory variables as dichotomous variables. Since there was an effect on the outcome variables by many of the exposure variables in univariate logistic regression, only variables with *P* values less than 0.05 by chi-square testing were selected for multivariate logistic regression. Separate logistic models were built for each outcome variable (i.e., mortality, SRC, skin response, and global improvement). Independent variables of interest were then entered as dichotomous explanatory variables into each backward, stepwise multivariate logistic model. Adjusted OR and 95% CI were used to model associations of each predictor with each outcome. The R^2 value was the

measure of how well the logistic regression model explained the outcome (18). The effect of explanatory variables on mortality and SRC was also analyzed by life-table (Kaplan-Meier) and Cox proportional hazards methods.

In the one situation where the outcome was continuous and normally distributed (change in skin score), linear regression was used. The effect of explanatory variables on change in skin score was analyzed by univariate and backward, stepwise multivariate linear regression (starting with the variables having P values less than or equal to 0.05 on univariate regression). The R^2 value was the measure of the proportion of variance in each outcome that was explained by the regression model.

RESULTS

Patient characteristics at baseline (n = 134). The details of our randomized, controlled trial of high-dose (1,000 mg daily) versus low-dose (125 mg every other day) penicillamine for the diffuse cutaneous scleroderma variant of SSc have been published elsewhere (16). In summary, the courses of the 3 major outcome variables in the drug trial were similar in the 2 dosage groups: the skin thickness score improved to a similar degree in both groups over the 2 years of study, and the number of patients who died or experienced renal crisis was also comparable in the 2 groups during a mean ± SD followup of 4.0 ± 1.1 years from entry. The data on all 134 patients who entered the trial were pooled for this analysis. In the analyses described below, only one outcome, physician-examiner's global assessment, was statistically related to patient allotment to the high or low dose of penicillamine in univariate analyses. In multivariate regression, the global assessment variable was not a significant contributor to the outcomes. During the followup period (verified in 133 patients; 1 patient was lost to followup), 18 (14%) developed renal crisis and 20 (15%) died.

The demographic and baseline disease-related characteristics of the 134 patients (of whom 104 were female) are shown in Tables 1 and 2. The mean ± SD age was 43.7 ± 12.4 years. The mean modified Rodnan skin thickness score was 21.0 ± 8.0, the mean duration of SSc prior to entry was 10 ± 4 months, and the mean HAQ-DI score was 1.04 ± 0.67. Tendon friction rubs were noted in 37% of the patients, large joint contractures in 56%, joint involvement in 38%, muscle involvement in 16%, renal involvement in 25% (no patient had experienced SRC prior to entry), heart involvement in 20%, and lung involvement in 54%. Thirty percent of the patients were taking prednisone, with the average dose being 7.5 ± 2.5 mg per day.

Table 1. Continuous parameters/variables at baseline in all 134 patients with systemic sclerosis*

Variable	Mean ± SD	Median
Demographic		
Age, years	43.7 ± 12.4	43.5
History		
Duration of Raynaud's phenomenon at entry, days	582 ± 1,209	289
Physical examination		
Maximum oral aperture, mm	45.3 ± 10.0	44.0
Handspread, right, mm	174.6 ± 28.5	175.0
Fist closure, right, mm	25.4 ± 20.8	25.0
Skin score (0–51)	21.0 ± 8.0	19.0
Tender joint count (0–8)	1.5 ± 2.4	0
Swollen joint count (0–8)	0.9 ± 1.6	0
Laboratory		
Hematocrit, male, %	41.4 ± 3.1	41.6
Hematocrit, female, %	38.3 ± 3.7	38.6
White blood cell count, ×1,000/mm^3	8.4 ± 2.3	8.2
Platelet count, ×1,000/mm^3	340 ± 103	330
Erythrocyte sedimentation rate, mm/hour	23.9 ± 17.1	20.0
Creatine kinase, % of upper limit of normal	81.7 ± 124.1	46.1
DLCO, % predicted	75.2 ± 18.3	76.0
Forced vital capacity, % predicted	83.5 ± 16.9	85.0
Creatinine clearance, ml/minute	93.2 ± 31.1	89.6
Serum creatinine, mg/dl	0.89 ± 0.19	0.90

* DLCO = diffusing capacity for carbon monoxide.

In patients with skin scores ≥20, the odds (OR) of having heart involvement at baseline, compared with subjects with skin scores <20, was 3.10 (95% CI 1.25–7.70, P = 0.015). The risk of having kidney or lung involvement at baseline, however, was similar in both skin score groups.

Table 2. Dichotomous parameters/variables at baseline in all 134 patients with systemic sclerosis

Variable	No. of patients	% of patients
Sex		
Female	104	78
Male	30	22
D-penicillamine dose		
High	66	49
Low	68	51
Prednisone use	40	30
Digital tip ulcer	15	11
Nondigital tip ulcer	21	16
Proximal muscle weakness	14	10
Muscle involvement	22	16
Tendon friction rub	49	37
Large joint contractures	75	56
Swollen joint count	40	30
Joint involvement	51	38
Renal involvement	34	25
Heart involvement	27	20
Lung involvement	72	54

Table 3. Baseline variables that correlated with or predicted mortality (n = 133)*

Variable	Correlations		Logistic regression (univariate)				Kaplan-Meier	
	Coefficient†	*P*	OR	95% CI	*P*	Group at risk	Log-rank *P*	Wilcoxon *P*
Heart involvement	0.2559	0.0032	4.23	1.532–11.668	0.005	Heart involvement	0.0007	0.0009
Lung involvement	0.2198	0.0116	4.07	1.280–12.954	0.017	Lung involvement	0.0080	0.001
Skin score	0.2135	0.0140	3.59	1.227–10.548	0.020	Skin score ≥20	0.0181	0.0335
Large joint contractures	0.1483	0.0897	3.59	1.130–11.425	0.030	Presence of large joint contractures	0.0159	0.0105
FVC	−0.2329	0.0077	3.46	1.178–10.173	0.024	FVC <85% predicted	0.0203	0.0021
Hematocrit	−0.2623	0.0024	3.34	1.136–9.814	0.028	Male <41.6%; female <38.6%	0.0184	0.0018
ESR	0.2671	0.0033	3.33	1.121–9.870	0.030	ESR ≥20 mm/hour	0.0176	0.0018
HAQ-DI	0.2664	0.0020	3.22	1.097–9.468	0.033	HAQ-DI ≥1.0	0.0338	0.0110
Oral aperture	−0.2509	0.0040	3.18	1.071–9.421	0.037	Oral aperture ≤45 mm	0.0283	0.0121
DLCO	−0.2333	0.0080	3.00	1.075–8.375	0.036	DLCO <76% predicted	0.0454	0.0387
Platelet count	0.1190	0.1760	1.09	0.412–2.891	0.176	Platelets ≥330,000/mm^3	0.0188	0.0125

* OR = odds ratio; 95% CI = 95% confidence interval; FVC = forced vital capacity; ESR = erythrocyte sedimentation rate; HAQ-DI = Health Assessment Questionnaire disability index; DLCO = diffusing capacity for carbon monoxide.
† Positive coefficients indicate that higher values of each variable at baseline predict an increased risk of mortality. Negative coefficients indicate that lower values of each variable at baseline predict an increased risk of mortality.

Baseline predictors of outcomes (n = 134). *Mortality.* During 4.0 ± 1.1 years of followup, 20 (15%) of the 133 patients in whom outcome was verified died. Using Spearman's correlations, univariate logistic regression, and life-table methods, 11 potential predictors of mortality were found at baseline (Table 3). Since FVC and DLCO were included in the composite lung involvement parameter and lung involvement had a stronger relationship with skin response, only the lung involvement parameter was explored further in multivariate logistic regression. When stepwise multivariate logistic regression was performed with the remaining 9 potential predictors, lung involvement (multivariate OR 6.09, 95% CI 1.61–23.07) and skin score ≥20 (OR 3.69, 95% CI 1.17–11.67) were the only 2 variables that contributed significantly to predicting mortality (combined R^2 = 0.1390). The addition of other explanatory variables did not appreciably increase the R^2 value.

SRC. At baseline, none of the patients had ever had an episode of SRC. During 4.0 ± 1.1 years of followup, 18 (14%) of the 133 patients in whom the outcome was verified developed SRC. Using Spearman's correlations, univariate logistic regression, and life-table methods, 6 potential predictors of SRC were found at baseline (Table 4). Stepwise multivariate logistic regression suggested that of the 6, large joint contractures (multivariate OR 9.41, 95% CI 1.16–76.07) and skin score ≥20 (OR 6.64, 95% CI 1.40–-31.46) were the only 2 variables that contributed significantly to predicting SRC (combined R^2 = 0.1603). The addition of other explanatory variables did not appreciably increase the R^2 value.

Patient characteristics over 2 years (n = 68). Over the course of the study, 66 patients withdrew before finishing the full 2 years. Two-year data were complete on only 68 patients. Improvement was noted in

Table 4. Baseline variables that correlated with or predicted SRC (n = 133)*

Variable	Correlations		Logistic regression (univariate)				Kaplan-Meier	
	Coefficient†	*P*	OR	95% CI	*P*	Group at risk	Log-rank *P*	Wilcoxon *P*
Large joint contractures	0.2626	0.0024	16.12	2.075–125.254	0.008	Presence of large joint contractures	0.0006	0.0007
Skin score	0.2824	0.0011	10.00	2.213–45.907	0.003	Skin score ≥20	0.0004	0.0007
Prednisone use	0.2251	0.0098	3.63	1.304–10.051	0.014	Prednisone use	0.0201	0.0202
Heart involvement	0.1791	0.0415	2.93	1.010–8.482	0.048	Heart involvement	0.0104	0.0075
Hematocrit	−0.2041	0.0194	2.04	0.714–5.800	0.183	Male <41.6%; female <38.6%	0.1256	0.1255
Oral aperture	−0.1858	0.0350	1.60	0.578–4.429	0.366	Oral aperture <45 mm	0.3128	0.3720

* OR = odds ratio; 95% CI = 95% confidence interval.
† Positive coefficients indicate that higher values of each variable at baseline predict an increased risk of scleroderma renal crisis (SRC). Negative coefficients indicate that lower values of each variable at baseline predict an increased risk of SRC.

Table 5. Continuous parameters/variables in the 68 systemic sclerosis patients who completed the 2-year study, according to skin "response"*

Variable	Baseline			Change score at 2 years		
	Skin responders	Skin nonresponders	*P*	Skin responders	Skin nonresponders	*P*
Demographic						
Age, years	43.7 ± 13.5	45.1 ± 13.5	0.686	–	–	–
History						
Duration of Raynaud's phenomenon at entry, days	776 ± 1,363	922 ± 2,074	0.733	–	–	–
Physical examination						
Maximum oral aperture, mm	48.2 ± 12.0	48.3 ± 8.0	0.971	4.2 ± 8.1	−1.9 ± 7.3	0.003†
Handspread, right, mm	179.2 ± 30.3	183.0 ± 29.1	0.620	0.8 ± 13.7	−15.2 ± 19.0	0.002†
Fist closure, right, mm	28.4 ± 22.5	17.6 ± 17.0	0.049†	−9.0 ± 16.5	6.2 ± 19.5	0.002†
Skin score (0–51)	20.8 ± 8.8	19.1 ± 7.9	0.484	−11.3 ± 6.0	2.9 ± 6.9	0.000†
Tender joint count (0–8)	2.0 ± 2.5	0.9 ± 1.5	0.052	−0.9 ± 2.1	0.4 ± 2.6	0.031
Swollen joint count (0–8)	0.7 ± 1.4	0.8 ± 1.3	0.774	−0.6 ± 1.5	−0.1 ± 2.1	0.268
Tender joint count (0–8)	2.0 ± 2.5	0.9 ± 1.5	0.052	−0.9 ± 2.1	0.4 ± 2.6	0.031
Swollen joint count (0–8)	0.7 ± 1.4	0.8 ± 1.3	0.774	−0.6 ± 1.5	−0.1 ± 2.1	0.268
Laboratory						
Hematocrit, %	39.9 ± 3.2	39.7 ± 3.4	0.807	0.4 ± 3.0	−0.6 ± 2.6	0.174
White blood cell count, ×1,000/mm^3	8.1 ± 2.2	8.7 ± 2.5	0.314	−1.4 ± 1.8	−0.4 ± 2.4	0.060
Platelet count, ×1,000/mm^3	318 ± 89	362 ± 103	0.073	−46 ± 68	−81 ± 93	0.085
Erythrocyte sedimentation rate, mm/hour	22.2 ± 15.1	20.2 ± 12.4	0.581	−4.3 ± 8.4	−1.2 ± 17.0	0.881
Creatine kinase, % of upper limit of normal	58.7 ± 55.30	123.5 ± 241.0	0.106	−11.3 ± 58.0	−9.5 ± 71.4	0.912
DLCO, % predicted	75.6 ± 16.4	76.1 ± 19.9	0.913	−0.6 ± 18.4	−3.1 ± 21.0	0.616
Forced vital capacity, % predicted	86.7 ± 20.7	80.8 ± 14.3	0.216	6.7 ± 13.2	1.9 ± 13.3	0.160
Creatinine clearance, ml/minute	87.7 ± 27.1	98.4 ± 37.7	0.188	−1.0 ± 33.4	−16.1 ± 34.1	0.088
Serum creatinine, mg/dl	0.9 ± 0.2	0.9 ± 0.2	1.000	−0.0 ± 0.2	−0.0 ± 0.2	1.000

* Values are the mean ± SD. Skin responders comprised 42 patients whose skin score decreased by ≥25% of the baseline value. Skin nonresponders comprised 26 patients whose skin score did not decrease by ≥25% of the baseline value. *P* values were calculated by unpaired *t*-test comparing mean values (baseline or change score at 2 years) of skin responders versus skin nonresponders. DLCO = diffusing capacity for carbon monoxide.
† $P < 0.05$.

a number of clinical and laboratory variables in the 68 patients. The modified Rodnan skin thickness score decreased by 5.9 ± 9.4 units, the HAQ-DI by 0.15 ± 0.59 units, the tender joint count by 0.4 ± 2.4 units, and fist closure by 3.8 ± 18.9 mm, whereas the FVC (expressed as % predicted) increased by 4.8 ± 13.4% and handspread by 5.3 ± 17.6 mm. During the 2-year study, the new onset of heart and lung involvement was not associated with the baseline skin score (data not shown).

Of the 68 patients who completed the full 2-year study, the skin thickness score decreased by ≥25% of the baseline value in 42 patients, and these patients were considered "skin responders" (Tables 5 and 6). The other 26 patients did not have such a decrease and were considered "skin nonresponders." At baseline, skin responders more frequently had impaired fist closure ($P = 0.049$) and large joint contractures ($P = 0.034$), while skin nonresponders more often had heart involvement ($P = 0.025$) and fingertip ulcers ($P = 0.011$). Baseline characteristics were otherwise not different in the 2 groups. At 2 years, however, skin responders as a group showed significant improvements, compared with nonresponders, in oral aperture, handspread, fist closure, HAQ-DI, and, not unexpectedly, skin thickness score. Although the new onset of heart, lung, kidney, muscle, and joint involvement over these 2 years was not significantly different between the 2 groups (Table 6), skin nonresponders were more likely to develop new contractures ($P = 0.037$). These data, therefore, support the conclusion that improvement in skin thickening (decrease in skin score) was associated with improvement in oral aperture, hand function (handspread, fist closure), and overall functional capacity (HAQ-DI).

Correlates of outcome over 2 years. *Skin score.* Because the course of the skin score over 2 years was normally distributed in the 68 patients, relationships to explanatory baseline and change variables were examined by linear regression. The baseline and change variables that had significant ($P < 0.05$) relationships to changes in the skin score (by Spearman's correlation coefficients and univariate regression R^2) are shown in Table 7. When these relationships were analyzed by stepwise multivariate linear regression, the skin score at baseline along with 4 change variables together ex-

Table 6. Frequency of dichotomous variables at baseline and the frequency of new-onset involvement (not present at baseline) at 2 years in 42 skin responders and 26 skin nonresponders

Variable	Baseline*			New occurrence of involvement at 2 years in subjects at risk†		
	Skin responder (n = 42)	Skin nonresponder (n = 26)	P‡	Skin responder	Skin nonresponder	P‡
Female sex	30 (71)	21 (81)	0.391	–	–	–
Prednisone use	10 (24)	5 (19)	0.660	–	–	–
Digital tip ulcer	3 (7)	8 (31)	0.011§	3/39 (8)	4/16 (25)	0.088
Nondigital tip ulcer	8 (19)	6 (23)	0.692	2/34 (6)	4/20 (20)	0.115
Proximal muscle weakness	3 (7)	0 (0)	0.163	3/39 (8)	1/26 (4)	0.519
Muscle involvement	4 (10)	2 (8)	0.811	4/38 (11)	2/24 (8)	0.070
Tendon friction rub	15 (36)	6 (23)	0.273	2/27 (7)	4/20 (20)	0.081
Large joint contractures	24 (57)	8 (31)	0.034§	4/18 (22)	10/18 (56)	0.037§
Swollen joint count	10 (24)	9 (35)	0.274	2/32 (6)	2/17 (12)	0.463
Joint involvement	20 (48)	8 (31)	0.173	4/22 (18)	3/18 (17)	0.934
Renal involvement	11 (26)	6 (23)	0.775	3/31 (10)	6/20 (30)	0.070
Heart involvement	3 (7)	7 (27)	0.025§	4/39 (10)	4/19 (21)	0.168
Lung involvement	19 (45)	16 (62)	0.195	4/23 (17)	3/10 (30)	0.400

* Values are the number (%) of patients.
† Values are the number (%) of new-onset cases/number of subjects at risk (i.e., did not have that involvement at baseline).
‡ *P* values were calculated by chi-square test comparing skin responder and skin nonresponder groups (baseline or change score at 2 years).
§ $P < 0.05$.

plained 57% of the variance of the change in skin score ($R^2 = 0.5671$). These included baseline skin score ($P = 0.0078$) and changes in large joint contractures ($P = 0.0072$), tender joint counts ($P = 0.0119$), handspread ($P = 0.0242$), and HAQ-DI ($P = 0.0244$). The addition of other explanatory variables did not appreciably increase the R^2 value.

Skin response. Of the 68 patients who completed the 2-year trial, 42 had a "response" in skin score (≥25% decrease). Using Spearman's correlations, only 3 potential baseline variables were predictive of a response in skin thickness: absence of heart involvement and of fingertip ulcers, and a platelet count ≥330,000/mm^3 (data not shown). Using univariate and multivariate logistic regression, the absence of heart involvement and of fingertip ulcers at baseline were, together, predictors of improvement in skin score (multivariate OR 7.26, 95% CI 1.65–32.07 for absence of fingertip ulcers and OR 6.23, 95% CI 1.37–28.31 for absence of heart involvement; $R^2 = 0.1701$).

Table 7. Baseline and 2-year change variables that correlated with change in skin score over the study period (n = 68)*

Variable	Correlation		Linear regression (univariate)		Direction of change associated with skin response in group at risk
	Coefficient†	P	R^2	P	
Change variable					
Handspread, right	−0.4672	0.0001	0.2469	0.000	Increasing handspread
HAQ-DI	0.4918	0.0000	0.2458	0.000	Decreasing HAQ-DI
Large joint contractures	0.4667	0.0001	0.1884	0.000	Fewer contractures
Oral aperture	−0.3480	0.0042	0.1686	0.001	Increasing oral aperture
Tender joint count	0.3370	0.0049	0.1663	0.001	Fewer tender joints
Swollen joint count	0.1998	0.1024	0.1460	0.001	Fewer swollen joints
Fist closure	0.4326	0.0004	0.1230	0.004	Tighter fist
Hematocrit	–	–	0.0645	0.040	Increasing hematocrit
WBC count	0.3614	0.0029	0.0786	0.023	Decreasing WBC
Baseline variable					
Skin score	−0.3777	0.0015	0.1354	0.002	Higher skin scores
Tender joint count	−0.2584	0.0334	0.0633	0.038	Tender joints

* HAQ-DI = Health Assessment Questionnaire disability index; WBC = white blood cell.
† Positive correlation indicates that higher values show stronger relationship to change in skin score. Negative correlation indicates that lower values show stronger relationship to change in skin score.

Table 8. Change in explanatory variables over 2 years that correlated with or were associated with physician's global assessment as "improved" over the same 2-year period (n = 50)*

Variable	Correlations		Logistic regression (univariate)			Direction of change in group at risk
	Coefficient†	*P*	OR	95% CI	*P*	
Change variable						
Skin score	−0.6852	0.000	24.62	5.475–110.711	0.000	Improving skin score
HAQ-DI	−0.5814	0.000	22.15	4.469–109.463	0.000	Improving HAQ-DI
WBC count	−0.4034	0.001	6.16	1.617–23.463	0.008	Decreasing WBC
Fist closure, right	−0.4337	0.000	4.20	1.181–14.937	0.027	Tighter fist
Oral aperture	0.4365	0.000	3.89	1.212–12.460	0.022	Wider aperture
Large joint contractures	−0.4600	0.000	5.62	0.677–46.685	0.110	Fewer contractures
Handspread, right	0.3588	0.003	3.22	0.817–12.677	0.095	Wider handspread
Skin ulcers	−0.3021	0.008	2.60	0.297–22.868	0.388	Fewer ulcers
FVC	0.2636	0.035	1.90	0.611–5.887	0.268	Increasing FVC
Tender joint count	−0.2637	0.030	0.92	0.249–3.381	0.896	Fewer tender joints
Baseline variable						
Penicillamine dose	0.3025	0.012	4.24	1.305–13.741	0.016	Lower dose

* See Tables 3 and 7 for definitions.
† Positive correlation indicates that higher values predict a higher likelihood of physician's global assessment as "improved." Negative correlation indicates that lower values predict a higher likelihood of physician's global assessment as "improved."

Global improvement. Of the 68 patients who completed the 2-year trial, 50 were judged by their center physicians to have "improved" globally. Using Spearman's correlations and univariate logistic regression, only 2 baseline variables were found to predict global improvement: higher hematocrit levels (≥41.6% in males and ≥38.6% in females) and the presence of fewer contractures. Higher hematocrit (multivariate OR 4.11, 95% CI 1.13–14.95) and the presence of fewer contractures at baseline (OR 4.46, 95% CI 1.23–16.14) were the only 2 variables that, together, best explained global improvement ($R^2 = 0.1565$). The addition of other explanatory variables did not appreciably increase the R^2 value.

In contrast, 11 change variables correlated with global improvement (Table 8). Two of these change variables (change in skin score and change in HAQ-DI) had exceptionally strong associations with global improvement and were analyzed together in multivariate logistic regression. Change in skin score (multivariate OR 18.91, 95% CI 3.06–116.98; $P = 0.002$) and change in HAQ-DI (OR 17.96, 95% CI 2.89–111.55; $P = 0.002$) were the 2 variables that, together, best explained the model ($R^2 = 0.455$). No other explanatory variables were significant at the $P < 0.05$ level when added to the model.

DISCUSSION

Although skin thickness is the most obvious and frequent characteristic of SSc, some investigators have downplayed its importance in the course of SSc. "Patients do not die from thick skin" is the refrain. In the last decade, however, it has become clear that the degree and extent of skin thickening definitely are important clues to risks for future morbidity and mortality in SSc.

The data derived from the present study cohort of 134 patients with early diffuse SSc support and add to the findings of previous reports (2,4–8) in which high skin scores have been linked to increased risks for morbidity and mortality in SSc. Our data clearly demonstrate that higher skin scores (in this case, a modified Rodnan skin score ≥20), together with lung involvement, are the most important identifiable risk factors for early mortality (multivariate OR 3.69 for skin score and 6.09 for lung involvement). This finding confirms the conclusions of previous investigators who have demonstrated that a high skin score is one of the major determinants predicting early mortality (6,7).

Our study also confirms the reports by Steen et al and other investigators, that high skin scores at first visit are an important risk factor for renal crisis (6,8). Our data show that although several variables are predictive of SRC, a skin score ≥20 (multivariate OR 6.64) and large joint contractures (OR 9.41) together are the best predictors. Our data also demonstrate that patients with higher skin scores (≥20) at baseline are more likely to have heart (but not lung) involvement at baseline. Our recently reported analysis of data from the same cohort of 134 SSc patients also showed that functional impairment in SSc at entry, as measured by the HAQ-DI, was greater in SSc patients with higher skin scores (4). These observations, therefore, link higher skin scores with

higher rates of mortality, SRC, heart involvement, and disability, and suggest that although thick skin may not be fatal per se, it clearly serves as a marker for patients who are destined for poor outcomes.

The same dissenting investigators have also remarked that skin softening should not be a major goal of therapy, again because "improving the skin is not improving visceral or functional outcomes," which are the "real" surrogates of morbidity and mortality. New data (ours and others) suggest that this view should be tempered. For example, Steen recently reported that the risk of 5- and 10-year mortality was significantly less in SSc patients with diffuse cutaneous scleroderma (first seen within 3 years of SSc onset) whose skin thickness scores decreased over a 2-year period compared with patients whose skin scores did not improve (12). Because we had only 1 death among the patients followed up for 2 years, we were not able to address this issue.

The present study explores the ramifications of skin softening further by evaluating whether skin softening is associated with improvement in physical, functional, and/or other outcomes over a 2-year study period. Our analyses consistently demonstrate that significant improvement in skin score is associated with improvement in hand function (as measured by improved handspread and fist closure), joint contractures (improved oral aperture and large joint contractures), inflammatory indices (lowered white blood cell count and increased hematocrit), arthritis signs (decreased joint tenderness and swelling counts), and overall functional capacity (improved HAQ-DI). In this analysis, regression showed that the baseline skin score coupled with changes in large joint contractures, tender joints, handspread, and HAQ-DI explained 57% of the changes in skin score over 2 years, a very good effect for a biologic model.

Our data cannot directly address the issue of whether skin improvement is causally related to improvement in physical and functional outcomes or to decreased morbidity and mortality. In the recently reported placebo-controlled trial of the antifibrotic therapy recombinant human relaxin, it was noted that softening of skin was also paralleled by improved patient function and finger motion (13). Together, these observations support the conclusion that softening of skin (naturally or by relaxin) is associated with (even if not causative of) improvements in other physical and functional outcomes.

These data (ours and others) suggest that changes in skin thickness reflect the evolution of the underlying biology of the fibrotic process in SSc over time, especially in patients with diffuse cutaneous scleroderma. Early in the disease, skin thickens as excess collagen and matrix are deposited. In a similar manner, the lungs, heart, gastrointestinal tract, tendons, periarticular structures, and other organs are targets of excess deposition of collagen and of matrix deposition. Later in the disease course, as the balance between deposition and resorption shifts in favor of collagen resorption, skin thins (or softens), the joints become more flexible, energy increases, and function improves. Softening of skin may therefore be a sign that the tempo of the disease is abating overall, that excess collagen is being removed, and that the body is beginning to heal itself.

In our study, only about half of the patients who entered actually completed the 2-year study. They formed the cohort used in the analysis over time. Unfortunately, our data cannot address the course of scleroderma in the patients who left the study and were not further evaluated. Our data do, however, clearly show that the condition of many of the patients (about two-thirds of the group who completed the study) improved over time. The improvements in skin thickening, hand function, inflammatory indices, arthritis signs, and overall functional ability were undoubtedly a reflection of the biologic phase of fibrosis, which, at this point, was in a resorptive phase.

Although patients were not queried about whether they had improved during the course of the study, we did query the investigators (physician's global assessment) about the status of the patients they cared for during the 2-year study. Fifty of the 68 SSc patients were judged by investigators to have improved. In this study, 2 major factors that contributed most to the investigator's judgment about whether or not patients had improved were change in skin score and change in the HAQ-DI. These 2 factors accounted for ~45% of the investigator's judgment about patient improvement.

These observations demonstrate that improvement in skin score in SSc patients with diffuse cutaneous scleroderma is associated with improvement in hand function, inflammatory indices, contractures, arthritis signs, overall functional ability, and investigator's global assessment of improvement. The improvements in physical, functional, and other outcomes detected in this and other studies support the argument that improvements in skin thickness are not limited to improvements in skin alone; they are clearly associated with improved functional outcomes and with a decrease in the risks of morbidity and mortality. The implications for patients and physicians alike is that the natural history for many

patients with early diffuse cutaneous scleroderma may involve a period of improvement after the initial few years of worsening.

REFERENCES

1. LeRoy EC, Black C, Fleischmajer R, Jablonska S, Krieg T, Medsger TA Jr, et al. Scleroderma (systemic sclerosis): classification, subsets, and pathogenesis. J Rheumatol 1988;15:202–5.
2. Poole JL, Steen VD. The use of the Health Assessment Questionnaire (HAQ) to determine physical disability in systemic sclerosis. Arthritis Care Res 1991;4:27–31.
3. Steen VD, Medsger TA Jr. The value of the Health Assessment Questionnaire and special patient-generated scales to demonstrate change in systemic sclerosis patients over time. Arthritis Rheum 1997;40:1984–91.
4. Clements PJ, Wong WK, Hurwitz EL, Furst DE, Mayes M, White B, et al. Correlates of the disability index of the Health Assessment Questionnaire: a measure of functional impairment in systemic sclerosis. Arthritis Rheum 1999;42:2372–80.
5. Poole JL. Grasp pattern variations seen in the scleroderma hand. Am J Occup Ther 1994;48:46–54.
6. 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.
7. Bryan C, Knight C, Black CM, Silman AJ. Prediction of five-year survival following presentation with scleroderma: development of a simple model using three disease factors at first visit. Arthritis Rheum 1999;42:2660–5.
8. Steen VD, Medsger TA Jr, Osial TA, Ziegler GL, Shapiro AP, Rodnan GP. Factors predicting development of renal involvement in progressive systemic sclerosis. Am J Med 1984;76:779–86.
9. Clements PJ, Medsger TA Jr. Organ involvement: skin. In: Clements PJ, Furst DE, editors. Systemic sclerosis. Baltimore: Williams & Wilkins; 1996. p. 389–407.
10. Medsger TA Jr. Systemic sclerosis (scleroderma): clinical aspects. In: Koopman WJ, editor. Arthritis and allied conditions: a textbook of rheumatology. Baltimore: Williams & Wilkins; 1997. p. 1433–64.
11. Seibold JR. Scleroderma. In: Kelley WN, Harris ED, Ruddy S, Sledge CB, editors. Textbook of rheumatology. Philadelphia: WB Saunders; 1997. p. 1133–62.
12. Steen V. Does improvement in skin thickening in diffuse scleroderma lead to improved survival? [abstract]. Arthritis Rheum 1998;41 Suppl 9:S240.
13. Seibold JR, Korn JH, Simms R, Clements PJ, Moreland LW, Mayes MD, et al. Recombinant human relaxin in the treatment of scleroderma: a randomized double-blind, placebo-controlled trial. Ann Intern Med 2000;132:871–9.
14. Subcommittee for Scleroderma Criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. Preliminary criteria for the classification of systemic sclerosis (scleroderma). Arthritis Rheum 1980;23:581–90.
15. Medsger TA Jr, Steen VD. Classification, prognosis. In: Clements PJ, Furst DE, editors. Systemic sclerosis. Baltimore: Williams and Wilkins; 1996. p. 51–64.
16. Clements PJ, Furst DE, Wong W-K, Mayes M, White B, Wigley F, et al. High-dose versus low-dose D-penicillamine in early diffuse systemic sclerosis: analysis of a two-year, double-blind, randomized, controlled clinical trial. Arthritis Rheum 1999;42:1194–203.
17. Clements P, Lachenbruch P, Seibold JR, White B, Weiner S, Martin RW, et al. Inter and intraobserver variability of total skin thickness score (modified Rodnan TSS) in systemic sclerosis. J Rheumatol 1995;22:1281–5.
18. Nagelkerke NJD. A note on a generalized definition of the coefficient of determination. Biometrika 1991;78:691–2.