

The Disability Index of the Health Assessment Questionnaire Is a Predictor and Correlate of Outcome in the High-Dose Versus Low-Dose Penicillamine in Systemic Sclerosis Trial

Philip J. Clements,¹ Weng Kee Wong,² Eric L. Hurwitz,² Daniel E. Furst,³ Maureen Mayes,⁴ Barbara White,⁵ Fredrick Wigley,⁶ Michael Weisman,⁷ Walter Barr,⁸ Larry Moreland,⁹ Thomas A. Medsger, Jr.,¹⁰ Virginia Steen,¹¹ Richard W. Martin,¹² David Collier,¹³ Arthur Weinstein,¹⁴ Edward Lally,¹⁵ John Varga,¹⁶ Steven R. Weiner,¹ Brian Andrews,¹⁷ Micha Abeles,¹⁸ and James R. Seibold¹⁹

Objective. To explore the clinical implications of a score of ≥ 1.0 on the Disability Index of the Health Assessment Questionnaire (HAQ DI) at the first patient visit, and to examine the implications of improvement in HAQ DI score over 2 years in a cohort of systemic

sclerosis (SSc) patients with diffuse cutaneous scleroderma.

Methods. SSc skin and visceral involvement was assessed in 134 SSc patients with diffuse scleroderma (mean \pm SD disease duration of 10 ± 4 months) when they entered a multicenter drug trial and again 2 years later. Mortality and the occurrence of scleroderma renal crisis were assessed for a mean \pm SD of 4.0 ± 1.1 years. Logistic and linear regression analyses were used to examine the relationship of the baseline HAQ DI score to morbidity, mortality, and visceral involvement, as well as the relationship of changes in the HAQ DI score to changes in physical examination, laboratory, and functional variables over 2 years.

Results. A baseline HAQ DI score of ≥ 1.0 was predictive of mortality (odds ratio 3.22, 95% confidence interval 1.097–9.468) over 4 years. Multivariate linear regression demonstrated that a model which included the erythrocyte sedimentation rate at baseline ($P = 0.005$) and changes at 2 years in the swollen joint count ($P = 0.002$), total skin score ($P = 0.005$), and white blood cell count ($P = 0.005$) best explained the change in HAQ DI score over 2 years ($R^2 = 0.528$). The HAQ DI score and total skin score at baseline were highly correlated (correlation coefficient 0.368), as were changes in the HAQ DI score and the total skin score over 2 years (correlation coefficient 0.492). Although the HAQ DI score was heavily influenced by hand dysfunction at baseline and at 2 years, improvement (reduction) in the HAQ DI score over 2 years was related to factors other than hand dysfunction.

Conclusion. A baseline HAQ DI score of ≥ 1.0 predicted mortality over 4 years. Improvement in the

Supported in part by grants from the Scleroderma Federation, United Scleroderma Foundation, FDA Orphan Drug Program, and Arthritis Foundation, CRC grants M01-RR-00865 and M01-RR-00827, and bequests from the estates of Winifred Krause, Morris Goldsmith, and Takako Ito.

¹Philip J. Clements, MD, MPH, Steven R. Weiner, MD: UCLA School of Medicine, Los Angeles, California; ²Weng Kee Wong, PhD, Eric L. Hurwitz, PhD: UCLA School of Public Health, Los Angeles, California; ³Daniel E. Furst, MD: Virginia Mason Research Center, Seattle, Washington; ⁴Maureen Mayes, MD, MPH: Wayne State University, Hutzel Hospital, Detroit, Michigan; ⁵Barbara White, MD: University of Maryland School of Medicine, Baltimore; ⁶Fredrick Wigley, MD: The Johns Hopkins University, Baltimore, Maryland; ⁷Michael Weisman, MD: Cedars-Sinai Medical Center, Los Angeles, California; ⁸Walter Barr, MD: Northwestern University Medical School, Chicago, Illinois; ⁹Larry Moreland, MD: The University of Alabama at Birmingham; ¹⁰Thomas A. Medsger, Jr., MD: University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania; ¹¹Virginia Steen, MD: Georgetown University Medical Center, Washington, DC; ¹²Richard W. Martin, MD: Michigan State University College of Human Medicine, Grand Rapids; ¹³David Collier, MD: University of Colorado Health Sciences Center, Denver; ¹⁴Arthur Weinstein, MD: The George Washington University Medical Center, Washington, DC; ¹⁵Edward Lally, MD: Brown University School of Medicine, Roger Williams General Hospital, Providence, Rhode Island; ¹⁶John Varga, MD: University of Illinois, Chicago; ¹⁷Brian Andrews, MD, PhD: University of California, Irvine; ¹⁸Micha Abeles, MD: University of Connecticut Health Center, Farmington; ¹⁹James R. Seibold, MD: UMDNJ–Robert Wood Johnson Medical School, New Brunswick, New Jersey.

Address correspondence and reprint requests to Philip J. Clements, MD, MPH, UCLA Medicine–Rheumatology, 32-59 Rehabilitation Center, Box 951670, Los Angeles, CA 90095-1670.

Submitted for publication July 13, 2000; accepted in revised form November 9, 2000.

HAQ DI score in these patients with diffuse scleroderma was associated with improvement in skin thickening, hand function, oral aperture, lung function, signs of arthritis, serum creatinine level, and the investigator's global assessment of improvement. The HAQ DI is a self-administered questionnaire that SSc patients can complete easily and rapidly and that gives the practicing physician important information about prognosis, patient status, and changes in disease course over time.

Systemic sclerosis (SSc) is a disorder characterized by overproduction and deposition of collagen in the skin and visceral organs, abnormalities of the microcirculation, and autoimmunity (1,2). Although it is well known that the disorder can be disabling, the degree of functional impairment has only recently been quantified using the Disability Index of the Health Assessment Questionnaire (HAQ DI) (3–9). We recently reported that moderate-to-severe functional impairment (HAQ DI score of ≥ 1.0) was already present within 18 months of SSc onset in 53% of 134 patients with diffuse cutaneous scleroderma (diffuse scleroderma) (3). This observation corroborates the findings of Poole, Steen, and others (4–9) who have also shown that moderate-to-severe functional impairment is frequent in SSc, especially in patients with diffuse scleroderma.

Recently, Steen and Medsger reported on the course of the HAQ DI score over a mean of 4 years in SSc patients followed up at the University of Pittsburgh (5). Their results showed that patients with diffuse scleroderma who ultimately died not only had higher first-visit HAQ DI scores than those of surviving patients with diffuse scleroderma, but also had a significant increase (worsening) in mean HAQ DI score prior to their deaths. Their analysis also showed that changes in the skin score were highly correlated with changes in the HAQ DI score (correlation coefficient 0.683) over 4 years. Few other published studies have addressed the issue of how functional impairment changes over time in patients with SSc, how these changes correlate with physical and laboratory changes, and what these changes imply regarding future morbidity or mortality.

In this study, we tested whether baseline data (derived from 134 SSc patients with diffuse scleroderma as they entered a randomized controlled trial of high-dose versus low-dose penicillamine for SSc) predicted survival over a 4-year period, and whether the courses of the HAQ DI score and the skin score were correlated in patients with diffuse scleroderma over 2 years of observation. More importantly, our present analysis explored changes in functional impairment in the subset of 68 SSc patients who completed the 2-year trial, as well as how

these changes correlated with other physical and laboratory parameters.

What patients consider an “important symptomatic difference” in functional level has been quantified in 2 studies using the HAQ DI (10,11). In both studies, the average patient with rheumatoid arthritis (RA) concluded that if her or his HAQ DI score was lower by ≥ 0.2 units than another patient's HAQ DI score, this difference would be enough for the first patient to say that she or he was at least “somewhat better” than the other patient. In the present study, we tested whether an improvement of ≥ 0.2 units in the HAQ DI score in SSc patients with diffuse scleroderma correlated with changes in other physical and laboratory parameters over 2 years of study.

PATIENTS AND METHODS

Patients. To be eligible to enter the low-dose versus high-dose penicillamine in early SSc trial, patients had to meet American College of Rheumatology (formerly, the American Rheumatism Association) criteria for SSc, have diffuse cutaneous scleroderma (cutaneous induration proximal to, as well as distal to, the elbow and/or knee, with or without facial skin thickening), and have a disease duration of ≤ 18 months from the onset of the first SSc manifestation (other than Raynaud's phenomenon) (12,13). All patients signed a consent form approved by the Institutional Review Board at each center. The details and results of the high-dose versus low-dose penicillamine drug trial have been published elsewhere (14).

Patients were excluded if they had only limited cutaneous scleroderma (cutaneous induration distal to, but not proximal to, the knees and elbows) or only one of the localized forms of scleroderma, or if they met any of the following criteria or had any of the following conditions: age < 18 or > 75 years, pregnancy, presence of another well-defined rheumatic disease, severe organ involvement as defined previously (14), another chronic debilitating illness (e.g., cancer), history of a 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 ≤ 10 mg/day prednisone (or equivalent) for at least 1 month prior to entry. D-penicillamine, azathioprine, cyclophosphamide, methotrexate, chlorambucil, potassium aminobenzoate, colchicine, or captopril had to have been discontinued for > 1 month prior to 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 score (total skin score) technique (3,14–16). Hand spread, fist closure, and vertical interlabial distance (oral aperture) were measured as previously described (3,14,16).

Organ system function was assessed at baseline and yearly thereafter for up to 2 years, as previously described (3,14,16). Assessments included pulmonary function tests (diffusing capacity for carbon monoxide [DLCO; percent of predicted] and forced vital capacity [FVC; recorded in milliliters

as well as percent of predicted]), posteroanterior chest radiograph; tests of 24-hour urine for creatinine clearance and protein level; serum creatine kinase (CK) level (recorded as the percent of the upper limit of normal); manual muscle testing; joint tenderness and swelling (each assessed in 8 joints [elbows, wrists, metacarpophalangeal joints, and knees; scored on a 0–1 scale] as well as the number of tender or swollen joints summed to give joint tenderness and joint swelling counts); 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; serum creatinine level; and routine urinalysis.

Patients also performed self assessments of their functional capacity by completing the 20-item HAQ DI (3). In this instrument, 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 0–3).

Assessment of organ system involvement. Involvement of the lung, kidney, muscle, joint, and heart were defined as previously described (3,14,16).

Change scores. Since complete data were only available for patients who finished the 2-year trial, calculation of change scores was possible only for those 68 patients. Change scores for these variables were derived by calculating the difference between the baseline and the 2-year values. Linear regression was employed in the analysis of normally distributed continuous or ordinal values. 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 above or below zero (no change).

Outcomes. Five outcomes of interest were studied. One outcome, change in HAQ DI score, was continuous. The other 4 outcomes (mortality, SRC, improvement in HAQ DI score, and global improvement) were dichotomous. The occurrence of mortality and SRC were assessed for a mean \pm SD of 4.0 ± 1.1 years after entry. Improved HAQ DI was defined as a decrease in the HAQ DI score of ≥ 0.2 units at 24 months compared with the value at entry. At 2 years, the physician examiners made a global assessment of each SSc patient’s disease course as improved, worsened, or unchanged compared with disease status at entry (using a 7-point Likert scale), as previously described (16). Because this ordinal outcome was not normally distributed (the data were heavily skewed toward improvement), the global assessment of a patient’s disease course over 2 years was dichotomized as “improved” if the investigator assessed the patient’s course as any grade of improved, and “not improved” if the patient’s course was assessed as unchanged or as any grade of worsened.

Because 1 patient was lost to followup, data needed for the prediction of mortality and SRC were available for only 133 patients. Because change score data were available for only the 68 patients who completed (received medication for the duration of) the 2-year study, the correlations of change in HAQ DI, improved HAQ DI, and global assessment with explanatory variables were possible only for this subset of 68 patients.

Statistical analysis. All of the statistical analyses were performed using Stata (Stata Corporation, College Station, TX) and SAS (SAS Institute, Cary, NC) software. All continuous data are reported as the mean \pm SD, unless otherwise

specified. Group means were compared using Student’s *t*-test (2-tailed, unpaired). Clinical subsets were developed based on dichotomous variables (occurring naturally or when continuous variables were made dichotomous by division above or below a cut point, usually the median). Since the data were frequently not normally distributed, Spearman (rather than Pearson) correlation coefficients were calculated. Spearman correlation coefficients were calculated with mortality, SRC, improved HAQ DI, and global improvement as binary outcome variables, and with change in HAQ DI as a continuous outcome variable. Other variables served as independent or explanatory variables (continuous where available, dichotomous otherwise).

Crude odds ratios (ORs) and 95% confidence intervals (95% CIs) were estimated using univariate logistic regression, with the outcome variable and the explanatory variables as dichotomous variables. Only variables with *P* values of <0.05 by chi-square testing during univariate analysis were selected for further multivariate logistic regression. Separate logistic models were built for each outcome variable (i.e., mortality, SRC, improved HAQ DI, and global improvement). Independent variables of interest were entered as dichotomous explanatory variables into each backward stepwise multivariate logistic model. Adjusted ORs and 95% CIs were used to model associations of multiple potential predictors with each outcome. R^2 was the measure of how well the regression explained the outcome (17). The effect of explanatory variables on mortality and SRC was also analyzed by life-table (Kaplan-Meier method) and Cox proportional hazards methods.

For change in HAQ DI, linear regression was used. The effect of explanatory variables on change in HAQ DI was analyzed by univariate and backward stepwise multivariate linear regression (using variables which had *P* values of ≤ 0.05 on univariate regression). R^2 was the measure of the proportion of variance in each outcome that was explained by the regression model.

RESULTS

Patient characteristics at baseline (n = 133 patients). The details of the randomized controlled trial of high-dose (1,000 mg/day) versus low-dose (125 mg/every other day) penicillamine for the diffuse scleroderma variant of SSc have been published elsewhere (14). In summary, the courses of the 3 major outcomes in the drug trial were similar in the 2 dosage groups. Skin thickness scores improved to similar degrees in both groups over the 2 years of the study, while the number of patients who died or experienced renal crisis were similar in the 2 groups followed up for a mean \pm SD of 4.0 ± 1.1 years. Since 1 patient was lost to followup, the data for the remaining 133 patients were pooled for this analysis. During the followup period of 4.0 ± 1.1 years, 18 patients developed renal crisis and 20 patients died.

The demographic and baseline disease-related characteristics of the original 134 patients studied (104 women and 30 men) have been described in detail

Table 1. Baseline variables that correlated with or predicted mortality (n = 133 patients)*

Variable	Correlations		Univariate logistic regression				Kaplan-Meier	
	Coefficient†	P	OR	95% CI	P	Group at risk	Log rank P	Wilcoxon P
Heart involvement	0.256	0.003	4.23	1.532–11.668	0.005	Heart involved	0.001	0.001
Lung involvement	0.220	0.012	4.07	1.280–12.954	0.017	Lung involved	0.008	0.001
Total skin score	0.214	0.014	3.59	1.227–10.548	0.020	Total skin score ≥ 20	0.018	0.034
Joint contractures	0.148	0.0907	3.59	1.130–11.425	0.030	Contractures	0.016	0.011
FVC, % of predicted	–0.233	0.008	3.46	1.178–10.173	0.024	FVC $< 85\%$ of predicted	0.020	0.002
Hematocrit, %	–0.262	0.002	3.34	1.136–9.814	0.028	Men with hematocrit of $< 41.6\%$	0.018	0.002
						Women with hematocrit of $< 38.6\%$		
ESR, mm/hour	0.267	0.003	3.33	1.121–9.870	0.030	ESR ≥ 20 mm/hour	0.018	0.002
HAQ DI score	0.266	0.002	3.22	1.097–9.468	0.033	HAQ DI score ≥ 1.0	0.034	0.011
Oral aperture, mm	–0.251	0.004	3.18	1.071–9.421	0.037	Oral aperture ≤ 45 mm	0.028	0.012
DLco, % of predicted	–0.233	0.008	3.00	1.075–8.375	0.036	DLco $< 76\%$ of predicted	0.045	0.039
Platelets/mm ³	0.119	0.176	1.09	0.412–2.891	0.176	$\geq 330,000$ platelets/mm ³	0.019	0.013

* OR = odds ratio; 95% CI = 95% confidence interval; FVC = forced vital capacity; ESR = erythrocyte sedimentation rate; HAQ DI = Disability Index of the Health Assessment Questionnaire; DLco = diffusing capacity for carbon monoxide.

† For positive correlations, higher (more abnormal) values at baseline predict greater risk of mortality. For negative correlations, lower (more abnormal) values at baseline predict greater risk of mortality.

elsewhere (3,14). Briefly, their age was 43.7 ± 12.4 years (mean \pm SD), their modified Rodnan skin thickness score was 21.0 ± 8.0 , their duration of SSc prior to entry was 10 ± 4 months, and their HAQ DI score was 1.04 ± 0.67 . Tendon friction rubs were noted in 37% of 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 patients were taking prednisone at a dosage of 7.5 ± 2.5 mg/day.

Baseline predictors of outcome (n = 133 patients). *Mortality.* During 4.0 ± 1.1 years of followup, 20 patients died (15%) of the 133 for whom the outcome was known.

Table 2. Frequency of dichotomous variables at baseline, proportion and frequency of patients who improved, and proportion and frequency of patients who developed new-onset involvement over 2 years*

Variable	Baseline, no. (%) of patients			Patients abnormal at baseline who improved, proportion (%)			Patients at risk who had new onset of involvement, proportion (%)†		
	HAQ DI			HAQ DI			HAQ DI		
	HAQ DI improved (n = 33)	not improved (n = 35)	P‡	HAQ DI improved	not improved	P‡	HAQ DI improved	not improved	P‡
Female	23 (70)	28 (80)	0.340	–	–	–	–	–	–
Prednisone use	8 (24)	7 (20)	0.673	–	–	–	–	–	–
Digital tip ulcer	4 (12)	7 (20)	0.378	3/4 (75)	2/5 (40)	0.242	1/29 (3)	6/28 (21)	0.436
Non-digital tip ulcer	7 (21)	7 (20)	0.902	6/7 (86)	2/6 (33)	0.103	0/26 (0)	6/28 (21)	0.013§
Proximal muscle weakness	2 (6)	1 (3)	0.520	2/2 (100)	1/1 (100)	1.000	1/31 (3)	3/34 (9)	0.314
Tendon rub	12 (36)	9 (26)	0.342	5/6 (83)	4/6 (67)	1.000	3/21 (14)	3/26 (12)	0.839
Contractures	19 (58)	13 (37)	0.092	9/19 (47)	5/13 (38)	0.725	7/14 (50)	7/22 (32)	0.280
Joint involvement	14 (42)	14 (40)	0.839	8/13 (62)	7/14 (50)	0.704	2/19 (11)	5/21 (24)	0.283
Muscle involvement	3 (9)	3 (9)	0.938	1/1 (100)	2/3 (67)	1.000	1/30 (3)	5/32 (16)	0.084
Renal involvement	7 (21)	10 (29)	0.484	4/7 (57)	7/10 (70)	0.644	3/26 (12)	6/25 (24)	0.264
Heart involvement	8 (24)	2 (6)	0.031§	6/8 (75)	1/2 (50)	1.000	2/25 (8)	6/33 (18)	0.273
Lung involvement	17 (52)	18 (51)	0.994	7/17 (41)	10/18 (56)	0.505	1/16 (6)	6/17 (35)	0.041§
Swollen joints	11 (33)	8 (23)	0.336	10/11 (91)	6/8 (75)	0.546	0/22 (0)	4/27 (15)	0.058

* HAQ DI = Disability Index of the Health Assessment Questionnaire.

† Proportions are the number of patients with new onset of involvement/number of patients at risk (i.e., without that involvement at baseline).

‡ By Fisher's exact test.

§ $P < 0.05$, by chi-square test.

Table 3. Baseline and change score variables (continuous) in the 68 systemic sclerosis patients who completed the 2-year study, classified according to improvement in HAQ DI score*

Variable	Baseline			Change scores at 2 years		
	HAQ DI improved (n = 33)†	HAQ DI not improved (n = 35)‡	P	HAQ DI improved (n = 33)†	HAQ DI not improved (n = 35)‡	P
Demographic						
Age, years	43.3 ± 14.4	45.1 ± 11.9	0.581	–	–	–
History						
Duration of Raynaud's phenomenon at entry, days	712 ± 1,481	957 ± 1,841	0.550	–	–	–
Physical examination						
Total skin score, 0–51	21.5 ± 7.7	18.8 ± 9.1	0.199	–9.2 ± 7.4	–2.7 ± 10.0	0.004§
HAQ DI score, 0–3	1.13 ± 0.60	0.85 ± 0.54	0.050§	–0.61 ± 0.35	0.28 ± 0.39	0.000§
Maximum oral aperture, mm	46.1 ± 11.3	50.2 ± 9.6	0.116	4.7 ± 7.2	–0.7 ± 8.6	0.008§
Right hand spread, mm	177.3 ± 24.5	183.7 ± 33.8	0.384	–3.3 ± 15.7	–7.1 ± 19.2	0.384
Right fist closure, mm	29.3 ± 18.4	20.2 ± 22.7	0.000§	–12.3 ± 16.2	3.8 ± 18.1	0.000§
Tender joint count, 0–8	1.6 ± 2.2	1.6 ± 2.3	1.000	–0.9 ± 1.9	0.0 ± 2.7	0.032§
Swollen joint count, 0–8	1.0 ± 1.6	0.5 ± 1.0	0.130	–0.9 ± 1.5	0.0 ± 1.8	0.032§
Laboratory						
Hematocrit, %	39.2 ± 3.1	40.4 ± 3.3	0.133	0.6 ± 2.8	–0.5 ± 2.9	0.123
White blood cells, ×1,000/mm ³	8.3 ± 2.1	8.4 ± 2.5	0.861	–1.5 ± 1.8	–0.6 ± 2.2	0.075
Platelets, ×1,000/mm ³	344 ± 98	326 ± 95	0.291	–63 ± 69	–56 ± 88	0.721
ESR, mm/hour	26.3 ± 15.8	17.3 ± 11.1	0.009§	–5.0 ± 14.1	–1.8 ± 10.1	0.291
Creatine kinase, % of upper limit of normal	86.1 ± 203.8	80.6 ± 94.3	0.008§	–0.7 ± 34.7	–19.3 ± 78.7	0.224
DLco, % of predicted	73.3 ± 16.1	78.2 ± 19.0	0.264	3.1 ± 20.9	–6.3 ± 16.7	0.047§
FVC, % of predicted	83.2 ± 21.4	85.7 ± 16.0	0.591	10.2 ± 13.5	–0.2 ± 11.2	0.001§
Creatinine clearance, ml/minute	95.4 ± 33.2	88.5 ± 30.5	0.382	–7.4 ± 30.9	–5.8 ± 37.6	0.851
Serum creatinine, mg/dl	0.9 ± 0.2	0.9 ± 0.2	1.000	0.0 ± 0.2	–0.1 ± 0.2	0.047§

* Except where indicated otherwise, values are the mean ± SD. See Table 1 for definitions.

† Patients whose HAQ DI score improved by ≥0.2 units from the baseline value.

‡ Patients whose HAQ DI score did not improve by ≥0.2 units from the baseline value.

§ $P \leq 0.05$, by 2-tailed unpaired *t*-test.

Using Spearman correlations, univariate logistic regression, and life-table methods, 11 potential predictors of mortality were found at baseline (Table 1), including HAQ DI score (univariate OR 3.22, 95% CI 1.097–9.468). Since FVC and DLco were included in the composite lung involvement parameter, and since lung involvement had a stronger relationship with mortality than FVC or DLco alone, only the composite lung involvement parameter was explored further in multivariate logistic regression. Backward stepwise multivariate logistic regression suggested that of the remaining 9 potential predictors, lung involvement (multivariate OR 6.09, 95% CI 1.613–23.073) and total skin score of ≥ 20 (multivariate OR 3.69, 95% CI 1.169–11.670) together contributed the most to predicting mortality ($R^2 = 0.139$). The addition of other variables did not appreciably increase the R^2 . Although the HAQ DI score was a significant predictor of mortality in univariate regression, it did not contribute significantly in multivariate regression.

SRC and other visceral involvements. At baseline, no patient had ever had an episode of SRC. The HAQ DI score did not correlate with the presence of heart, lung, or renal involvement at baseline. During 4.0 ± 1.1 years of followup, 18 of the 133 patients at risk (14%) developed

SRC. The baseline HAQ DI score predicted neither SRC over 4 years nor new-onset heart or lung involvement over 2 years.

Patient characteristics over 2 years (n = 68 patients). Sixty-six patients withdrew before completing the full 2 years of the study. Two-year data were complete, therefore, only for the 68 patients who completed the 2-year study (completers). Improvement in completers was noted in a number of clinical and laboratory variables. The modified Rodnan skin thickness score decreased by 5.9 ± 9.4 units, the HAQ DI score by 0.17 ± 0.59 units, the tender joint count by 0.4 ± 2.4 units, and fist closure by 3.8 ± 18.9 mm, while FVC (percent of predicted) increased by 4.8 ± 13.4 percentage points and hand spread by 5.3 ± 17.6 mm. Over 2 years of followup, new onset of heart involvement occurred in 8 of 58 patients at risk (14%), while new onset of lung involvement occurred in 7 of 33 patients at risk (21%) (Table 2).

The HAQ DI score decreased (improved) by ≥ 0.2 units compared with the baseline value in 33 patients, and these patients were considered to be "HAQ DI improved" (Tables 2 and 3). The other 35 patients did not show such an improvement and were considered to be "HAQ DI not improved." In HAQ DI–

Table 4. Baseline and 2-year change variables that correlated with change in HAQ DI over the same 2-year period*

	Correlation		Linear regression		Association with improved HAQ DI‡
	Coefficient†	P	R ²	P	
Change variable					
Total skin score	0.492	0.000	0.246	0.000	Decreased skin score
Right fist closure	0.524	0.000	0.240	0.000	Tighter fist
Oral aperture	−0.452	0.000	0.194	0.000	Greater oral aperture
WBC count	0.445	0.000	0.186	0.000	Decreased WBC count
Joint swelling	0.334	0.006	0.137	0.001	Fewer swollen joints
FVC	−0.384	0.002	0.113	0.007	Increased FVC
Right hand spread	−0.217	0.082	0.108	0.008	Increased hand spread
Tender joints	0.276	0.024	0.101	0.009	Fewer tender joints
Contractures	0.255	0.039	0.072	0.029	Fewer contractures
Tendon rubs	0.205	0.097	0.068	0.033	Fewer tendon rubs
Skin ulcers	0.242	0.050	0.064	0.041	Fewer skin ulcers
Baseline variable					
ESR	−0.299	0.016	0.148	0.002	Higher ESR
HAQ DI score	−0.252	0.040	0.126	0.003	Higher HAQ DI score
Swollen joints	−0.237	0.053	0.109	0.006	More swollen joints
Total skin score	−0.309	0.011	0.068	0.034	Higher total skin score
Right hand spread	0.254	0.039	0.065	0.039	Decreased hand spread
Hematocrit	0.224	0.069	0.061	0.043	Lower hematocrit level

* WBC count = white blood cell count (see Table 1 for other definitions).

† For positive correlations, higher values show stronger relationship with change in HAQ DI. For negative correlations, lower values show stronger relationship with change in HAQ DI.

‡ For change variables, the direction of the change is shown. For baseline variables, the baseline level is shown.

improved patients, baseline values for the HAQ DI, fist closure, erythrocyte sedimentation rate (ESR), and CK were higher (more abnormal), and heart involvement more frequent, than in the HAQ DI-not improved group ($P = 0.000$ to $P = 0.050$) (Tables 2 and 3). Otherwise, baseline characteristics were similar in the 2 groups.

At 2 years, HAQ DI-improved patients showed significantly more improvement than HAQ DI-not improved patients in total skin score, HAQ DI score, oral aperture, tender and swollen joint counts, fist closure, DLco (percent of predicted), FVC (percent of predicted), and serum creatinine level ($P = 0.000$ to $P = 0.047$) (Table 3). Also, HAQ DI-not improved patients were more likely to have developed new onset of non-digital tip cutaneous ulcers and new-onset lung involvement at 2 years than were HAQ DI-improved patients ($P = 0.013$ and $P = 0.041$, respectively) (Table 2). Therefore, these data support the conclusion that improvement in the HAQ DI score was associated with improvement in multiple other important aspects of disease involvement, including skin thickness, hand function, and oral aperture, as well as lung, kidney, and joint involvement.

Correlates of outcome over 2 years (n = 68 patients). *HAQ DI.* The baseline and change variables which were related to changes in HAQ DI scores (by Spearman correlation coefficient and univariate regression R^2) are shown in Table 4. When these relationships were analyzed by backward stepwise multiple linear regression, 1 baseline variable and 3 change variables together explained 53% of

the variance of the change in HAQ DI ($R^2 = 0.528$). These were ESR at baseline ($P = 0.005$) and changes in the swollen joint count ($P = 0.002$), total skin score ($P = 0.005$), and white blood cell count ($P = 0.005$). The addition of other variables did not increase the R^2 appreciably.

Improvement in HAQ DI. No baseline variable predicted improved the HAQ DI at 24 months in the 68 patients who completed the trial.

Global improvement. Of the 68 patients who completed the 2-year trial, 50 were judged to have “improved” the globally. Using Spearman correlations and univariate logistic regression, the absence of large joint contractures (multivariate OR 4.46, 95% CI 1.230–16.136) and the presence of higher hematocrit levels (defined as $\geq 41.6\%$ in men and $\geq 38.6\%$ in women) (multivariate OR 4.11, 95% CI 1.131–14.950) at baseline were the only variables that together best explained global improvement ($R^2 = 0.157$). Addition of other variables did not improve the prediction.

Conversely, 10 change variables correlated with global improvement (Table 5). Two of the change variables (change in total skin score and change in HAQ DI) had exceptionally strong associations with global improvement. When analyzed in multivariate logistic regression, improvement in the total skin score (multivariate OR 18.91, 95% CI 3.058–116.981) and improvement in the HAQ DI (multivariate OR 17.96, 95% CI 2.890–111.550) were the only 2 variables that indepen-

Table 5. Change over 2 years in variables that correlated with or were associated with physician's (MD) global assessment of "improved" over the same 2-year period*

Change in variable	Correlation		Univariate logistic regression				Change associated with MD global assessment of "improved"
	Coefficient†	P	OR	95% CI	P		
Total skin score	-0.685	0.000	24.62	5.475-110.711	0.000		Improved skin score
HAQ DI score	-0.581	0.000	22.15	4.469-109.463	0.000		Improved HAQ DI score
Right fist closure	-0.434	0.000	4.20	1.181-14.937	0.027		Tighter fist
Oral aperture	0.437	0.000	3.89	1.212-12.460	0.022		Increased aperture
WBC count	-0.403	0.001	6.16	1.617-23.463	0.008		Decreased WBC count
Contractures	-0.460	0.000	5.62	0.677-46.685	0.110		Fewer contractures
Right hand spread	0.359	0.003	3.22	0.817-12.677	0.095		Greater hand spread
Skin ulcers	-0.302	0.008	2.60	0.297-22.868	0.388		Fewer ulcers
FVC	0.264	0.035	1.90	0.611-5.887	0.268		Increased FVC
Tender joints	-0.264	0.030	0.92	0.249-3.381	0.896		Fewer tender joints

* WBC count = white blood cell count (see Table 1 for other definitions).

† For positive correlations, higher values predict an MD global assessment of "improved." For negative correlations, lower values predict an MD global assessment of "improved."

dently contributed to the model ($R^2 = 0.445$). Addition of other variables did not improve the prediction.

Correlates of HAQ DI score at baseline and at 2 years in completers (n = 68 patients). We were concerned that the relationship of the baseline HAQ DI score with variables in the 68 completers might be different from the relationship we found previously in the 134 patients of the initial cohort. Using stepwise logistic regression in these 68 patients, we found that the oral aperture and hand spread at baseline (Table 6) best explained the model ($R^2 = 0.374$). No other variables remained in the model.

We also examined the relationship of other explanatory variables to the HAQ DI score at 2 years in the completers (Table 6). We recalculated medians for continuous variables at 2 years in these 68 patients. We then converted continuous variables to binary variables (using cut points at or near the median) while maintaining naturally occurring dichotomous variables as binary variables. Stepwise multivariate logistic regression showed that a model including oral aperture, hand spread, tender joints, and fist closure together best explained the relationship ($R^2 = 0.586$).

Relationship between HAQ DI score and total skin score. The HAQ DI score and total skin score at baseline were highly correlated (correlation coefficient of 0.368, $P = 0.000$) (3), as were the courses of the HAQ DI score and total skin score over the 2-year study period (correlation coefficient 0.4918, $P = 0.000$). We explored this relationship further by comparing the courses of change in HAQ DI and its 8 components in patients whose skin score decreased by $\geq 25\%$ over 2 years (skin improved) with those courses in patients whose skin score did not improve by $\geq 25\%$ over the 2 years (skin not improved). While no

baseline differences were found (Table 7), there were significantly greater improvements over 2 years in the arising, walking, hygiene, and reaching components of the HAQ DI in skin-improved patients compared with skin-not-improved patients (Table 7). Conversely, there were no differences between skin-improved and skin-not-improved patients in the courses of the dressing, eating, gripping, or activity components of the HAQ DI over 2 years.

DISCUSSION

Although the level of functional impairment in SSc has now been documented in several reports (3-9), few published studies have documented how function in SSc changes over time. Steen and Medsger recently reported their analysis of HAQ DIs that were completed annually by 1,250 SSc patients attending the University of Pittsburgh Scleroderma Clinic over a mean of 4.0 years (5). Their major findings were that HAQ DI scores in diffuse scleroderma correlated directly with skin

Table 6. Variables that had a significant ($P < 0.05$) relationship to HAQ DI at baseline and at 2 years by stepwise multivariate logistic regression in the 68 patients who completed the 2-year study*

Variables	OR	95% CI	P
At baseline			
Oral aperture	6.882	1.955-24.233	0.003
Hand spread	5.820	1.784-18.993	0.004
At 2 years			
Oral aperture	9.681	1.724-54.353	0.010
Hand spread	10.785	1.952-59.586	0.006
Tender joints	21.208	2.757-163.108	0.003
Fist closure	10.347	1.769-60.513	0.010

* See Table 1 for definitions.

Table 7. Baseline and change score values at 2 years for the 8 component scores of the HAQ DI (ordinal) in the 68 systemic sclerosis patients who completed the 2-year study, classified according to improvement in final skin score*

HAQ DI variable	Baseline values			Change score values at 2 years		
	Skin improved (n = 42)†	Skin not improved (n = 26)‡	P	Skin improved (n = 42)†	Skin not improved (n = 26)‡	P
Dressing	0.857 ± 0.718	0.923 ± 0.744	0.718	−0.146 ± 0.823	0.154 ± 1.008	0.187
Arising	0.881 ± 0.633	0.654 ± 0.689	0.169	−0.463 ± 0.552	0.038 ± 0.824	0.009§
Eating	1.048 ± 0.764	1.154 ± 1.047	0.631	−0.073 ± 0.848	−0.154 ± 0.849	0.726
Walking	0.667 ± 0.570	0.692 ± 0.549	0.856	−0.220 ± 0.652	0.192 ± 0.849	0.029§
Hygiene	1.190 ± 1.042	1.038 ± 1.076	0.566	−0.293 ± 0.716	0.423 ± 0.809	0.001§
Reaching	1.071 ± 0.808	1.115 ± 0.864	0.833	−0.463 ± 0.977	0.077 ± 1.093	0.039§
Gripping	0.881 ± 0.633	1.192 ± 0.895	0.129	−0.244 ± 0.734	−0.231 ± 0.951	0.950
Activity	1.119 ± 0.803	1.385 ± 0.852	0.200	−0.390 ± 0.703	−0.077 ± 1.017	0.176
Overall	0.964 ± 0.550	1.019 ± 0.649	0.710	−0.287 ± 0.486	−0.053 ± 0.664	0.019§

* Values are the mean ± SD. HAQ DI = Disability Index of the Health Assessment Questionnaire.

† Patients whose final skin score decreased by ≥25% of the baseline value.

‡ Patients whose final skin score did not decrease by ≥25% of the baseline value.

involvement, scleroderma heart and kidney disease, tendon friction rubs, finger contractures, and proximal muscle weakness at entry; over time, changes in HAQ DI scores correlated with changes in total skin scores and was a good predictor of survival.

Our recently published analysis of the relationship of baseline HAQ DI score to other physical and laboratory parameters at entry (3) and our present analysis of changes in HAQ DI over time confirm several of Steen and Medsger's findings relative to the relationship of HAQ DI to other variables. We found that the HAQ DI score was directly correlated with skin and joint involvement, scleroderma heart disease, tendon friction rubs, hand contractures, and proximal muscle weakness at entry. Also, changes in the HAQ DI were highly correlated with changes in skin score over 2 years (correlation coefficient 0.4918).

We also found that the HAQ DI score was predictive of survival. We realize, however, that this relationship between functional status and mortality is not unique to SSc; such a relationship has been noted previously in RA, in general medical patients, and in congestive heart failure (18–20). In the present study, when we further explored the relationship of the HAQ DI score and other variables to the prediction of survival using multivariate logistic regression, we found that a combination of lung involvement and total skin score of ≥20 at baseline was superior to the HAQ DI score (alone or in combination) as a predictor of survival.

More importantly, we explored the relationship of changes in HAQ DI scores to changes in other physical examination and laboratory variables over time in the cohort of 68 patients who completed the 2-year study. The question of what represents a meaningful change or difference in HAQ DI from the patient's point

of view has been addressed and answered in 2 recent cross-sectional studies of RA patients (10,11). Both studies reported that a difference in HAQ DI score of ≥0.2 units was judged by patients to be symptomatically important (10,11). Although no similar study in SSc patients has been reported, we have assumed for this analysis that SSc patients would have rated themselves in a similar manner.

The HAQ DI score decreased by ≥0.2 units in 33 of the 68 completers, and they might therefore have rated themselves at least "somewhat better" (an important symptomatic difference) over the study period. Our analysis showed that patients whose HAQ DI score decreased by ≥0.2 units (HAQ DI-improved patients) also showed greater improvements in numerous other physical examination and laboratory parameters than did the HAQ DI-not improved patients (Tables 2 and 3). These improvements included skin involvement (total skin score), joint involvement (tender and swollen joint counts, tendon friction rubs), oral aperture, lung involvement (DLco, FVC), serum creatinine level, and hand function (fist closure).

We queried physician investigators about how each of their patients fared over the 2-year study period (physician's global assessment). In multivariate logistic regression, only 2 variables remained as independent contributors to the physician's global assessment: change in total skin score (multivariate OR 18.91) and change in HAQ DI score (multivariate OR 17.96), with $R^2 = 0.445$. This suggests that the physician's assessment of the patient's course is greatly influenced by improvements in skin thickening and in function.

In our recent analysis of HAQ DI scores at baseline (3), we hypothesized that the major factors that determine the HAQ DI score early in the disease course

might not be those that influence the HAQ DI later in time. At 2 years, we found that hand function was still a very important influence on the HAQ DI score. We also found, however, that the improvements in HAQ DI recorded in skin-improved patients over 2 years were attributable primarily to the arising, walking, hygiene, and reaching components of the HAQ DI and not to the dressing, eating, gripping, or activity components. In reviewing the 20 questions that make up the HAQ DI, we noted that only 2 of the 9 questions that ask about arising, walking, hygiene, and reaching are directly related to hand function, while 8 of the 11 questions that ask about dressing, eating, gripping, or activity are directly related to hand function. This suggests that although the total HAQ DI score (at baseline and at 2 years) was heavily influenced by hand function, the improvement in function that occurred over 2 years was largely related to improvement in factors other than hand dysfunction (i.e., possible improvements in large joint flexibility, muscle strength, joint pain, and energy level).

In summary, our analyses further elucidate the performance characteristics and clinical usefulness of the HAQ DI in SSc patients with recent-onset diffuse scleroderma (3,5,14). Although our data cannot specifically address how severe heart, lung, or kidney involvement might have affected the HAQ DI score, the HAQ DI appears to have several attributes which are useful in the clinical management of SSc patients with early diffuse scleroderma. The HAQ DI can be used to document functional impairment at a given point in time (i.e., the initial visit) and how it changes over time. Baseline HAQ DI scores of ≥ 1.0 predict a greater risk of mortality. Improvement (decline) in the HAQ DI score suggests that a number of other SSc physical and laboratory involvements are probably improving as well. Finally, our data suggest that the assessment of patient function using the HAQ DI (a self-administered questionnaire that SSc patients can easily complete in less than 5 minutes), coupled with physician assessment of the total skin score (16), provides the practicing physician with important information about prognosis, patient status, and changes in disease course over time in SSc patients with early diffuse scleroderma.

REFERENCES

1. 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.
2. Seibold JR. Scleroderma. In: Kelley WN, Harris ED, Ruddy S, Sledge CB, editors. *Textbook of rheumatology*. Philadelphia: WB Saunders; 1997. p. 1133–62.
3. 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.
4. 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.
5. 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.
6. Moser DK, Clements PJ, Brecht M-L, Weiner SR. Predictors of psychosocial adjustment in systemic sclerosis: the influence of formal education level, functional ability, hardness, uncertainty, and social support. *Arthritis Rheum* 1993;36:1398–405.
7. Roca RP, Wigley FM, White B. Depressive symptoms associated with scleroderma. *Arthritis Rheum* 1996;39:1035–40.
8. Merkel PA, Martin RW, Herlyn K. Health status measurements in patients with Raynaud's phenomenon and scleroderma [abstract]. *Arthritis Rheum* 1997;40:S255.
9. Callahan LF, Smith WJ, Pincus T. Self-report questionnaires in five rheumatic diseases: comparisons of health status constructs and associations with formal education level. *Arthritis Care Res* 1989;2:122–31.
10. Redelmeier DA, Lorig K. Assessing the clinical importance of symptomatic improvement: an illustration in rheumatology. *Arch Intern Med* 1993;153:1337–42.
11. Wells GA, Tugwell P, Kraag GR, Baker PR, Groh J, Redelmeier DA. Minimum important difference between patients with rheumatoid arthritis: the patient's perspective. *J Rheumatol* 1993;20:557–60.
12. Medsger TA Jr, Steen VD. Classification, prognosis. In: Clements PJ, Furst DE, editors. *Systemic sclerosis*. Baltimore: Williams and Wilkins; 1996. p. 51–64.
13. 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.
14. 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.
15. 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.
16. Clements PJ, Hurwitz EL, Wong WK, Seibold JR, Mayes M, White B, et al. Skin thickness score is a predictor and correlate of outcome in systemic sclerosis: high-dose versus low-dose penicillamine trial. *Arthritis Rheum* 2000;43:2445–54.
17. Nagelkerke NJD. A note on a generalized definition of the coefficient of determination. *Biometrika* 1991;78:691–2.
18. Pincus T, Brooks RH, Callahan LF. Prediction of long-term mortality in patients with rheumatoid arthritis according to simple questionnaire and joint count measures. *Ann Intern Med* 1994;120:26–34.
19. Inouye SK, Peduzzi PN, Robison JT, Hughes JS, Horwitz RI, Concato J. Importance of functional measures in predicting mortality among older hospitalized patients. *JAMA* 1998;279:1187–93.
20. Edwards BS, Rodeheffer RJ. Prognostic features in patients with congestive heart failure and selection criteria for cardiac transplantation. *Mayo Clin Proc* 1992;67:485–92.