Observational Study of Treatment Outcome in Early Diffuse Cutaneous Systemic Sclerosis

ARIANE L. HERRICK, MARK LUNT, NINA WHIDBY, HOLLY ENNIS, ALAN SILMAN, NEIL McHUGH, and CHRISTOPHER P. DENTON

ABSTRACT. ***Objective.*** Randomized clinical trials in early diffuse cutaneous systemic sclerosis (dcSSc) are challenging. We used an observational approach to estimate the relative effectiveness of different current treatment approaches, capturing entry and outcome data in a standardized way.

Methods. Patients with dcSSc within 3 years of the onset of skin thickening were included. Standardized entry and followup data were collected in relation to the first disease-modifying treatment at baseline and 4-6 weeks, then 3, 6, 12, 18, 24, 30, and 36 months. The 5 different protocols were (1) intravenous cyclophosphamide followed by mycophenolate mofetil (MMF); (2) antithymocyte globulin followed by MMF; (3) MMF alone; (4) no disease-modifying treatment; (5) other immunosuppressant treatment. The primary outcome measure was the modified Rodnan skin score (mRSS). Inverse probability of treatment weights were used to allow for differing patient characteristics between groups.

Results. The study included 147 patients from 12 centers. Numbers of patients starting on Protocols 1 to 5 were 29, 25, 61, 19, and 13, respectively. mRSS decreased over time from 24 (IQ 19–32) at baseline to 15.5 (IQ 9–24.5) at 3 years. Although there were differences in the magnitude of the change for different protocols, there were no significant differences between protocols in the rate of change of mRSS over time ($p = 0.43$). When inverse probability weights were applied, the results remained nonsignificant ($p = 0.41$).

Conclusion. Using this observational approach, there were no obvious differences in outcome between groups after allowing as far as possible for baseline differences in treatment allocations. (J Rheumatol First Release Dec 1 2009; doi:10.3899/jrheum.090668)

Key Indexing Terms:
DIFFUSE CUTANEOUS SYSTEMIC SCLEROSIS
TREATMENT
OBSERVATIONAL STUDY
OUTCOME

Early diffuse cutaneous systemic sclerosis (dcSSc) is associated with high morbidity and mortality. Patients with early diffuse disease often experience a rapid onset and progression of skin thickening that spreads to involve proximal limb and/or trunk[1]; they are at high risk of internal organ involvement including pulmonary fibrosis and accelerated hypertension/renal crisis.

Effective treatments are available for control of specific organ-based manifestations of the systemic sclerosis (SSc) disease process[2]. For example, proton pump inhibitors for upper gastrointestinal symptoms, angiotensin converting enzyme inhibitors for renal involvement, and cyclophosphamide, recently shown to confer modest benefit in 2 randomized controlled clinical trials of SSc-related pulmonary fibrosis[3,4]. However, despite major advances in our understanding of the molecular and cellular pathology of the underlying disease process[5,6], there is still no known effective disease-modifying therapy[7].

Studies investigating immunosuppressive treatment of dcSSc are most appropriate in early disease (within 3 years of onset of skin thickening) because it is within this time that the disease progresses most rapidly and that immunosuppressive therapy is most likely to be effective and justified. Recent years have seen a number of well-designed controlled clinical trials in early diffuse disease, but none has identified an effective treatment: trials of interferon-alpha[8], D-penicillamine[9], and anti-transforming growth factor-ß antibody therapy[10] have all been disappointing. Further, despite a trend in favor of a reduction in skin score with methotrexate (MTX), this agent is also of limited efficacy[11].

From the University of Manchester; the Arthritis Research Campaign, Chesterfield; the Royal National Hospital for Rheumatic Diseases, Bath; the Royal Free Hospital, London, United Kingdom.

Supported by the Scleroderma Society and by the Arthritis Research Campaign.

A.L. Herrick, Reader in Rheumatology, FRCP; M. Lunt, Senior Lecturer in Biostatistics, PhD; N. Whidby, Research Assistant, MB ChB; H. Ennis, Research Assistant, PhD, University of Manchester; A.J. Silman, Medical Director, Arthritis Research Campaign, FRCP; N. McHugh, Professor of Rheumatology, MB ChB, Royal National Hospital for Rheumatic Diseases; C.P. Denton, Professor of Experimental Rheumatology, FRCP, Royal Free Hospital.

Address correspondence to Dr. A.L. Herrick, ARC Epidemiology Unit, Stopford Building, University of Manchester, Manchester M13 9PL, United Kingdom. E-mail: ariane.herrick@manchester.ac.uk

Accepted for publication July 10, 2009.

Stem cell transplantion is currently being evaluated[12,13], but even if effective is likely to be restricted to a minority of severe cases with evidence of significant internal organ involvement.

The scarcity of controlled clinical trials of early diffuse disease reflects their intrinsic difficulty. DcSSc is rare, meaning that trials have to be multicenter and often international. In addition, many clinicians have reservations about recruiting patients with potentially life-threatening disease into trials including a placebo arm. The situation is complicated by the fact that many patients already have significant internal organ involvement when being assessed for inclusion, and may be ineligible depending on how strictly the inclusion and exclusion criteria have been defined, thus randomized trials may exclude those patients in whom a disease-modifying therapy is most needed. Possibly of greater relevance is the fact that SSc is an "orphan" disease, and industry-sponsored studies are infrequent and investigator-initiated studies are difficult to fund.

Given that, in practice, patients with this disorder are treated with a variety of different agents, it is possible if entry and outcome data are captured in a standardized way to estimate the relative effectiveness of these agents in an observational way. Clearly this approach is non-randomized, but recent analytical methods suggest that there is potential to adjust for differences in treatment choice (confounding by indication), which would give an estimate of treatment effects. Against this background, the UK Scleroderma Study Group embarked upon an observational study to examine different treatments for which there was some evidence (although not from controlled trials) for efficacy, but at the same time giving clinicians the option to make their decision just to observe patients without prescribing immunosuppressant therapy. The study commenced in 2000, and the choice of treatment arms reflected (1) that the combination of antithymocyte globulin (ATG) and mycophenolate mofetil (MMF) had been associated with some clinical benefit in a small open study[14]; and (2) that cyclophosphamide was believed by many clinicians to confer benefit with a number of open studies at that time suggesting efficacy in patients with SSc-related lung disease. The observational study was designed to be all-inclusive, reflecting clinical practice and the heterogeneity of patients with early dcSSc. We investigated whether currently used approaches to suppress disease activity as early as possible after presentation with dcSSc beneficially affect outcome.

MATERIALS AND METHODS

Study design. This was an observational cohort study of patients with early dcSSc that involved collecting standardized entry and followup data in relation to the first disease-modifying treatment offered within the context of the study. Most of the data were collected prospectively, although in centers with comprehensive clinical databases, some retrospective data entry was allowed as long as for each patient, the baseline visit date was within 3 years of onset of skin thickening, at or after commencement of the study in 2000. Patients were included into the study between August 2000 and July 2007.

To enhance recruitment, all UK rheumatologists were encouraged to either refer patients to one of the UK specialist centers where patients were treated as per the proposed protocols, or to treat the patients per protocol themselves and document/investigate patients along standard guidelines. Clinicians selected the protocol of their choice for each patient. Patients were assessed at baseline, at 4-6 weeks and then at 3, 6, 12, 18, 24, 30, and 36 months.

Inclusion criteria were dcSSc (skin involvement proximal to elbow, knee, face, neck[1]) and within 3 years of the onset of skin thickening. If any patient had a contraindication to any of the study drugs, then this precluded that patient being prescribed that particular treatment. If a patient had received any immunosuppressant or antifibrotic drug within the previous 1 month (e.g., cyclosporine, penicillamine), then 1 month was allowed to elapse before that patient entered the study. Corticosteroids were not a contraindication to entry, but the dose was kept constant wherever possible.

The recommended treatment protocols, decided after discussion among members of the UK Scleroderma Study Group, were
(1) intravenous (IV) cyclophosphamide followed by MMF. IV cyclophosphamide (15 mg/kg) was given monthly for 6 months, followed by MMF for 6 months (500 mg bd, increased after 2 weeks if tolerated to 1 g bd); (2) ATG followed by MMF. ATG infusions were given daily for 5 days, commencing at 2.5 mg/kg/day. One month after entry, daily MMF was commenced and given for 11 months (500 mg bd, increased after 2 weeks if tolerated to 1 g bd); (3) MMF alone, for 12 months (500 mg bd increased after 2 weeks if tolerated to 1 g bd); (4) no disease-modifying treatment; (5) other immunosuppressant treatment. Reflecting that this was an observational study, at the outset it was accepted that patients might be prescribed other immunosuppressants (e.g., MTX) as a result of physician/patient preference, and all such patients were aggregated into this last group. This option also allowed for changing therapeutic practice as the study evolved over time.

From its inception the study was discussed with the Research Ethics Committee and because this was not a randomized controlled trial, but simply an anonymized collection of otherwise routinely collected clinical data, from patients treated according to physician choice of "best guess" from a list of currently agreed approaches in the absence of a proven effective treatment, the investigators were formally advised that ethical approval was not required and that it was not necessary for patients to sign informed consent.

Patients. One hundred forty-seven patients with dcSSc (from 11 centers in the UK, excepting 4 patients from Slovenia), all of whom satisfied the American College of Rheumatology (ACR) criteria[15], were included in the study. Demographic characteristics including age, gender, ethnicity, smoking habit, antibody status [anti-topoisomerase-1 (anti-Scl-70), anti-RNA polymerase] and presence of visceral organ involvement were recorded for all patients. Specifically, clinicians were asked to document, in addition to skin involvement, the presence/absence of pulmonary hypertension (as estimated on echocardiography or measured at right heart catheterization), pulmonary fibrosis, and cardiac involvement. Pulmonary fibrosis was defined as radiological evidence of basal fibrosis with reduction in transfer factor (and usually confirmed on computed tomography scan). Pulmonary arterial hypertension was defined as an estimated pulmonary artery systolic pressure of > 30 mmg Hg on echocardiography, or a raised mean pulmonary artery pressure at right heart catheterization. Cardiac involvement was defined as a conduction defect on electrocardiogram, arrhythmia, impaired left ventricular function, or clinically evident congestive cardiac failure. Renal involvement was defined by an elevated plasma creatinine, impaired creatinine clearance, or a history of accelerated hypertension (renal crisis), and reported as plasma creatinine.

Specifically, baseline characteristics recorded/measured included (1) functional ability as measured by a modification of the disability index of the Health Assessment Questionnaire (HAQ-DI)[16] and the 11-item sclero-

derma Functional Questionnaire[17]. This scores on the scale 0–3, where 0 = able to perform in normal manner, 1 = can manage with some alteration of style, 2 = can only manage with difficulty, 3 = impossible to achieve, with a maximum score of 33; and (2) pulmonary function as measured by forced vital capacity and carbon monoxide diffusing capacity.

Outcome measures. The primary outcome measure was the modified Rodnan skin score (mRSS)[18] in which the skin was assessed clinically at 17 body sites on a 0–3 scale, where 0 = uninvolved, 1 = mildly thickened, 2 = thickened, 3 = severely thickened. The maximum score is 51. The skin score was measured at each of the time points specified above.

Statistical analysis. Since this was an observational, not a randomized study, there was a possibility that observed differences in outcome between the treatment protocols were due to differing patient characteristics between the treatment groups, rather than different effects of the treatments themselves (confounding by indication). To examine this possibility, we needed to compare the baseline characteristics between the treatment groups to see if there were differences, and also examine the associations between the baseline characteristics and the skin score. Only variables that differed between protocols and affected the skin score were considered as potential confounders in the subsequent analysis.

Examination of differences in baseline characteristics between the treatment subgroups. Baseline characteristics for treatment protocols were compared using the Kruskal-Wallis test for continuous variables and the Fisher's exact test for categorical variables. Continuous variables are presented as the median and interquartile range (IQR). Categorical variables are expressed as frequencies and percentages in each category.

Examination of the influence of baseline characteristics on skin score at baseline and over time. The association between baseline predictors and skin score was examined using linear regression. All available skin score measures were used, using a robust standard error to allow for correlations between different measurements in the same individual. The predictor variables in this model were time (measured in years), 1 baseline predictor, and the interaction between the baseline predictor and time. The coefficient of the time variable gives the mean change in skin score per year, the coefficient of the baseline predictor measures the effect of the predictor on the skin score at baseline, and the interaction term measures the effect of the predictor on changes in the skin score over time.

Examination of skin score change between the different treatment groups. To allow for the fact that the patient characteristics may differ between the different protocols, inverse probability of treatment (IPT) weights were used[19]. While this method is most commonly used to compare 2 exposures, it has been extended to cover multiple groups by Imbens[20]. First, multinomial logistic regression is used to calculate, for each subject, the probability that they would be assigned to each protocol given their baseline characteristics. A subject's weight is calculated as 1/ptreat, where ptreat is the calculated probability of assignment to whichever protocol the subject was, in reality, assigned. This weighting will, in the long run, balance all covariates between the protocols. However, some subjects received very high weights, giving them great influence in the final result. Since these subjects are by definition unusual (the weight is large, so ptreat is small, i.e., they were unlikely to receive the treatment they did receive), the analysis was rerun after trimming subjects with a weight greater than 20[21].

The differences between protocols were again assessed using linear regression, with time as the predictor variable. A separate intercept and slope was fitted to each protocol, and the change in skin score with time was used to assess the effect of treatment, with a large decrease per unit time representing an effective treatment. Using the IPT weights with this regression equation should give an estimate of the change in skin score over time in each protocol without confounding by the baseline predictors.

A further potential cause of confounding was differential loss to followup between the protocols, either through death, subjects dropping out of the study, or subjects recruited less than 3 years before the analysis. For this reason, we looked at mortality in each of the protocol arms, and the numbers of subjects who did not complete 3 years of followup. In addition, there was a considerable amount of missing data for some of the baseline characteristics, in particular the HAQ-DI and functional ability scores as measured by the 11-item Functional Questionnaire[17]. Since a complete case analysis could have introduced bias, multiple imputation was used to enable all subjects to be included in the analysis. Values were imputed using ice[22], an implementation of imputation by chained equations[23] in STATA 9.2. Its companion package, mim, was used to analyze the imputed datasets separately, and produce appropriate effect estimates and standard errors using "Rubin's rules"[24]. The statistical analysis was conducted using STATA 9.2.

RESULTS

There were 147 patients from 12 different centers. Ninety percent came from 3 centers: the Royal Free Hospital, London (96 patients/65%), Salford Royal Hospital (23 patients/16%) and the Royal National Hospital for Rheumatic Diseases, Bath (12 patients/8%). Only 3 of the 147 patients had a baseline visit retrospective to commencement of the study in August 2000.

As this was not a randomized study, the number of participants starting on each treatment protocol differed: 29 on protocol 1 (IV cyclophosphamide followed by MMF), 25 on protocol 2 (ATG followed by MMF), 61 on protocol 3 (MMF), 19 on protocol 4 (no disease-modifying/immunosuppressant treatment), and 13 on protocol 5 (other immunosuppressant therapy). Of the 19 patients on protocol 4, 5 were on active treatment (all MMF) within 6 months, 4 more within 1 year (2 MMF, 2 MTX), and 1 more within 2 years (azathioprine).

Baseline characteristics of patients

Table 1 shows the key clinical characteristics of patients as a single cohort and subdivided into the 5 treatment groups.

The majority of patients (78%) were of white origin and 16% were current smokers. There were differences between treatment groups in age (p = 0.01; patients on Protocol 5 were younger than those on Protocol 4) and in steroid use at baseline. The differences in gender distribution and in autoantibody status between protocols were not statistically significant (Table 1).

MRSS and functional ability. There were significant differences between groups (p = 0.0001) in mRSS, which was highest in Protocol 2 and lowest in Protocols 4 and 5 (Table 1). Scores for the HAQ-DI and the Functional Questionnaire did not differ significantly between the groups (p = 0.41 and p = 0.10, respectively).

Organ involvement. There were significant differences between treatment groups for presence of pulmonary hypertension, pulmonary fibrosis, and cardiac problems (p = 0.001, p = 0.008, and p = 0.024, respectively; Table 1). The proportions of patients with pulmonary and cardiac problems were highest in Protocol 1. The baseline differences for the different laboratory measurements (hemoglobin, platelet count, erythrocyte sedimentation rate, plasma creatinine, and pulmonary function tests) for different treatment proto-

Table 1. Baseline characteristics and antibody status of patients with systemic sclerosis by treatment group. Values are median (IQR) unless otherwise indicated.

Characteristic	Protocol 1 n = 29	Protocol 2 n = 25	Protocol 3 n = 61	Protocol 4 n = 19	Protocol 5 n = 13	p**	Total (overall) n = 147
Age, yrs	55.1 (49.9–64.1)	52.7 (45.7–57.8)	54.4 (43.0–61.3)	55.6 (50.1–67.2)	40.9 (32.3–51.5)	0.01	53.1 (43.3–61)
Women (%)	18 (62)	20 (80)	44 (72)	13 (68)	8 (62)	0.60	103 (70)
Antitopoisomerase (anti-Scl-70) no. (%)*	8 (33)	5 (20)	14 (24)	6 (33)	4 (33)	0.72	37 (27)
Anti-RNA polymerase III, no. (%)*	1 (8)	1 (4)	9 (22)	1 (14)	0 (0)	0.15	13 (12)
On corticosteroids, no. (%)	9 (31)	17 (68)	16 (26)	7 (37)	8 (62)	0.002	57 (39)[1]
Pulmonary hypertension, no. (%)	6 (21)	0 (0)	0 (0)	2 (11)	0 (0)	0.001	8 (6)
Pulmonary fibrosis, no. (%)	16 (57)	7 (28)	11 (18)	6 (32)	3 (23)	0.008	43 (30)
Cardiac involvement, no. (%)	8 (28)	1 (4)	3 (5)	2 (11)	1 (8)	0.024	15 (10)
mRSS, (0–51)	24 (20.5–30)	32 (30–41)	23.5 (18–30)	20.5 (18–22)	21 (13–27)	0.0001	24 (19–32)
HAQ-DI score, 0–3	1.19 (0.56–2.00)	1.63 (1.38–2.50)	1.50 (0.75–2.13)	1.00 (0.25–2.00)	1.50 (0.00–2.63)	0.41	1.50 (0.75–2.13)
Functional Questionnaire, 0–33	11.5 (3–18)	18 (10–21)	14.5 (7–21)	10 (1–23)	10 (0–11)	0.10	13.5 (6–19)
Hemoglobin, g/l	127 (114–134)	117 (114–130)	124 (116–137)	123 (112–131)	121 (114–134)	0.60	123 (114–134)[2]
Platelets × 10^9/l	312 (257–350)	324 (270–414)	329 (297–399)	370 (284–441)	370 (305–432)	0.29	328 (278–403)
ESR, mm/h	22 (7–37)	21 (8.5–36)	22 (10–34.5)	28 (13.5–36)	13 (9–18)	0.72	22 (10–36)
Plasma creatinine, μmol/l	70 (57–87)	68 (64–82)	70 (63–83)	78 (72–104)	63.5 (58–72)	0.067	71 (63–84)
Pulmonary function tests							
FVC, % predicted	76.0 (64.0–91.0)	93.3 (86.2–101.5)	87.8 (70.1–97.0)	88.7 (81.0–97.0)	84.7 (89.9–100.5)	0.16	87.7 (70.4–97.0)
DLCO, % predicted	58.5 (45.0–79.0)	76.1 (64.3–81.6)	71.5 (54.0–87.0)	69.4 (48.6–85.0)	81.0 (65.5–91.4)	0.091	71.0 (54.0–84.0)

* Percentages related to the numbers tested. ** Significance p: Fisher's exact test, for categorical variables; Kruskal-Wallis test, for continuous variables. [1] Mean dose 15.7 mg (SD 10.2). [2] 24/76 women (32%) and 18/39 men (46%) were anemic (hemoglobin < 115 g/l for women and 135 g/l for men. IQR: interquartile range; mRSS: modified Rodnan skin score; HAQ-DI: Health Assessment Questionnaire-Disability Index; ESR: erythrocyte sedimentation rate; FVC: forced vital capacity; DLCO: diffusion capacity for carbon monoxide.

cols were not statistically significant (Table 1). Although 9 patients in Protocol 3 were known to be anti-RNA polymerase III antibody positive, none of these 9 developed renal crisis.

Progression through the study

Figure 1 shows how patients progressed through the study. Overall, 68 patients completed 3 years of followup, 23 patients died before reaching 3 years of followup, 25 patients had not yet reached a 3 year time point, 10 patients withdrew from the study, and 21 were lost to followup by the end of the year.

A total of 17 subjects changed protocol, 2 of them changing twice (from Protocol 2 to 4 and then to 5, and from Protocol 4 to 3 and then to 5). The 19 protocol changes were (1) from Protocol 1, 1 change to Protocol 4; (2) from Protocol 2, 1 change to Protocol 4; (3) from Protocol 3, 1 change to Protocol 1, 2 to Protocol 4, 2 to Protocol 5; (4) from Protocol 4, 7 changes to Protocol 3, 3 to protocol 5; and (5) from Protocol 5, 2 changes to Protocol 3.

Out of 10 patients who withdrew, 2 moved abroad, 3 were unable or did not wish to travel to their specialist unit, 3 declined treatment, 1 experienced adverse effects of treatment and did not wish further followup, and 1 patient experienced a deterioration in health and did not wish to continue treatment.

Nine patients developed renal involvement during followup. Two of them were diagnosed as having renal crisis 2 months after baseline (both were on prednisolone at baseline, one 10 mg daily and the other 5 mg).

Of the 23 patients who died, 6 died of cancer, of whom 3 had lung cancer, 1 had oropharynx cancer, 1 esophageal cancer, and 1 pancreatic cancer. Ten patients died of cardiac problems and multiorgan involvement related to SSc (3 were reported to have died of myocardial infarction, 1 cardiomyopathy, 1 pulmonary arterial hypertension, 1 adult respiratory distress syndrome, 1 cardiorespiratory failure, 1 congestive cardiac failure, and 2 "disease progression") and 1 died from sepsis secondary to bilateral pneumonia. The causes of death for the remaining 6 patients are unknown. The 23 patients who died were enrolled into treatment groups as follows: 8 into Protocol 1, 4 into Protocol 2, 3 into Protocol 3, 7 into Protocol 4, and 1 into Protocol 5. Survival was significantly better with Protocol 3 than the other protocols, but did not differ significantly among the other protocols.

Of the 1323 possible measurement occasions, skin scores were recorded at 893 (67%). The mean number of followup visits per person did not differ among the different protocols.

Influence of baseline variables on mRSS. To assess the influence of baseline variables on mRSS, and hence their potential for confounding, linear regression analysis was carried out.

Those variables that were associated with baseline mRSS

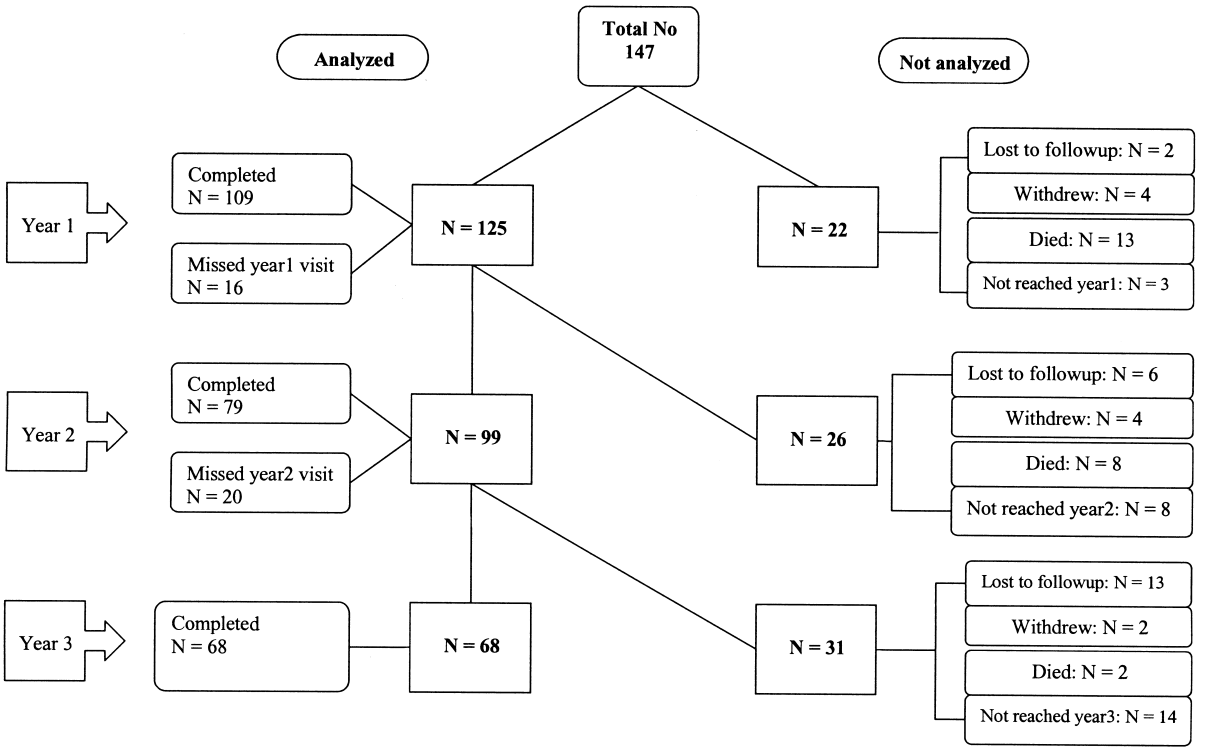

Figure 1. Completion of 3-year followup, by year.

Table 2. Effect of baseline confounders on baseline mRSS and change over time. Values are expressed as coefficient (95% CI).

Measurement	Effect on Skin Score at Baseline, n = 751	Effect on Skin Score Slope, n = 751
Age, per 10 years*	1.00 (0.07, 1.93)	–0.83 (–1.48, –0.18)
HAQ-DI score, per unit change	3.32 (1.67, 4.97)	–0.62 (–1.81, 0.57)
Functional Questionnaire, per unit change	0.31 (0.14, 0.49)	–0.08 (–0.20, 0.04)
Hemoglobin, per g/l	–0.122 (–0.202, –0.041)	0.028 (–0.0223, 0.079)
Platelets, per 10^9/l	0.022 (0.007, 0.036)	0.005 (–0.006, 0.016)
Plasma creatinine, per μmol/l	0.007 (–0.009, 0.023)	–0.015 (–0.022, –0.008)
Pulmonary hypertension	–0.05 (–3.97, 3.86)	–1.88 (–7.34, 3.59)
Pulmonary fibrosis	–0.66 (–3.99, 2.68)	–1.04 (–2.91, 0.82)
Cardiac involvement	0.18 (–4.30, 4.65)	–1.26 (–4.56, 2.04)

* Changes in skin score assessed per 10-year unit. HAQ-DI: Health Assessment Questionnaire-Disability Index.

were age, HAQ-DI, Functional Questionnaire score, hemoglobin and platelet count, while age and plasma creatinine were associated with change in mRSS over time (Table 2). Specifically, age was positively associated with mRSS score at baseline and mRSS reduced with age over time, i.e., older patients improving more quickly. Although baseline mRSS was increased by 3.32 per unit increase in HAQ-DI at baseline (95% CI: 1.67, 4.97) and by 0.31 per unit increase in Functional Questionnaire score (95% CI: 0.14, 0.49), there was no significant influence of the HAQ and Functional Questionnaire score on the skin scores over time, i.e., the changes in the skin score over time were similar in those with the higher and the lower HAQ and Functional Questionnaire scores (p = 0.30 for HAQ score, p = 0.18 for functional ability score).

There were no significant differences in the mRSS at baseline in those with higher and lower levels of plasma creatinine; however, the influence of the level of creatinine on the skin score over time was statistically significant, i.e., those with the greater levels of plasma creatinine had a greater reduction, although the difference was small (Table 2).

Those with low hemoglobin and high levels of platelets at baseline were more likely to have higher total skin scores at baseline, but the changes in the total skin score over time did not differ.

Changes in mRSS over time in the different treatment groups. The total skin score decreased over time from 24 (IQR 19–32) at baseline to 15.5 (IQR 9–24.5) at 3 years (Figure 2). There were differences in the magnitude of the change for different treatment protocols (Figure 3). There was a considerable reduction in the mRSS for protocols 2, 3, and 4 at the 3-year followup with a smaller decrease in Protocol 5 and some fluctuation in Protocol 1. When analysis was restricted to those 68 subjects who had completed

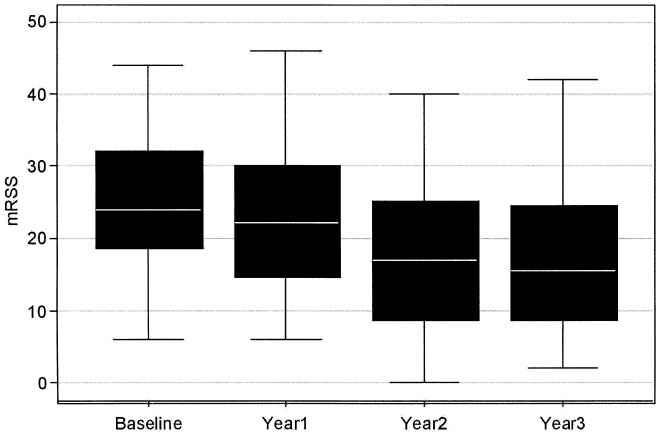

Figure 2. Modified Rodnan skin score (mRSS) during followup. Shaded box: interquartile range; light bar: median; whiskers: the most extreme observations.

the 3-year followup, the results changed very little. There was a very strong correlation between the change in the mRSS and the baseline skin score r = –0.56), but all comparisons between protocols were adjusted for baseline skin score.

Of the 58 patients for whom information from the final followup was available, 13 (22%) had a higher skin score at the end of the study than at the start. Of the 133 subjects with at least 1 measurement post-baseline, 29 (22%) had increased their skin score.

Linear regression was performed on the multiply imputed dataset in order to see whether the mRSS changed significantly over time for each protocol, and whether any changes differed significantly between treatment protocols. In these models, the outcome variable was mRSS and the independent variables were treatment protocols, time (in months), and their interaction. The regression was carried out for 4 protocols. Protocol 5 was excluded from the analysis because of the small number of participants on this type of treatment. The analysis revealed significant reductions in the mRSS over time for protocols 2, 3, and 4 but not for Protocol 1 (Table 3).

Although the skin score decreased significantly over time with all protocols except Protocol 1, there were no significant differences between protocols in the rate of change of the skin score with time (p = 0.43). When inverse probability weights were applied, the results remained nonsignificant (p = 0.41). Finally, when all subjects with weights more than 20 were removed from the analysis, to avoid giving inordinate influence to unusual subjects, the difference between protocols remained nonsignificant (p = 0.28). With the weighting, the improvement in skin score increased very slightly for Protocols 2–4, but decreased for Protocol 1.

Adverse effects. Sixteen (55%) of the patients in Protocol 1 reported adverse effects, 10 (40%) in Protocol 2, 27 (44%) in Protocol 3, 8 (42%) in Protocol 4, and 5 (39%) in Protocol 5. The adverse effects reported in Protocol 4 (no disease-modifying treatment) were due to patients changing protocol [mainly to Protocol 3 (MMF)].

DISCUSSION

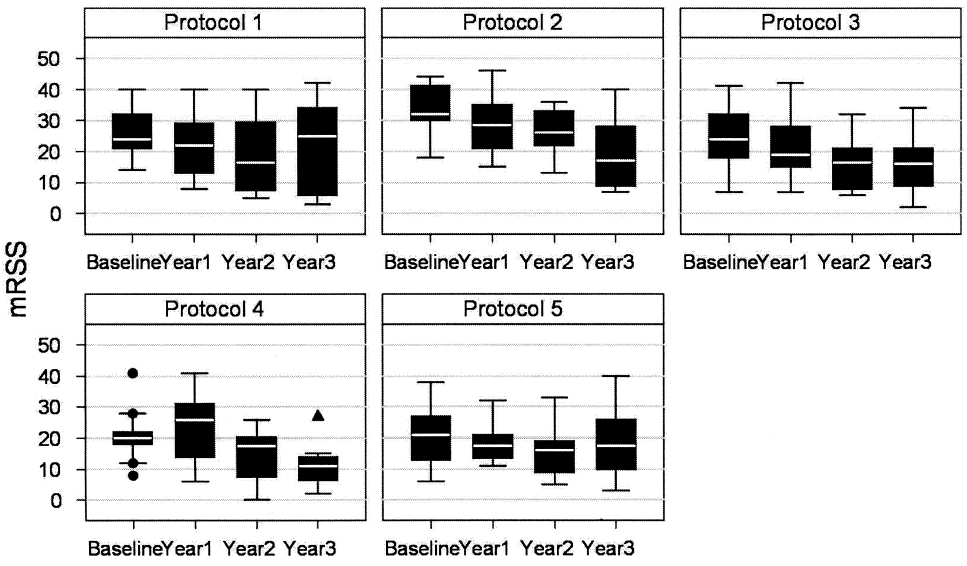

Figure 3. Modified Rodnan skin score (mRSS) during followup, by treatment protocol. Light bars show medians; shaded boxes show interquartile range. The whiskers show the most extreme observation less than 1.5 times the length of the shaded boxes. Dots show the individual observations beyond the whiskers. Protocol 1. IV cyclophosphamide followed by mycophenolate mofetil (MMF); Protocol 2. Antithymocyte globulin followed by MMF; Protocol 3. MMF alone; Protocol 4. No disease-modifying treatment; Protocol 5. Other immunosuppressant treatment.

Table 3. Changes in mRSS over time in the different treatment protocols. Values are expressed as coefficient (95% CI).

	Protocol 1, n = 29	Protocol 2, n = 25	Protocol 3, n = 61	Protocol 4, n = 19	p
Present dataset	–1.81 (–4.08, 0.460)	–4.46 (–6.69, –2.23)	–3.10 (–4.27, –1.93)	–2.86 (–4.61, –1.11)	0.43
Present dataset (all subjects with weights ≥ 20 removed)	n = 27 –0.786 (–3.55, 1.98)	n = 23 –4.61 (–7.57, –1.65)	n = 60 –3.49 (–4.76, –2.22)	n = 16 –2.91 (–4.48, –1.34)	0.28

mRSS: modified Rodnan skin score.

The study methodology permitted an evaluation of the relative effectiveness of different treatment regimens based on standard data collection in otherwise routine clinical practice. Such an approach has recently been advocated by the head of the UK's drug approval body, NICE (National Institute for Clinical Excellence, an organization committed to basing drug approvals on the highest standards of evidence), as an appropriate substitute for randomized trials in rare disorders such as SSc, for which there is a serious lack of data from randomized controlled trials[25]. Overall, there was a reduction in skin score with all of the protocols considered (although nonsignificant in Protocol 1), but the differences between protocols were not statistically significant. Adjusting for possible confounding by weighting had very little effect on the estimates, and the differences between protocols remained nonsignificant. This is understandable, given that the data suggest that none of the baseline variables had a very strong effect on the skin score, and hence the variables would produce only a modest confounding effect (Table 2).

The only noticeable change with weighting was a reduction in the effect of Protocol 1. Again, this is understandable since pulmonary hypertension, pulmonary fibrosis, and cardiac problems were overrepresented in this protocol group, and these variables were associated with larger (but not significantly larger) reductions in skin score over time.

There are a number of limitations to the study reflecting both its nonrandomized nature and the fact that it was embedded into routine clinical practice, with the real-world consequences of missing data and some loss to followup. First, there is the possibility of residual confounding. There may have been patient and/or disease-related variables other than those that were measured and adjusted for in the analysis that both influenced the choice of therapy and independently influenced outcome, and whose effects were also independent of the variables analyzed. For example, all patients treated with ATG were from the same center, so it may be that the apparent effect of ATG treatment contained a center-specific effect.

Second, subjects could change protocols during the study. Several patients changed protocol during the first 12 months of the study and the reasons for changing may have been related to the skin score. Also, patients continued on their initial treatment for variable lengths of time. Thus our analysis is effectively an intention-to-treat analysis, rather than a measure of the effect of a given protocol over 3 years. In principle, it is possible to measure the effect of receiving treatment using marginal structural models, but this requires measuring all potential confounders at the time that the decision to change treatment is made. We did not have this information. Therefore our conclusions relate to starting choice. In particular, many subjects on Protocol 4 changed to active treatment during the study, so the measured outcome in this group cannot be taken as the effect of no treatment.

There is also a potential bias effect due to subjects not completing the 3 years of followup. We minimized this by including all available skin scores in the linear regression model, so that even subjects who did not complete 3 years contributed data until they dropped out. There was no evidence that the dropout rate differed among protocols, so this bias is likely to have had only minor impact.

The choice of treatments was extensively discussed prior to commencement of the study. In a study of this design, there are inevitably a large number of possible protocols. It would be appropriate to include MTX in future observational studies, in view of the reported trend toward improvement in skin score[11] and recent recommendations from the European League Against Rheumatism Scleroderma Trials and Research group, supporting its use in dcSSc for skin fibrosis[26]. Although there has so far been only minimal experience with biologic treatments in SSc, this experience has been disappointing and therefore more "standard" immunosuppressants including MTX merit further study.

The findings of our study in terms of falls in skin score were comparable to those of other studies. In our study, skin score across all 147 patients fell from 24 at baseline to 15.5 at 3 years, a fall of 8.5. This compares to mean skin score falls over 24 months in the high- vs low-dose penicillamine study of 4.8 from 20.4 in the high-dose group, and 6.9 from 19.9 in the low-dose group[9]. ATG followed by MMF was associated with a very marked fall in skin score. The ATG/MMF group had a very high baseline mRSS (score = 32) and all patients came from the same center. Treatment with ATG is included in some cell transplantation protocols, but has otherwise not been further studied in patients with SSc.

Since our observational study began, 2 studies have reported benefit from MMF in diffuse SSc (although neither

was a controlled trial). Liossis, *et al* in a small open study of 5 patients with diffuse SSc[27] reported that MMF and small doses of prednisolone conferred benefit in SSc-related alveolitis. A large retrospective study including 109 patients treated with MMF reported that MMF was well tolerated and that patients had a 5-year survival rate of 95.4% from disease onset[28]. The authors concluded that MMF was at least as effective as other therapies although the retrospective cohort design has many drawbacks compared to the present study that includes the same center. Twenty-seven of the patients recruited toward the end of the retrospective study[28] were also included in the present study. Data collection and analysis of the 2 studies were independent of each other.

Regarding cyclophosphamide, the Scleroderma Lung Study of 158 patients with SSc-related lung disease[3] reported that in addition to conferring a modest benefit in lung function, cyclophosphamide was associated with a fall in skin score and an improvement in the HAQ.

While having several limitations, our study does have the advantage that a large number of patients were studied with different severities of disease, and so the patient cohort is more likely to be representative of those patients encountered in clinical practice than those included in controlled clinical trials. Our findings do not allow us to make any new conclusions regarding recommendations for treatment of dcSSc, other than that there were no important differences between the different immunosuppressive regimens studied. It is difficult to make any conclusion about the relative merits of immunosuppression vs no immunosuppression because of the small number of patients in the "no treatment" group, many of whom started on immunosuppression during the study period. Nor is it possible to make any conclusions about side effects of the different treatments, as these were not recorded in detail unless they led to withdrawal from the study. Our study confirmed the high mortality in patients with early diffuse disease.

A number of indirect benefits were gained from this study. We consider that it raised awareness among UK rheumatologists about the importance of identifying patients with dcSSc early, and those recruited were carefully monitored in line with best clinical practice. Perhaps the most important conclusion is that although clinical trials, with near-complete data capture, at present remain the gold standard for evaluating efficacy and safety, the longterm feasibility of prospective cohort studies such as this one offer promise in the future evaluation of therapies in clinical practice. Clinicians and scientists must work together to develop and test new therapies for early diffuse disease.

ACKNOWLEDGMENTS

We are grateful to all those clinicians who notified patients for the study, and to Sue Brown, Rachel Ochiel, Rachel Vincent, Helen Wilson, Elizabeth Wragg, Sandra Zimba, and all other nurses who assisted in data collection.

REFERENCES

1. LeRoy EC, Black C, Fleischmajer R, Jablonska S, Kreig T, Medsger TA, et al. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. J Rheumatol 1988;15:202-5.
2. Denton CP, Black CM. Scleroderma – clinical and pathological advances. Best Pract Res Clin Rheumatol 2004;18:271-90.
3. Tashkin DP, Elashoff R, Clements PJ, Goldin J, Roth MD, Furst DE, et al. Cyclophosphamide versus placebo in scleroderma lung disease. N Engl J Med 2006;354:2655-66.
4. Hoyles RK, Ellis RW, Wellsbury J, Lees B, Newlands P, Roberts C, et al. A multicenter, prospective, randomized, double-blind, placebo-controlled trial of corticosteroids and intravenous cyclophosphamide followed by oral azathioprine for the treatment of pulmonary fibrosis in scleroderma. Arthritis Rheum 2006;54:3962-70.
5. Jimenez SA, Derk CT. Following the molecular pathways toward an understanding of the pathogenesis of systemic sclerosis. Ann Intern Med 2004;140:37-50.
6. Charles C, Clements P, Furst DE. Systemic sclerosis: hypothesis-driven treatment strategies. Lancet 2006;367:1683-91.
7. Denton CP, Black CM. Scleroderma and related disorders: therapeutic aspects. Balliere's Clin Rheumatol 2000;14:17-35.
8. Black CM, Silman AJ, Herrick AL, Denton CP, Wilson H, Newman J, et al. Interferon-alpha does not improve outcome at one year in patients with diffuse cutaneous scleroderma: results of a randomized double-blind placebo-controlled trial. Arthritis Rheum 1999;42:299-305.
9. Clements PJ, Furst DE, Wong WK, 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.
10. Denton CP, Merkel PA, Furst DE, Khanna D, Emery P, Hsu VM, et al. Recombinant human anti-transforming growth factor ß1 antibody therapy in systemic sclerosis: a multicenter, randomized, placebo-controlled phase I/II trial of CAT-192. Arthritis Rheum 2007;56:323-33.
11. Pope JE, Bellamy N, Seibold JR, Baron M, Ellman M, Carette S, et al. A randomized, controlled trial of methotrexate versus placebo in early diffuse scleroderma. Arthritis Rheum 2001;44:1351-8.
12. Vonk MC, Marjanovic Z, van den Hoogen FH, Zohar S, Schattenberg AV, Fibbe WE, et al. Long-term follow-up results after autologous haematopoietic stem cell transplantation for severe systemic sclerosis. Ann Rheum Dis 2008;67:98-104.
13. Nash RA, McSweeney PA, Crofford LJ, Abidi M, Chen C, Godwin JD, et al. High-dose immunosuppressive therapy and autologous hematopoetic cell transplantation for severe systemic sclerosis: long-term follow-up of the US multicenter pilot study. Blood 2007;110:1388-96.
14. Stratton RJ, Wilson H, Black CM. Pilot study of anti-thymocyte globulin plus mycophenolate mofetil in recent-onset diffuse scleroderma. Rheumatol 2001;40:84-8.
15. Masi AT, Rodnan GP, Medsger TA, Altman RD, D'Angelo WA, Fries JF, et al. Preliminary criteria for the classification of systemic sclerosis (scleroderma). Arthritis Rheum 1980;23:581-90.
16. Steen VD, Medsger TA. The value of the health assessment questionnaire and special patient-generated scales to demonstrate change in systemic sclerosis over time. Arthritis Rheum 1997;40:1984-91.
17. Silman A, Akesson A, Newman J, Henriksson H, Sandquist G, Nihill M, et al. Assessment of functional ability in patients with scleroderma: a proposed new disability assessment instrument. J Rheumatol 1998;25:79-83.
18. Clements P, Lachenbruch P, Siebold J, White B, Weiner S, Martin R, et al. Inter and intraobserver variability of total skin thickness score (modified Rodnan TSS) in systemic sclerosis. J Rheumatol

1995;22:1281-5.
19. Sato T, Matsuyama Y. Marginal structural models as a tool for standardization. Epidemiology 2003;14:680-6.
20. Imbens GW. The role of the propensity score in estimating dose-response functions. Biometrica 2000:87:706-10.
21. Glynn RJ, Schneeweiss S, Sturmer T. Indications for propensity scores and review of their use in pharmacoepidemiology. Basic Clin Pharmacol Toxicol 2006;98:253-9.
22. Royston P. Multiple imputation of missing values: update of ice. Stata J 2005;5:527-36.
23. van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. Stat Med 1999;18:681-94.
24. Rubin DB. Multiple imputation for non-response in surveys. New York: J. Wiley and Sons; 1987.
25. Rawlins MD. The Harveian Oration of 2008: On the evidence for decisions about the use of therapeutic interventions. London: Royal College of Physicians; 2008.
26. Avouac J, Kowal-Bielecka O, Landewé RB, Chwiesko S, Miniati I, Czirják L, et al. European League Against Rheumatism (EULAR) Scleroderma Trial and Research group (EUSTAR) recommendations for the treatment of systemic sclerosis: Methods of elaboration and results of systematic literature research. Ann Rheum Dis 2009;68:629-34.
27. Liossis SNC, Bounas A, Andonopoulos AP. Mycophenolate mofetil as first-line treatment improves clinically early scleroderma lung disease. Rheumatology 2006;45:1005-8.
28. Nihtyanova SI, Brough GM, Black CM, Denton CP. Mycophenolate mofetil in diffuse cutaneous systemic sclerosis – a retrospective analysis. Rheumatology 2007;46:442-5.