EXTENDED REPORT

Treatment outcome in early diffuse cutaneous systemic sclerosis: the European Scleroderma Observational Study (ESOS)

Ariane L Herrick,[1,2] Xiaoyan Pan,[3] Sébastien Peytrignet,[3] Mark Lunt,[3] Roger Hesselstrand,[4] Luc Mouthon,[5] Alan Silman,[6] Edith Brown,[7] László Czirják,[8] Jörg H W Distler,[9] Oliver Distler,[10] Kim Fligelstone,[11] William J Gregory,[12] Rachel Ochiel,[11] Madelon Vonk,[13] Codrina Ancuța,[14] Voon H Ong,[15] Dominique Farge,[16] Marie Hudson,[17] Marco Matucci-Cerinic,[18] Alexandra Balbir-Gurman,[19] Øyvind Midtvedt,[20] Alison C Jordan,[21] Paresh Jobanputra,[21] Wendy Stevens,[22] Pia Moinzadeh,[23] Frances C Hall,[24] Christian Agard,[25] Marina E Anderson,[26] Elisabeth Diot,[27] Rajan Madhok,[28] Mohammed Akil,[29] Maya H Buch,[30] Lorinda Chung,[31] Nemanja Damjanov,[32] Harsha Gunawardena,[33] Peter Lanyon,[34] Yasmeen Ahmad,[35] Kuntal Chakravarty,[36] Søren Jacobsen,[37] Alexander J MacGregor,[38] Neil McHugh,[39] Ulf Müller-Ladner,[40] Gabriela Riemekasten,[41] Michael Becker,[42] Janet Roddy,[43] Patricia E Carreira,[44] Anne Laure Fauchais,[45] Eric Hachulla,[46] Jennifer Hamilton,[47] Murat İnanç,[48] John S McLaren,[49] Jacob M van Laar,[50] Sanjay Pathare,[51] Susannah Proudman,[52] Anna Rudin,[53] Joanne Sahhar,[54] Brigitte Coppere,[55] Christine Serratrice,[56] Tom Sheeran,[57] Douglas J Veale,[58] Claire Grange,[59] Georges-Selim Trad,[60] Christopher P Denton[15]

Handling editor Tore K Kvien

▶ Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/annrheumdis-2016-210503)

For numbered affiliations see end of article.

Correspondence to
Professor Ariane Herrick, The University of Manchester, Salford Royal NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester M13 9PT, UK; ariane.herrick@manchester.ac.uk

ABSTRACT

Objectives The rarity of early diffuse cutaneous systemic sclerosis (dcSSc) makes randomised controlled trials very difficult. We aimed to use an observational approach to compare effectiveness of currently used treatment approaches.

Methods This was a prospective, observational cohort study of early dcSSc (within three years of onset of skin thickening). Clinicians selected one of four protocols for each patient: methotrexate, mycophenolate mofetil (MMF), cyclophosphamide or 'no immunosuppressant'. Patients were assessed three-monthly for up to 24 months. The primary outcome was the change in modified Rodnan skin score (mRSS). Confounding by indication at baseline was accounted for using inverse probability of treatment (IPT) weights. As a secondary outcome, an IPT-weighted Cox model was used to test for differences in survival.

Results Of 326 patients recruited from 50 centres, 65 were prescribed methotrexate, 118 MMF, 87 cyclophosphamide and 56 no immunosuppressant. 276 (84.7%) patients completed 12 and 234 (71.7%) 24 months follow-up (or reached last visit date). There were statistically significant reductions in mRSS at 12 months in all groups: −4.0 (−5.2 to −2.7) units for methotrexate, −4.1 (−5.3 to −2.9) for MMF, −3.3 (−4.9 to −1.7) for cyclophosphamide and −2.2 (−4.0 to −0.3) for no immunosuppressant (p value for between-group differences=0.346). There were no statistically significant differences in survival between protocols before (p=0.389) or after weighting (p=0.440), but survival was poorest in the no immunosuppressant group (84.0%) at 24 months.

Conclusions These findings may support using immunosuppressants for early dcSSc but suggest that overall benefit is modest over 12 months and that better treatments are needed.

Trial registration number NCT02339441.

INTRODUCTION

The diffuse cutaneous subtype of systemic sclerosis (dcSSc) is rare (SSc incidence is around 10–20/million/year,[1] of whom approximately 25% will have diffuse disease) but carries high morbidity and mortality due to early internal organ involvement and rapidly progressive, painful skin thickening. Also, 5-year and 10-year survival rates, although improving, are in the order of 68% and 50%, respectively.[2 3]

At present, there is no drug known to favourably influence disease course. Randomised controlled trials (RCTs) have historically been confounded by disease rarity (only small numbers of patients are recruited, often over long periods) and strict entry criteria meaning that severe cases are often excluded.[4] These strict criteria further restrict sample sizes and limit generalisability. Therefore, although RCTs represent a gold standard for assessing drug efficacy, results may not be applicable to

real-life clinical settings.[5] Small trials run the risk of being underpowered, thus potentially yielding false-negative results.[6] The past three decades have seen a number of promising treatments for early dcSSc failing to meet efficacy end points in RCTs: examples include methotrexate (multinational, 71 patients)[7] and anti-transforming growth factor β1 antibody therapy (multinational, 45 patients).[8]

A further difficulty in recruiting into RCTs of early dcSSc is that many clinicians have reservations about placebo therapy in a potentially life-threatening disease and favour immunosuppression, consistent with the European League Against Rheumatism (EULAR) recommendations, which advocate methotrexate for skin manifestations[9] in early dcSSc, although this agent has been shown to be of only limited efficacy.[7] Immunosuppressants are potentially hazardous, especially in patients prone to internal organ disease and infection.

Against this background, our aim was to compare, using an observational approach, the effectiveness of standard treatment approaches (mainly immunosuppressant treatments but including a 'no immunosuppressant' option to reflect that some patients or clinicians may choose this approach) in the early management of patients with dcSSc, capturing entry and outcome data in a systematic way. Modern statistical approaches allow robust interrogations of prospective observational studies, as an adjunct to, or even substitute for, RCTs in rare diseases,[10] although the potential of these novel approaches has not yet been realised.[11]

METHODS
Study design
The European Scleroderma Observational Study (ESOS) was a prospective, observational cohort study (ClinicalTrials.gov identifier: NCT02339441), in which standardised data were collected at study entry and at follow-up visits, and entered electronically by investigators at each centre into an electronic case record form. All data were checked by the project coordinator and any inconsistencies were discussed with the chief investigator and (if appropriate) the local principal investigator. The main inclusion criteria were early dcSSc (skin involvement proximal to elbow, knee, face, neck[12] and within three years of the onset of skin thickening) and age >18 years. Exclusion criteria were previous stem cell transplantation, previous immunosuppressant treatment for >4 months or use of any immunosuppressant drug other than methotrexate, mycophenolate mofetil (MMF) or cyclophosphamide within the month prior to study entry.

Clinicians selected the protocol of their choice for each patient. The recommended treatment protocols, as decided by the Steering Committee to reflect international best clinical practice, were

1. Methotrexate (oral or subcutaneous with a target dose of 20–25 mg weekly).
2. MMF (500 mg twice daily for 2 weeks increasing to 1 g twice daily).
3. Cyclophosphamide.
   Possible regimens included:
   i. Intravenous. Minimum monthly dose 500 mg/m^2 with a recommended duration of 6–12 months.
   ii. Oral. 1–2 mg/kg/day with a recommended duration of 12 months. Patients treated with cyclophosphamide were then usually 'transferred' to a maintenance immunosuppressive drug (methotrexate, MMF or azathioprine) as per the treating clinician's choice.

4. No immunosuppressant treatment, to give the option of including patients in whom immunosuppression was not felt indicated or appropriate (or declined by the patient).

Patients were assessed at baseline, with subsequent visits scheduled three-monthly for 24 months (or between 12 and 24 months for those patients recruited after September 2013).

To have 80% power to detect a difference between two treatment arms of five modified Rodnan skin score (mRSS) units at 12 months would require 63 patients per protocol. Allowing 20% loss to follow-up, and varying numbers recruited to the different protocols, recruitment target was 316 patients.

Patients
Patients were recruited between July 2010 and September 2014. Demographic characteristics including age, gender, smoking habit, ethnicity, antibody status (anti-topoisomerase-1 (anti-Scl70), anti-RNA III polymerase, anticentromere) and presence of visceral organ involvement were recorded for all patients. The algorithms to determine the presence of different types of organ involvement are summarised in online supplementary table S1.

Outcome measures
The primary outcome measure, assessed at each visit, was the change in mRSS over time. All mRSS assessments were performed by those experienced in skin scoring. The mRSS is assessed clinically at 17 body sites on a 0–3 scale (maximum score 51) and measures the extent of skin thickening.[13] It is the most commonly used primary outcome measure in RCTs of dcSSc,[4 7 8] reflecting disease severity and predicting mortality.[14] All other outcomes/recorded variables were mainly part of routine clinical practice and are summarised in online supplementary table S2. Secondary end points included pulmonary function (forced vital capacity (FVC: % predicted) and carbon monoxide diffusing capacity (DLCO: % predicted)), quality of life[15–18] (including the Health Assessment Questionnaire Disability Index (HAQ-DI)[15] and Cochin Hand Function Scale[18]), occurrence of side effects and survival.

Statistical analysis
In an observational study, patient characteristics differ between groups and any differences in outcomes might be driven by those characteristics rather than the treatments (confounding by indication). In each of the analyses (for the different outcome measures), all variables associated with the outcome were considered as confounders.[19 20]

Differences between protocols at baseline
Kruskal-Wallis test was applied for continuous variables and Fisher's test for categorical variables.

Influence of baseline characteristics on mRSS at baseline and over time
The association between baseline variables and mRSS was assessed by simple linear regressions, entering each characteristic separately as a predictor of mRSS. To examine how each variable affected the progression of mRSS, the regression equation was modified by adding a term for time and its interaction with the baseline predictor value.

Differences in the changes between groups for all outcomes
Inverse probability of treatment (IPT) weights equalise the distributions of confounders between the treatment groups, thus removing confounding by indication.[21] Treatment probabilities

were computed using multinomial logistic regressions, with the baseline values of the selected confounders as predictors.[22] Censoring weights rebalance the data such that the distributions of confounders remain unchanged throughout the study. For each observation, the probability of remaining uncensored given the baseline values of the confounders, the initial protocol and a cubic spline for time was calculated using a pooled logistic regression model.[23] Multiplying both weights yielded the IPT and inverse probability of censoring (IPTC) weights. Weights >20 were truncated at that value.[24]

Treatment effects were assessed using IPTC-weighted linear regression models, which include an intercept, a time term, indicator variables for treatment groups and interactions between time and treatments. The model followed an intention-to-treat approach. Differences in the interaction terms reflected differences in the evolution of outcome.

Cochin hand function data were log-transformed (after adding one to each value) to correct for a highly left-skewed distribution. CIs for the difference of logs were back-transformed, yielding a percentage difference between predicted baseline and 12-month levels.

Because of missing data at baseline for confounders, multiple imputation by chained equations was applied with STATA V.13.1. Imputations were performed separately for each different outcome model. Moreover, each analysis was restricted to the subset of patients with available outcome data at baseline.

Survival analysis
Kaplan-Meier curves, adjusted using IPT weights, provide estimates of the cumulative probability of surviving in each of the protocols. An IPT-weighted Cox regression, including indicator variables for the protocols, was used to test for differences in survival between protocols. Both overall and adverse event-free survival were examined.

RESULTS
In total, 326 patients from 50 centres (19 countries) were recruited into the study (figure 1): 160 from mainland Europe and the Middle East, 134 from the UK, 15 from Australia and 17 from North America (six centres from Australia and North America joined after the initial recruitment wave). Not being a randomised study, the number of patients starting on each protocol differed: 65 (19.9%) methotrexate, 118 (36.2%) MMF, 87 (26.7%) cyclophosphamide and 56 (17.2%) no immunosuppressant treatment. Median (IQR) doses are shown in online supplementary table S3.

Baseline characteristics of patients
The median mRSS (21, IQR 16–27) and its distribution did not differ across all four treatment groups (p=0.306) (table 1). There were significant differences between treatment groups in gender (patients in the cyclophosphamide group less likely to be female, p=0.003) and duration of skin thickening (the 'no immunosuppressant' group had the longest, p=0.001). Also, patients in the cyclophosphamide group were more likely to have had previous immunosuppression (p=0.007) or steroid treatment (p=0.001). At baseline, 94 (28.8%) patients were taking oral corticosteroids, with a median dose of 10 mg/day (range 2.5–60 mg/day).

Organ involvement
There were significant differences between groups for presence of pulmonary fibrosis, cardiac, renal and muscle involvement. Patients on cyclophosphamide were more likely to have

pulmonary fibrosis (p=0.036 across groups) or cardiac involvement (p=0.009 across groups). Patients in the 'no immunosuppressant' group were more likely to have renal involvement (p=0.039), and the methotrexate group had more frequent muscle involvement (p=0.002).

Functional ability
Scores for the HAQ-DI, Functional Assessment of Chronic Illness Therapy (FACIT) fatigue and Short-Form 36 (SF36) physical and mental indexes did not differ significantly between groups. However, there were significant differences across groups in the cochin hand function scale (CHFS), which was poorest in the cyclophosphamide group (p=0.025).

Concomitant medications
As anticipated in a study of patients with early dcSSc, there was substantial use of concomitant medications (see online supplementary table S4).

Progression through the study
Figure 1 shows how patients progressed through the study. Overall, 276 patients (84.7%) remained in the study at 12 months of follow-up and 234 (71.7%) completed 24 months (or reached the last study visit date of 30 September 2015).

Changes in protocol
A total of 60 (18.4%), 12 (3.7%) and 1 (0.3%) patients changed protocol one, two or three times during the study. Among patients still in the study, adherence to initial protocol at 24 months for the different cohorts was 76.2% (methotrexate), 79.7% (MMF), 79.2% (cyclophosphamide) and 73.3% (no immunosuppressant) (see online supplementary figure S1). In the no immunosuppressant cohort, 10 out of 56 patients commenced an immunosuppressant (figure 1).

Withdrawals and deaths
In total, 35 patients (10.7%) died and 42 (12.9%) withdrew from the study (including lost to follow-up). Of the 35 deceased patients, 31 cases were primarily attributed to SSc-related causes (26 most likely primarily cardiorespiratory, 2 renal crises, 2 gastrointestinal (one aspiration) and 1 peritonitis (on peritoneal dialysis following renal crisis)), 3 died of cancer (1 nasopharyngeal, 1 rectal, 1 colorectal) and in 1 case the cause was unknown.
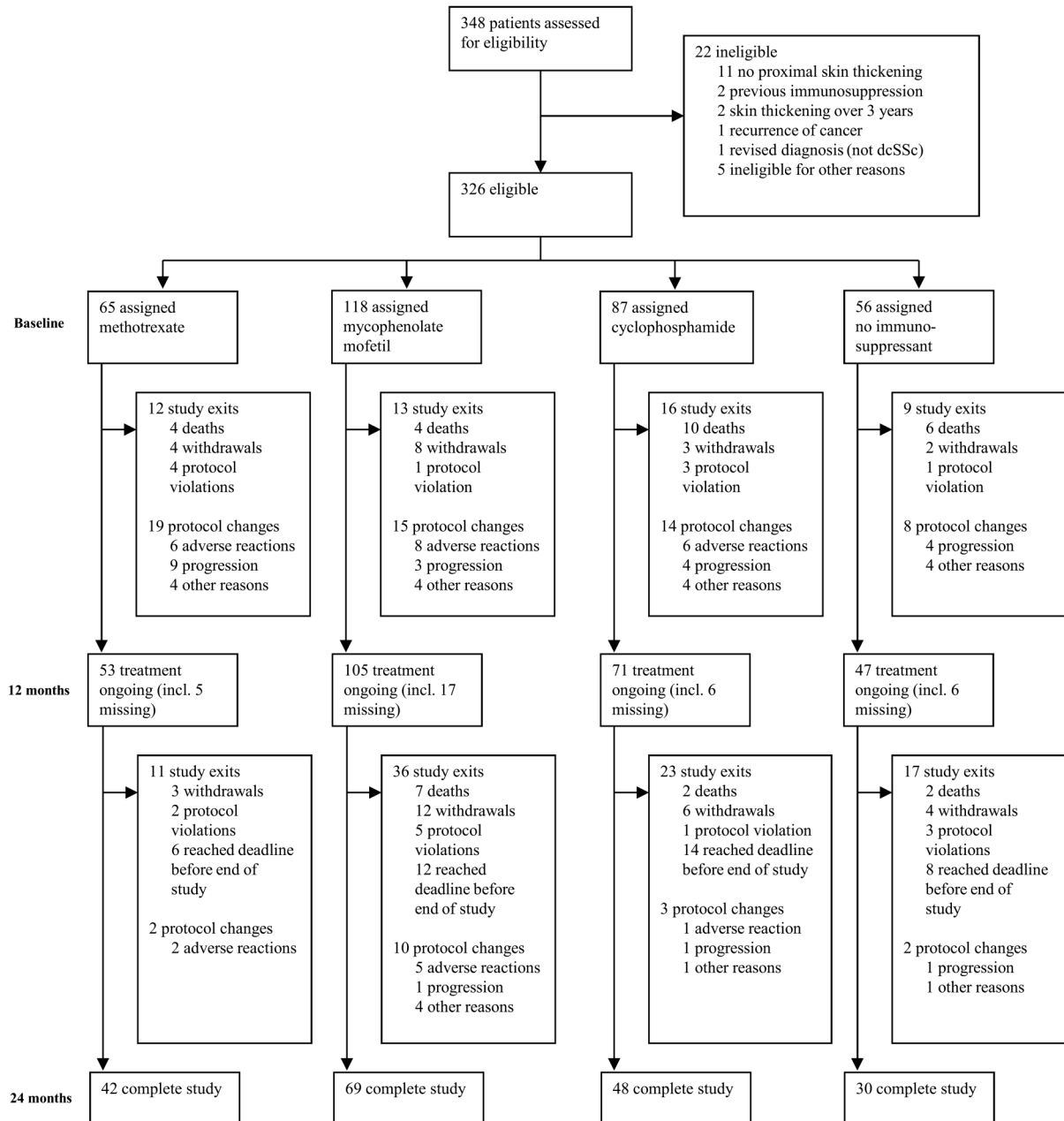
Influence of baseline variables on the initial skin score and on skin score trajectory
Table 2 summarises the effect of different characteristics on the initial mRSS and its subsequent trajectory, as analysed with linear regression.

Using the associations described by table 2, the confounders identified for the skin score were age, duration of skin thickening, current or previous steroid use, anti-topoisomerase, anti-RNA polymerase III, pulmonary fibrosis, pulmonary hypertension, cardiac, renal and muscle involvement, as well as HAQ-DI, Cochin hand function and FACIT fatigue scores (see online supplementary table S5 for lists of confounders and online supplementary tables S6– S13 for each model's confounder selection process).

Changes in skin score over time in the different treatment groups
The mean change in mRSS after 12 and 24 months was −2.9 and −6.7 units. Based on a weighted regression model, there were

348 patients assessed for eligibility

22 ineligible
11 no proximal skin thickening
2 previous immunosuppression
2 skin thickening over 3 years
1 recurrence of cancer
1 revised diagnosis (not dcSSc)
5 ineligible for other reasons

326 eligible

Baseline

65 assigned methotrexate

118 assigned mycophenolate mofetil

87 assigned cyclophosphamide

56 assigned no immuno-suppressant

12 study exits
4 deaths
4 withdrawals
4 protocol violations

19 protocol changes
6 adverse reactions
9 progression
4 other reasons

13 study exits
4 deaths
8 withdrawals
1 protocol violation

15 protocol changes
8 adverse reactions
3 progression
4 other reasons

16 study exits
10 deaths
3 withdrawals
3 protocol violation

14 protocol changes
6 adverse reactions
4 progression
4 other reasons

9 study exits
6 deaths
2 withdrawals
1 protocol violation

8 protocol changes
4 progression
4 other reasons

12 months

53 treatment ongoing (incl. 5 missing)

105 treatment ongoing (incl. 17 missing)

71 treatment ongoing (incl. 6 missing)

47 treatment ongoing (incl. 6 missing)

11 study exits
3 withdrawals
2 protocol violations
6 reached deadline before end of study

2 protocol changes
2 adverse reactions

36 study exits
7 deaths
12 withdrawals
5 protocol violations
12 reached deadline before end of study

10 protocol changes
5 adverse reactions
1 progression
4 other reasons

23 study exits
2 deaths
6 withdrawals
1 protocol violation
14 reached deadline before end of study

3 protocol changes
1 adverse reaction
1 progression
1 other reasons

17 study exits
2 deaths
4 withdrawals
3 protocol violations
8 reached deadline before end of study

2 protocol changes
1 progression
1 other reasons

24 months

42 complete study

69 complete study

48 complete study

30 complete study

(i) Withdrawals include patients who left the study for the following reasons: unable to travel, changed hospital, enrolled in a randomized clinical trial, no longer willing to participate or lost to follow-up with no given reason. Among the 8 patients who were lost to follow-up in the first year, 5 patients were due for their 24-month visit after the study deadline.

(ii) Protocol violations brought about a premature exit from the study. There were 20 protocol violations mid-study: 9 who started rituximab (as their main treatment), 5 who received a stem cell transplantation, 4 who started azathioprine (not as part of a post-cyclophosphamide transfer) , 1 who started tocilizumab and 1 who entered a randomized clinical trial.

(iii) Protocol changes designate patients who changed immunosuppressants, started taking or stopped receiving one (in non-observance to their initial assigned protocol). Changes that were a part of a post-cyclophosphamide transfer were not considered protocol changes. If a patient deviated from her/his initial protocol more than once, only the initial change (along with its reason) was recorded in the flow-chat.

(iv) There were 58 post-cyclophosphamide transfers: 40 patients went on to mycophenolate mofetil, 9 on to azathioprine, 5 on to methotrexate and 4 on to no immunosuppressant.

(v) In the no immunosuppressant cohort, 10 out of 56 patients deviated from their protocol at any point: 2 switched to methotrexate, 5 to mycophenolate mofetil and 3 to cyclophosphamide. 7 of those 10 cases were linked to pulmonary complications, one to associated inflammatory arthritis, one to skin thickening and one because of preference.

Figure 1 Progression of patients through the study.

Table 1 Baseline characteristics and differences between protocols

Characteristic	Protocol 1 Methotrexate n=65	Protocol 2 Mycophenolate mofetil, n=118	Protocol 3 Cyclophosphamide n=87	Protocol 4 No immunosuppressant n=56	P Value*	Total n=326	Missing at baseline, no. (%)
Age, years	52.5 (41.5–61.1)	50 (40.7–61.1)	54.1 (45.7–60.5)	52.3 (42.3–60.3)	0.324	52.3 (43–60.8)	0 (0%)
Female, no. (%)	50 (76.9%)	94 (79.7%)	49 (56.3%)	40 (71.4%)	0.003	233 (71.5%)	0 (0%)
Months since onset of skin thickening	10.3 (5.5–20.8)	12.6 (7.8–21.8)	10.2 (5.9–14.5)	16.5 (8.7–27)	0.001	11.9 (7–21)	18 (5.5%)
Years since onset of first non-Raynaud's manifestation	1.3 (0.7–2.1)	1.4 (0.9–2)	1.0 (0.6–1.4)	1.7 (0.9–2.5)	0.001	1.2 (0.8–2)	6 (1.8%)
Years since onset of Raynaud's phenomenon	1.3 (0.7–2.5)	1.9 (1.1–3.1)	1.4 (0.9–2.4)	2.2 (1.1–3.7)	0.014	1.7 (1–2.9)	22 (6.7%)
Previous immunosuppressant use, no. (%)†	3 (4.6%)	5 (4.2%)	15 (17.2%)	3 (5.4%)	0.007	26 (8%)	0 (0%)
Current or previous steroid use, no. (%)	33 (50.8%)	39 (34.2%)	51 (58.6%)	18 (32.1%)	0.001	141 (43.8%)	4 (1.2%)
Current smoker (%)	11 (16.9%)	17 (15.3%)	18 (21.7%)	11 (20.8%)	0.646	57 (18.3%)	14 (4.3%)
History of cancer, no. (%)	5 (7.8%)	4 (3.4%)	4 (4.7%)	7 (12.5%)	0.121	20 (6.2%)	4 (1.2%)
Caucasian, no. (%)	54 (83.1%)	92 (78.0%)	77 (88.5%)	49 (87.5%)	0.201	272 (83.4%)	0 (0%)
mRSS (0–51)	21 (17–24)	21 (16–27)	22 (17–29)	20 (15.5–26)	0.306	21 (16–27)	0 (0%)
Haemoglobin (g/L)	127 (118–136)	126 (118–137)	130 (116–140)	130 (118–139)	0.721	128 (118–137)	13 (4%)
White blood count (×10⁹/L)	7.7 (6.7–9.4)	7.9 (6.4–9.5)	8.9 (7.3–10.6)	8.0 (6.7–9.4)	0.029	8 (6.8–9.9)	14 (4.3%)
Platelets (×10⁹/L)	300 (254–337)	298 (246–370)	309 (259–359)	281 (250–337)	0.459	298 (253–358)	15 (4.6%)
ESR (mm/hour)	17 (10–29)	18 (8–30)	21 (9–48)	20 (10–35)	0.341	18 (8–34)	77 (23.6%)
CRP (mg/L)	4.0 (2.2–7.5)	5.0 (1.1–11)	5.9 (3.2–20.0)	4.8 (3.0–11.7)	0.026	5 (2.1–11.8)	90 (27.6%)
Anti-topoisomerase (anti-Scl70), no. (%)	20 (31.3%)	49 (42.6%)	39 (45.3%)	18 (33.3%)	0.228	126 (39.5%)	7 (2.1%)
Anti-RNA polymerase III, no. (%)	9 (18.8%)	23 (23.2%)	9 (13.0%)	9 (19.6%)	0.433	50 (19.1%)	64 (19.6%)
Anticentromere, no. (%)	9 (14.1%)	6 (5.4%)	4 (4.7%)	3 (5.7%)	0.147	22 (7%)	12 (3.7%)
Organ involvement							
Pulmonary fibrosis, no. (%)	8 (12.3%)	13 (11%)	21 (24.1%)	5 (8.9%)	0.036	47 (14.4%)	0 (0%)
FVC (% predicted)	93.7 (81–106)	90 (75–102)	82.5 (68.5–96.5)	90 (75–100)	0.026	89 (75–102)	19 (5.8%)
DLCO (% predicted)	73 (62–83)	64.5 (48–77)	57 (41–73)	64 (52–75)	<0.0005	64 (50–78)	35 (10.7%)
Pulmonary hypertension, no. (%)	4 (6.3%)	7 (5.9%)	10 (11.5%)	5 (8.9%)	0.488	26 (8%)	1 (0.3%)
sPAP or RVSP (mm Hg)‡	27 (23–32)	27 (20–33)	30 (23–37)	27 (23–35)	0.472	29 (21–34)	124 (38%)
Cardiac involvement, no. (%)	5 (7.7%)	8 (6.8%)	19 (21.8%)	7 (13.0%)	0.009	39 (12%)	2 (0.6%)
Renal involvement, no. (%)§	1 (1.5%)	13 (11%)	10 (11.5%)	8 (14.3%)	0.039	32 (9.8%)	0 (0%)
eGFR (mL/min)	90 (63–90)	84.5 (60–90)	90 (60–90)	80 (60–90)	0.339	85 (60–90)	92 (28.2%)
Renal crisis	0 (0%)	8 (6.8%)	4 (4.6%)	4 (7.1%)	0.110	16 (4.9%)	0 (0%)
Plasma creatinine (µmol/L)	63.5 (55–72)	69 (55–83)	64.5 (53–86)	64 (56–77)	0.422	65 (55–79.5)	70 (21.5%)

Continued

Table 1 Continued

Characteristic	Protocol 1 Methotrexate n=65	Protocol 2 Mycophenolate mofetil, n=118	Protocol 3 Cyclophosphamide n=87	Protocol 4 No immunosuppressant n=56	P Value*	Total n=326	Missing at baseline, no. (%)
Any GI involvement, no. (%)	25 (38.5%)	32 (27.1%)	30 (34.5%)	26 (46.4%)	0.078	113 (34.7%)	0 (0%)
Muscle involvement, no. (%)	12 (18.5%)	9 (7.6%)	10 (11.5%)	0 (0%)	0.002	31 (9.5%)	0 (0%)
Current digital ulcers	10 (15.4%)	17 (14.4%)	17 (19.5%)	11 (19.6%)	0.705	55 (16.9%)	0 (0%)
Patient questionnaire data							
HAQ-DI (0–3)	1.1 (0.4–1.8)	1 (0.5–1.6)	1 (0.4–1.9)	0.7 (0.1–1.5)	0.400	1 (0.4–1.8)	19 (5.8%)
FACIT fatigue score (0–52)	33 (21.5–42.5)	30 (20–37)	31 (17–40)	37 (21–44)	0.165	31 (20–41)	16 (4.9%)
SF36 physical score (0–100)	39.2 (30.9–45.1)	36.4 (30.8–43)	36 (27.2–44.5)	39.2 (32–48.4)	0.203	37.4 (29.9–45)	15 (4.6%)
SF36 mental score (0–100)¶	37.2 (33.3–44)	36.8 (32.2–41.9)	39.5 (35.7–44.7)	40.8 (35.9–44.1)	0.029	38.3 (34.3–44)	15 (4.6%)
Cochin hand function score (0–90) **	11.5 (2–30.5)	10 (4–24)	16 (5–40)	6.5 (0–23)	0.025	11 (3–29)	96 (29.4%)

Median (IQR) unless otherwise indicated.
*p indicates significance of Kruskal-Wallis test (for continuous variables) or Fisher's exact test (for categorical variables).
†Of the 26 patients who had previously received immunosuppressant therapy, in 2 patients this was for cancer.
‡86 patients had a sPAP/RVSP value assumed to be normal and thus not measured. If those cases are omitted, only 38 values of sPAP/RVSP are missing (11.7%). Median values are 'falsely' high because calculation omits unmeasured (normal) values.
§Renal involvement is defined as renal crisis and/or moderate-to-severe renal impairment.
¶Despite the significant p-value for the Kruskal-Wallis test, post hoc tests reject any between-group differences in the SF36 mental scores.
** Cochin hand function scores were not performed in all centres because of translational issues.
CRP, C reactive protein; DLCO, carbon monoxide diffusing capacity; eGFR, estimated glomerular filtration rate; ESR, erythrocyte sedimentation rate; FACIT, Functional Assessment of Chronic Illness Therapy; FVC, forced vital capacity; GI, gastrointestinal; HAQ-DI, Health Assessment Questionnaire Disability Index; mRSS, modified Rodnan skin score (17 sites); RVSP, right ventricular systolic pressure; SF36, Short-Form 36; sPAP, systolic pulmonary artery pressure.

statistically significant reductions in mRSS in all four treatment groups at 12 months (−4.0 (−5.2 to −2.7) units for methotrexate, −4.1 (−5.3 to −2.9) for MMF, −3.3 (−4.9 to −1.7) for cyclophosphamide and −2.2 (−4.0 to −0.3) for the no immunosuppressant group), but the differences between treatments were not significant (p=0.346) (table 3 and figure 2).

Changes in secondary outcomes over time in the different treatment groups
Lung function
After adjusting for potential confounders, the change rates of FVC and DLCO were not significantly different in the four treatment groups (p=0.460 and 0.505) (table 3).

However, in a subset of patients with pulmonary fibrosis or suspected pulmonary fibrosis (cases confirmed on high-resolution CT (HRCT) irrespective of FVC or DLCO, or with one of the following if HRCT not performed: FVC or DLCO under 55% predicted or definite bibasal shadowing on X-ray), there was a significant difference in the change rate of FVC over time (p=0.035). Patients initially prescribed cyclophosphamide demonstrated 7.4% absolute increase in FVC (% predicted) compared with 2.0% decrease for methotrexate, 3.2% increase for MMF and 4.0% increase for the 'no immunosuppressant' group (table 3).

Functional ability and hand function
Changes over time for the HAQ-DI and CHFS did not differ between protocols (p=0.130 and 0.073), regardless of adjusting (table 3).

Development of internal organ involvement
This is described in online supplementary figure S2.

Comparison of survival between treatment protocols
Survival was lowest in the no immunosuppressant group at both 12 and 24 months but differences between protocols were not statistically significant either before (p=0.389) or after weighting (p=0.440). In the adjusted model, at 24 months, those in the no immunosuppressant group had a predicted survival rate of 84.0% compared with 94.1% for methotrexate, 88.8% for MMF and 90.1% for cyclophosphamide (figure 3). Patients with lung involvement (pulmonary fibrosis and/or hypertension) at baseline had significantly poorer survival than those without: at 24 months, their predicted survival rate was 74.6% versus 91.7% (p<0.0005) and similarly for cardiac involvement, 71.6% versus 90.7% (p<0.0005).

Adverse effects
Of the 75, 182 and 101 patients who were ever on methotrexate, MMF or cyclophosphamide, respectively, 29 (38.7%), 40 (22.0%) and 23 (22.8%) were reported to have had side effects, necessitating drug discontinuation in 9 (12.0%), 14 (7.7%) and 5 (4.5%) patients, respectively. A survival analysis on protocol exits due to adverse effects showed no differences in the tolerability of the three treatments (p=0.212) (see online supplementary figure S3).

DISCUSSION
Our main findings were, first, that there were no significant differences in outcome between the four treatment protocols (methotrexate, MMF, cyclophosphamide, no immunosuppression), although there may be a signal in favour of immunosuppression for early dcSSc. Although skin score improved in all treatment groups, this was least in the no immunosuppressant category, who also had the highest mortality. Second, ESOS

Table 2 Associations between baseline characteristics and skin score

Characteristic	(A) Association with baseline mRSS		(B) Effect on mRSS evolution		
	Baseline predictor coefficient (95% CI)	p[(1)]	Time slope (12 months)	Time–predictor interaction coefficient (95% CI)	p[(2)]
Age, per 10 years	0.9 (0.3 to 1.6)	0.004	−1.8	−0.3 (−0.6 to −0.1)	0.007
Female	−1.8 (−3.7 to 0.2)	0.072	−3.5	0.1 (−0.5 to 0.8)	0.680
Months since onset of skin thickening	−0.1 (−0.2 to 0)	0.156	−4.3	0.1 (0 to 0.1)	0.002
Previous immunosuppressant use	−2.1 (−5.3 to 1.2)	0.205	−3.5	0.1 (−1 to 1.3)	0.808
Current or previous steroid use	1.2 (−0.6 to 3.0)	0.191	−2.9	−1.2 (−1.8 to −0.6)	<0.0005
mRSS (0–51), per 5 units			0.9	−1 (−1.2 to −0.8)	<0.0001
White blood count (×109/L)	0.5 (0.2 to 0.8)	0.002	−2.2	−0.1 (−0.3 to 0)	0.014
ESR (mm/hour)	0.1 (0 to 0.1)	0.003	−4.0	0 (0 to 0)	0.079
CRP (mg/L)	0.1 (0 to 0.2)	0.003	−4.4	0.1 (0 to 0.1)	0.001
Anti-topoisomerase (anti-Scl70)	−2.6 (−4.4 to −0.8)	0.005	−4.3	2.2 (1.6 to 2.8)	<0.0001
Anti-RNA polymerase III	4.5 (2.1 to 6.9)	<0.0005	−3.0	−2.1 (−2.9 to −1.2)	<0.0001
Anticentromere	−0.4 (−3.9 to 3.1)	0.816	−3.4	0.5 (−0.8 to 1.9)	0.456
Organ involvement					
Pulmonary fibrosis	2.9 (0.4 to 5.4)	0.021	−3.5	0.3 (−0.6 to 1.2)	0.534
FVC (% predicted)	−0.1 (−0.1 to 0)	0.013	−5.5	0 (0 to 0)	0.005
DLCO (% predicted)	0 (−0.1 to 0)	0.105	−3.2	0 (0 to 0)	0.689
Pulmonary hypertension	2.5 (−0.7 to 5.8)	0.128	−3.3	−2.1 (−3.3 to −0.8)	0.001
Cardiac involvement	2.5 (−0.2 to 5.2)	0.075	−3.5	0 (−1.0 to 0.9)	0.929
Renal involvement*	2.2 (−0.7 to 5.2)	0.140	−3.3	−1.6 (−2.6 to −0.5)	0.004
Any GI involvement, no. (%)	2.2 (0.3 to 4.0)	0.021	−3.4	−0.1 (−0.8 to 0.5)	0.664
Muscle involvement	1.2 (−1.8 to 4.2)	0.425	−3.2	−2 (−2.9 to −1.0)	<0.0005
Current digital ulcers	2.5 (0.1 to 4.8)	0.038	−3.3	−0.9 (−1.8 to 0)	0.047
Patient questionnaire data					
HAQ-DI (0–3)	3.5 (2.5 to 4.6)	<0.0001	−3.0	−0.4 (−0.8 to 0)	0.039
FACIT fatigue score (0–52), per 10 units	−1.3 (−2.0 to −0.6)	<0.0005	−3.1	−0.1 (−0.3 to 0.2)	0.484
SF36 physical score (0–100)	−0.2 (−0.3 to −0.1)	<0.0001	−3.3	0 (0 to 0)	0.873
SF36 mental score (0–100)	0.1 (−0.1 to 0.2)	0.297	−1.8	0 (−0.1 to 0)	0.081
Cochin hand function score (0–90), per 10 units	1.4 (0.9 to 1.9)	<0.0001	−3.7	0 (−0.2 to 0.2)	0.979

Example for interpretation of results: the presence of anti-RNA polymerase III is associated with (A) a higher mRSS by 4.5 units at baseline and (B) losing an extra 2.1 units per year compared with an average of −3.0 units per year for all patients.
*Renal involvement is defined as renal crisis and/or moderate-to-severe renal impairment.
p[(1)]: Significance p value for characteristic coefficient in linear regression of baseline mRSS on baseline predictor.
p[(2)]: Significance p value for interaction coefficient between time and baseline characteristic in a longitudinal regression model.
CRP, C reactive protein; DLCO, carbon monoxide diffusing capacity; ESR, erythrocyte sedimentation rate; FACIT, Functional Assessment of Chronic Illness Therapy; FVC, forced vital capacity; GI, gastrointestinal; HAQ-DI, Health Assessment Questionnaire Disability Index; mRSS, modified Rodnan skin score (17 sites); SF36, Short-Form 36.

confirms the relative effectiveness of cyclophosphamide in patients with pulmonary fibrosis.[25 26]

An important point when interpreting our findings (and therefore a note of caution) is that the 'no immunosuppressant' group was not a control group. Patients in this group had a longer disease duration than the other three groups and were more likely to have renal involvement.

Our findings lend support to two recently published studies (the Autologous Stem Cell Transplantation International Scleroderma trial (ASTIS) trial of autologous stem cell transplantation[27] and the Scleroderma Lung Study (SLS) II (comparing MMF and cyclophosphamide),[26] which suggest benefit, including in mRSS, from immunosuppression (as did SLS 1[25]). In ASTIS, those patients randomised to cyclophosphamide had an 8.8 unit fall in mRSS (from 25.8) at 24 months (compared with 3.3 in ESOS over 12 months), but the cyclophosphamide protocol was more intense, and the patients had more severe disease (patients with the highest mRSS at baseline tend to improve most quickly[4] as also demonstrated by our own findings (table 2)). MRSS fell by 19.9 units in those patients

randomised to stem cell transplantation[27] (and therefore intensive immunosuppression). In SLS 1,[25] patients with dcSSc randomised to cyclophosphamide experienced a 5.3 unit fall in mRSS at 12 months (compared with 3.3 in ESOS), whereas mRSS fell by 1.7 on placebo (compared with 2.2 units in the ESOS 'no immunosuppressant' group). In SLS II,[26] mRSS at 24 months fell 4.9 units on MMF (compared with 4.1 units in ESOS at 12 months) and by 5.4 after 12 months treatment with cyclophosphamide, although these values are not directly comparable because they relate to patients with limited cutaneous and dcSSc combined.

The methodological strength of ESOS, which built upon experience gained in a previous, smaller observational study,[28] was its design: its standardised protocols emulated the conditions of a clinical trial, and although not randomised, patients were enrolled into four homogenous treatment arms with well-defined interventions and a systematic record of protocol changes and exits. Entry criteria were deliberately inclusive: RCTs often exclude patients with internal organ involvement and for whom immunosuppression is most likely to be

Table 3 Predicted yearly changes in outcomes and survival rates according to initial protocol, with and without adjusting (95% CI)

Outcome	Model specification	p Value	Methotrexate	Mycophenolate mofetil	Cyclophosphamide	No immunosuppressant
mRSS (0–51)	No adjusting, n=326	0.252	−4.4 (−5.7 to −3.2)	−3.8 (−4.9 to −2.8)	−3.5 (−5.0 to −2.0)	−2.4 (−3.9 to −1)
	Adjusting for confounding (weighted model), n=326	0.346	−4.0 (−5.2 to −2.7) (n=65)	−4.1 (−5.3 to −2.9) (n=118)	−3.3 (−4.9 to −1.7) (n=87)	−2.2 (−4.0 to −0.3) (n=56)
FVC (% predicted)	No adjusting, n=307	0.045	−1.7 (−4.4 to 1.0)	2.1 (−0.1 to 4.3)	4.0 (0.5 to 7.5)	2.7 (−0.8 to 6.2)
	Adjusting for confounding (weighted model), n=307	0.460	−0.5 (−3.7 to 2.6) (n=59)	2.0 (−0.7 to 4.6) (n=111)	3.3 (−0.6 to 7.2) (n=84)	2.0 (−1.6 to 5.6) (n=53)
	Subset with PF on HRCT, weighted, n=129†	0.035	−2.0 (−5.9 to 2.0) (n=19)	3.2 (−0.6 to 7.0) (n=31)	7.4 (2.2 to 12.7) (n=57)	4.0 (−1 to 9.0) (n=22)
DLCO (% predicted)	No adjusting, n=291	0.703	−1.2 (−3.5 to 1.0)	0.3 (−1.7 to 2.3)	0.6 (−2.0 to 3.3)	−0.3 (−4.2 to 3.7)
	Adjusting for confounding (weighted model), n=291	0.505	−1.6 (−3.8 to 0.6) (n=51)	0.8 (−1.5 to 3.1) (n=110)	0.1 (−3.1 to 3.2) (n=79)	−0.1 (−3.7 to 3.4) (n=51)
	Subset with PF on HRCT, weighted, n=116†	0.809	−0.8 (−5.4 to 3.8) (n=12)	1.9 (−1.8 to 5.6) (n=31)	1.8 (−1.9 to 5.4) (n=53)	1.6 (−2.6 to 5.7) (n=20)
HAQ-DI (0–3)	No adjusting, n=307	0.070	−0.1 (−0.2 to 0.1)	0 (−0.1 to 0.1)	−0.2 (−0.3 to −0.1)	0.1 (−0.1 to 0.2)
	Adjusting for confounding (weighted model), n=307	0.130	−0.1 (−0.3 to 0) (n=59)	0 (−0.1 to 0.1) (n=113)	−0.1 (−0.3 to 0) (n=81)	0.1 (−0.1 to 0.2) (n=54)
Cochin hand function (0–90)‡	No adjusting, n=230	0.072	−1.1 (−3.5 to 2.2)	−0.3 (−1.6 to 1.2)	−3.1 (−5.1 to −0.5)	1.1 (−0.3 to 3.0)
	Adjusting for confounding (weighted model), n=230	0.073	−1.4 (−3.5 to 1.4) (n=36)	−0.6 (−2.0 to 1.1) (n=103)	−2.4 (−4.7 to 0.6) (n=49)	1.7 (-0.1 to 4.0) (n=42)
Survival	No adjusting, n=326	0.389	(12 months) 93.5% (24 months) 93.5%	(12 months) 96.5% (24 months) 89.3%	(12 months) 88.1% (24 months) 85.4%	(12 months) 88.9% (24 months) 85.1%
	Adjusting for confounding (weighted model), n=326	0.440	(12 months) 94.1% (24 months) 94.1%	(12 months) 96.1% (24 months) 88.8%	(12 months) 91.7% (24 months) 90.1%	(12 months) 88.6% (24 months) 84.0%

Significance p: Fisher's test for equality of change rates between protocols, for each outcome variable.
*Results are reported in terms of changes after 12 months. However, all study data (from baseline to the 24-month end point) were used in estimation. To obtain 24-month changes, multiply results above by 2.
†For the subanalysis involving the subset of patients with pulmonary fibrosis at baseline, patients with definite bibasal pulmonary fibrosis confirmed on HRCT were included, irrespective of FVC value. If no HRCT scan was performed at baseline, an FVC<55%, DLCO<55% predicted or definite bibasal shadowing on X-ray was also a basis for inclusion.
‡Changes expressed in units for the Cochin regression are an approximation derived from the 95% CI of percentage changes between baseline and 12 months (on a scale shifted by one unit), applied to the predicted baseline values for each group in the original scale.
DLCO, carbon monoxide diffusing capacity; FVC, forced vital capacity; HAQ-DI, Health Assessment Questionnaire Disability Index; HRCT, high-resolution CT; mRSS, modified Rodnan skin score (17 sites); PF, pulmonary fibrosis.
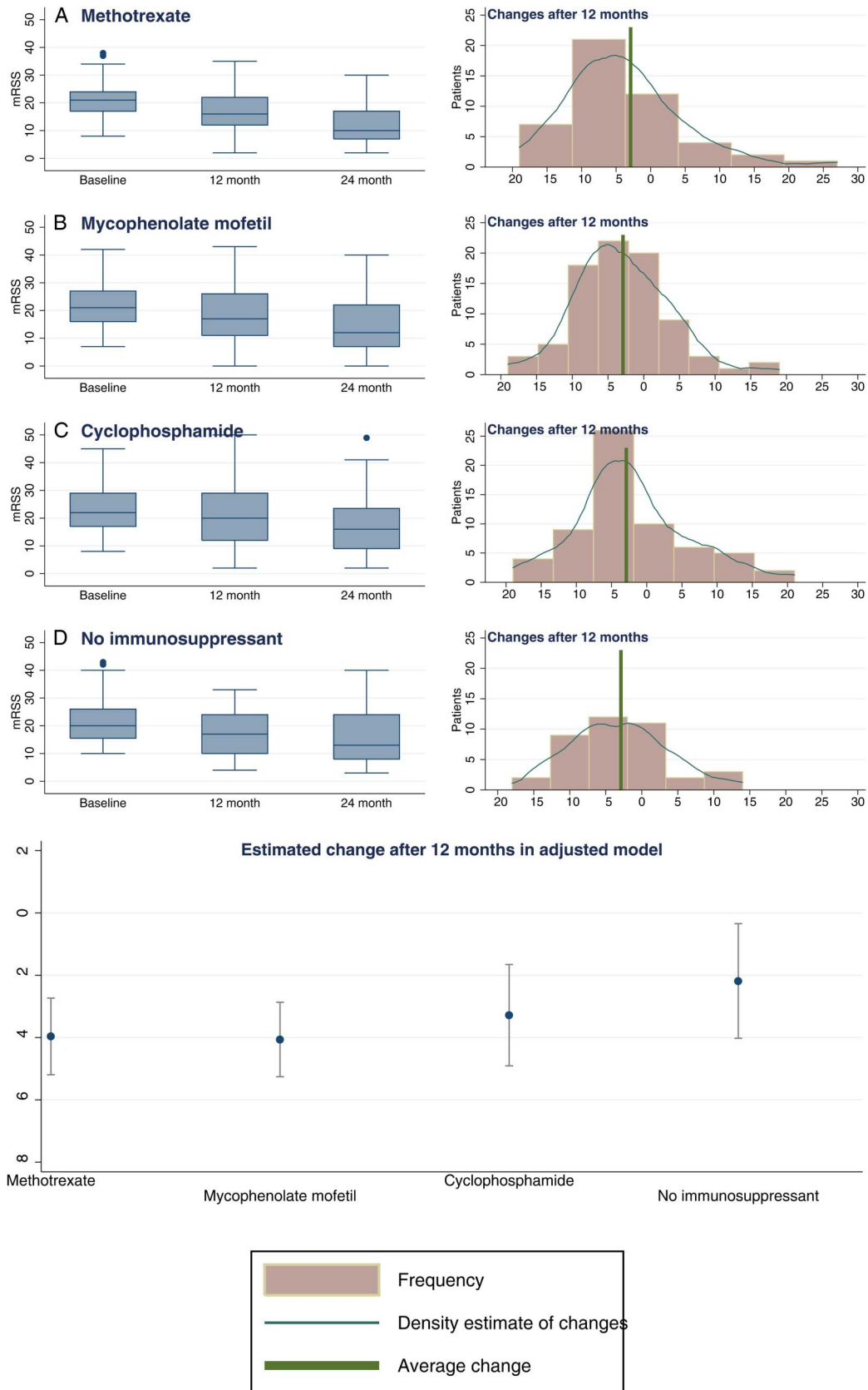
Figure 2 Modified Rodnan skin score (mRSS) during baseline and follow-up visits, by initial protocol. For each group of patients, according to their initial protocol, the distribution of the skin score is illustrated on the left-hand side by box and whisker plots (indicating the median and IQR) at baseline, 12 and 24 months. On the right-hand side, the distribution of individual 1-year changes in the skin score is described by histograms and a kernel density estimate. In addition, a vertical green line indicates the value of the average 1-year change in the skin score, irrespective of treatment choice. The bottom panel in the figure describes the estimated changes in mRSS (with 95% CI) according to initial protocol, based on the results from the adjusted model (described in table 3).
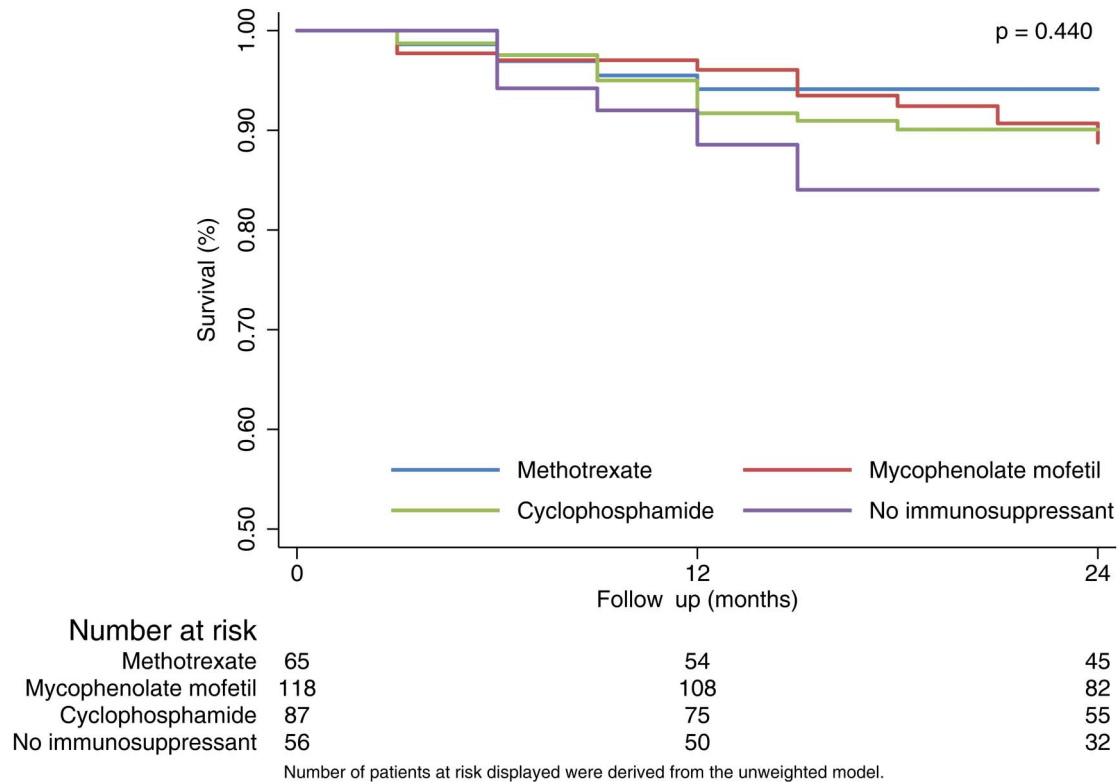
Figure 3 Kaplan-Meier estimated survival curves by treatment group.

beneficial. By recruiting 326 patients from 50 centres, ESOS represents a large cohort of patients with very early dcSSc (median duration of skin thickening 11.9 months): its data will serve as a benchmark when designing and interpreting future clinical trials. This is especially relevant with a number of novel treatment approaches currently being explored including biological agents. For example, in a recent RCT of tocilizumab,[29] mRSS fell over 24 weeks by 3.9 units from 26 in the 43 tocilizumab-treated patients and by 1.2 units from 26 in the 44 placebo-treated patients, this latter fall comparable to the ESOS 'no immunosuppressant' response. In comparing between these studies, the higher baseline mRSS in the tocilizumab study should be borne in mind.

The main weakness of observational studies is that each patient's outcome on her/his treatment arm cannot be completely disentangled from her/his initial characteristics. For instance, ESOS has verified that patients with lung and cardiac involvement tend to be prescribed cyclophosphamide. However, adjusting using IPT weights minimises the problem of confounding by indication.

In conclusion, observational studies offer a rich population-wide perspective assessing treatment effects in a real-world setting. ESOS achieved its aim of following a large international cohort of patients with early dcSSc over 2 years, each of whom was treated according to one of four protocols. The message for clinicians is that there is a weak signal to support using immunosuppressants for early dcSSc (and in particular cyclophosphamide for patients with pulmonary fibrosis). However, it is clear that there remains a pressing need for the development of more effective and targeted treatments.

Author affiliations
[1]Centre for Musculoskeletal Research, The University of Manchester, Salford Royal NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK
[2]NIHR Manchester Musculoskeletal Biomedical Research Unit, Central Manchester NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK
[3]Centre for Musculoskeletal Research, The University of Manchester, Manchester Academic Health Science Centre, Manchester, UK
[4]Department of Rheumatology, Lund University, Lund, Sweden
[5]Service de Médecine Interne, Hôpital Cochin, Centre de Référence pour les Vascularites Nécrosantes et la Sclérodermie Systémique, Université Paris Descartes, Assistance Publique-Hôpitaux de Paris (AP-HP), Paris, France
[6]Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK
[7]Member of Steering Committee, contact via Professor Herrick, The University of Manchester, Manchester, UK
[8]Department of Rheumatology and Immunology, Medical Center, University of Pécs, Pecs, Hungary
[9]Department of Internal Medicine 3, University of Erlangen-Nuremberg, Erlangen, Germany
[10]Department of Rheumatology, University of Zurich, Zurich, Switzerland
[11]Royal Free London NHS Foundation Trust, London, UK.
[12]Rehabilitation Services, Salford Royal NHS Foundation Trust, Salford, UK
[13]Department of the Rheumatic Diseases, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands
[14]Rheumatology 2 Department, "Grigore T. Popa" University of Medicine and Pharmacy, Clinical Rehabilitation Hospital, Iași, Romania
[15]UCL Division of Medicine, Centre for Rheumatology and Connective Tissue Diseases, London, UK
[16]Unité Clinique de Médecine Interne, Maladies Auto-immunes et Pathologie Vasculaire, UF 04, Hôpital Saint-Louis, AP-HP Assistance Publique des Hôpitaux de Paris, INSERM UMRS 1160, Paris Denis Diderot University, France
[17]Jewish General Hospital, Lady Davis Institute and McGill University, Montreal, Canada
[18]Department Experimental and Clinical Medicine, Division of Rheumatology AOUC, University of Florence, Florence, Italy
[19]Shine Rheumatology Unit, Rambam Heath Care Campus; Rappaport Faculty of Medicine, Haifa, Israel
[20]Rheumatology Unit, Oslo University Hospital Rikshospitalet, Oslo, Norway
[21]Queen Elizabeth Hospital Birmingham, UHB Foundation Trust, Birmingham, UK
[22]St Vincent's Hospital, Melbourne, Australia
[23]Department for Dermatology, University of Cologne Kerpenerstr. 62, Köln, Germany
[24]Cambridge University NHS Hospital Foundation Trust, Cambridge, UK
[25]Department of Internal Medicine, Hôtel-Dieu Hospital, University of Nantes, Nantes, France

[26]University of Liverpool, Aintree University Hospital, Liverpool, UK
[27]Service de Médecine Interne, Hôpital Bretonneau Tours Cedex, France
[28]Centre for Rheumatic Diseases, Glasgow Royal Infirmary, Glasgow, UK
[29]Sheffield Teaching Hospitals, Sheffield, UK
[30]Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds and NIHR Leeds Musculoskeletal Biomedical Research Unit, Leeds Teaching Hospitals NHS Trust, UK
[31]Stanford University, Stanford, California, USA.
[32]University of Belgrade School of Medicine, Institute of Rheumatology, Belgrade, Serbia
[33]Clinical and Academic Rheumatology, North Bristol NHS Trust, Bristol, UK
[34]Nottingham University Hospitals NHS Trust, and Nottingham NHS Treatment Centre, Nottingham, UK
[35]Peter Maddison Rheumatology Centre, Llandudno, UK
[36]Queens Hospital, Romford, UK
[37]University of Copenhagen, Copenhagen Lupus and Vasculitis Clinic, Center for Rheumatology and Spine Diseases, Rigshospitalet, Copenhagen, Denmark
[38]Norwich Medical School, University of East Anglia, Norwich, UK
[39]Royal National Hospital for Rheumatic Diseases, Bath, UK
[40]Department of Rheumatology and Clinical Immunology, Justus-Liebig University Giessen, Bad Nauheim, Germany
[41]Department of Rheumatology, University of Lübeck, Lübeck, Germany
[42]Department of Rheumatology and Clinical Immunology, University Hospital Charité Berlin, Berlin, Germany
[43]Department of Rheumatology, Royal Perth Hospital, Perth, Australia
[44]Servicio de Reumatologia. Hospital Universitario 12 de Octubre, Madrid, Spain
[45]Internal Medicine Unit, Limoges University Hospital, France
[46]Département de Médecine Interne et Immunologie Clinique, Centre National de Référence Maladies Systémiques etAuto-immunes Rares, Université de Lille, Inserm, U995, FHU Immune-Mediated Inflammatory Diseases and Targeted Therapies, Lille, France
[47]Gateshead Hospitals Foundation Trust, Gateshead, UK
[48]Department of Internal Medicine, Division of Rheumatology, Istanbul University, Istanbul, Turkey
[49]Fife Rheumatic Diseases Unit, Whyteman's Brae Hospital, Kirkcaldy, UK
[50]Department of Rheumatology and Clinical Immunology, UMC Utrecht, Utrecht, The Netherlands
[51]James Cook University Hospital, Middlesbrough, UK
[52]Rheumatology Unit, Royal Adelaide Hospital, and Discipline of Medicine, University of Adelaide, Adelaide, South Australia
[53]Department of Rheumatology and Inflammation Research, The Sahlgrenska Academy at Gothenburg University, Gothenburg, Sweden
[54]Monash Centre for Inflammatory Diseases, Monash University, Clayton, Melbourne, Australia
[55]Department of Internal Medicine, Hôpital Edouard Herriot, Lyon, France
[56]Department of Internal Medicine, Foundation Hospital Saint Joseph, Marseille, France
[57]Cannock Chase Hospital, Cannock, UK
[58]St Vincent's University Hospital, Dublin, Ireland
[59] Department of Internal Medicine, Centre Hospitalier Lyon Sud, Pierre Benite, France
[60]Internal Medecine, Ambroise Paré Hospital, Boulogne Billancourt, France

REFERENCES

1 Nikpour M, Stevens WM, Herrick AL, et al. Epidemiology of systemic sclerosis. *Best Practice Res Clin Rheumatol* 2010;24:857–69.
2 Rubio-Rivas M, Royo C, Simeon CP, et al. Mortality and survival in systemic sclerosis : systematic review and meta-analysis. *Sem Arthritis Rheum* 2014;44:208–19.
3 Nihtyanova SI, Schreiber BE, Ong VH, et al. Prediction of pulmonary complications and long-term survival in systemic sclerosis. *Arthritis Rheum* 2014;66:1625–35.
4 Merkel PA, Silliman NP, Clements PJ, et al. Patterns and predictors of change in outcome measures in clinical trials in scleroderma: An individual patient meta-analysis of 629 subjects with diffuse cutaneous systemic sclerosis. *Arthritis Rheum* 2012;64:3420–9.
5 Silverman SL. From randomized controlled trials to observational studies. *Am J Med* 2009;122:114–20.
6 Halpern SD, Karlawish JT, Berlin JA. The continuing unethical conduct of underpowered clinical trials. *JAMA* 2002;288:358–62.
7 Pope JE, Bellamy N, Seibold JR, et al. A randomized, controlled trial of methotrexate versus placebo in early diffuse scleroderma. *Arthritis Rheum* 2001;44:1351–8.
8 Denton CP, Merkel PA, Furst DE, et al. Recombinant human anti-transforming growth factor β1 antibody therapy in systemic sclerosis: a multicentre, randomized, placebo-controlled Phase I/II trial of CAT-192. *Arthritis Rheum* 2007;56:323–3.
9 Kowal-Bielecka O, Landewé R, Avouac J, et al. EULAR recommendations for the treatment of systemic sclerosis: a report from the EULAR Scleroderma Trials and Research Group (EUSTAR). *Ann Rheum Dis* 2009;68:620–8.
10 Rawlins MD. *The Harveian Oration of 2008: On the evidence for decisions about the use of therapeutic interventions*. London: Royal College of Physicians, 2008.
11 Gagne J, Thompson L, O'Keefe K, et al. Innovative research methods for studying treatments for rare diseases: methodological review. *BMJ* 2014;349:g6802.
12 LeRoy EC, Black C, Fleischmajer R, et al. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. *J Rheumatol* 1988;15:202–5.
13 Clements P, Lachenbruch P, Siebold J, et al. Inter- and intraobserver variability of total skin thickness score (modified Rodnan TSS) in systemic sclerosis. *J Rheumatol* 1995;22:1281–5.
14 Clements PJ, Hurwitz EL, Wong WK, et al. Skin thickness score as a predictor and correlate of outcome in systemic sclerosis. *Arthritis Rheum* 2000;43:2445–54.
15 Steen VD, Medsger TA. 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.
16 Webster K, Cella D, Yost K. The Functional Assessment of Chronic Illness Therapy (FACIT) Measurement System: properties, applications, and interpretation. *Health Qual Life Outcomes* 2003;1:79.
17 Harel D, Thombs BD, Hudson M, et al. Measuring fatigue in SSc: a comparison of the Short Form-36 Vitality subscale and Functional Assessment of Chronic Illness Therapy–Fatigue scale. *Rheumatol* 2012;51:2177–85.
18 Rannou F, Poiraudeau S, Berezne A, et al. Assessing disability and quality of life in systemic sclerosis: construct validities of the Cochin Hand Function Scale, Health

Assessment Questionnaire (HAQ), Systemic Sclerosis HAQ, and Medical Outcomes Study 36-Item Short Form Health Survey. *Arthritis Rheum* 2007;57:94–102.

19 Brookhart MA, Schneeweiss S, Rothman KJ, *et al*. Variable selection for propensity score models. *Amer J Epidemiol* 2006;163:1149–56.

20 Austin PC, Grootendorst P, Normand S-LT, *et al*. Conditioning on the propensity score can result in biased estimation of common measures of treatment effect: a Monte Carlo study. *Stat Med* 2007;26:754–68.

21 Sato T, Matsuyama Y. Marginal structural models as a tool for standardization. *Epidemiology* 2003;14:680–6.

22 Imbens G. The role of the propensity score in estimating dose-response functions. *Biometrika* 2000;87:706–10.

23 Fewell Z, Hernan MA, Wolfe F, *et al*. Controlling for time-dependant confounding using marginal structural models. *Stata J* 2004;4:402–20.

24 Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. *Amer J Epidemiol* 2008;168:656–64.

25 Tashkin DP, Elashoff R, Clements PJ, *et al*. Cyclophosphamide versus placebo in scleroderma lung disease. *N Eng J Med* 2006;354:2655–66.

26 Tashkin DP, Roth MD, Clements PJ, *et al*. Mycophenolate mofetil versus oral cyclophosphamide in scleroderma-related interstitial lung disease (SLS II): a randomsed controlled, double-blind, parallel group trial. *Lancet Respir Med* 2016;4:708–19.

27 Van Laar JM, Farge D, Sont JK, *et al*. Autologous hematopoietic stem cell transplantation vs intravenous pulse cyclophosphamide in diffuse cutaneous systemic sclerosis. *JAMA* 2014;311:2490–8.

28 Herrick A, Lunt M, Whidby N, *et al*. Observational study of treatment outcome in early diffuse cutaneous systemic sclerosis. *J Rheumatol* 2010;37;116–24.

29 Khanna D, Denton CP, Jahreis A, *et al*. Safety and efficacy of subcutaneous tocilizumab in adults with systemic sclerosis (faSScinate): a phase 2, randomised controlled trial. *Lancet* 2016;387:2630–40.