EXTENDED REPORT

Patterns and predictors of skin score change in early diffuse systemic sclerosis from the European Scleroderma Observational Study


Handling editor Tore K Kvien

ABSTRACT

Objectives Our aim was to use the opportunity provided by the European Scleroderma Observational Study to (1) identify and describe those patients with early diffuse cutaneous systemic sclerosis (dcSSc) with progressive skin thickness, and (2) derive prediction models for progression over 12 months, to inform future randomised controlled trials (RCTs).

Methods The modified Rodnan skin score (mRSS) was recorded every 3 months in 326 patients. ‘Progressors’ were defined as those experiencing a 5-unit and 25% increase in mRSS score over 12 months (±3 months). Logistic models were fitted to predict progression and, using receiver operating characteristic (ROC) curves, were compared on the basis of the area under curve (AUC), accuracy and positive predictive value (PPV).

Results 66 patients (22.5%) progressed, 227 (77.5%) did not (33 could not have their status assessed due to insufficient data). Progressors had longer disease duration (median 8.1 vs 12.6 months, P<0.001) and lower mRSS (median 19 vs 21 units, P=0.030) than non-progressors. Skin score was highest, and peaked earliest, in the anti-RNA polymerase III (PolIII+) subgroup (n=50). A first predictive model (including mRSS, duration of skin thickening and their interaction) had an accuracy of 60.9%, AUC of 0.666 and PPV of 33.8%. By adding a variable for PolIII positivity, the model reached an accuracy of 71%, AUC of 0.711 and PPV of 41%.

Conclusions Two prediction models for progressive skin thickening were derived, for use both in clinical practice and for cohort enrichment in RCTs. These models will inform recruitment into the many clinical trials of dcSSc projected for the coming years.

Trial registration number NCT02339441.

INTRODUCTION

Patients with the diffuse cutaneous subtype of systemic sclerosis (dcSSc) have high morbidity and mortality, associated with the degree of severity of skin fibrosis/thickening as assessed by the modified Rodnan skin score (mRSS).1 2 The mRSS, as well as being a key clinical tool that clinicians use to monitor patients in everyday clinical practice, is usually the primary end point in randomised controlled trials (RCTs) of dcSSc. These trials pose particular challenges first because dcSSc is a rare disease, and second because mRSS tends to rapidly progress over time (usually within the first 3–5 years), but then to ‘plateau’ and often subsequently fall,3 probably contributing to why several treatments associated with benefit in open-label or observational studies have not conferred benefit in RCTs.4–7 Ideally, we need to be able to predict which patients are likely to progress in terms of mRSS and recruit from this subset into RCTs. Most RCTs have restricted inclusion to patients with early disease (some within 18 months of onset of skin thickening),1 4–8 others within 3–5 years.9–12 More recently, it has been suggested that an upper mRSS cut-off could further enrich the cohort for...
worsening skin.\textsuperscript{13,14} with 22 as a proposed level.\textsuperscript{15} However, the stricter the inclusion criteria, inevitably the more difficult it will be to recruit. This is a key issue: recent advances are driving new approaches to therapy, and recruitment is now increasingly difficult with competing studies.

The European Scleroderma Observational Study (ESOS)\textsuperscript{16} was a prospective observational study of treatment outcome in 326 patients with early dcSSc. Patients were assessed every 3 months for 12–24 months (most for 24 months), with mRSS documented at each visit. Thus, ESOS provided a unique opportunity to perform a detailed study of mRSS trajectory over time in a large multinational cohort with very early disease (median disease duration from onset of skin thickening: 11.9 months). Our aim was twofold: (1) for the practising clinician, to identify and describe (in the ESOS cohort) patients with progressive skin thickness; and (2) for the clinical trialist, to derive predictive models for progression over 12 months, in order to inform/ maximise recruitment into future RCTs.

METHODS
ESOS study design and patients
This is described fully elsewhere\textsuperscript{15}: patients with early dcSSc were recruited into a prospective, observational cohort study comparing the effectiveness of four different treatment protocols. The main inclusion criteria were early dcSSc (skin involvement extending proximal to elbow or knee and/or involving trunk),\textsuperscript{16} and within 3 years of the onset of skin thickening as judged by physician at screening visit) and age >18 years. Patients attended every 3 months for 12–24 months. The primary outcome measure was the mRSS. Demographic and clinical characteristics including age, gender, smoking habit, ethnicity, antibody status (antitopoisomerase-1 (anti-Scl-70, ‘TOPO’), anti-RNA polymerase III (‘Pol3’), anti-centromere (‘ACA’)) and presence of visceral organ involvement were recorded for all patients.\textsuperscript{15} There were 326 patients from 50 centres (19 countries) who were recruited: 65 started on methotrexate, 118 on mycophenolate mofetil, 87 on cyclophosphamide and 56 no immunosuppressant. Four patients who were found postrecruitment to have a baseline duration of skin thickening >36 months (up to 44.6) were retained (a subsidiary analysis verified the robustness of our predictive models to their inclusion). Because progression status did not significantly differ between treatment groups, mRSS trajectories were analysed irrespective of treatment protocol (online supplementary table S1). Each patient gave written informed consent.

Definition of progressive patients
Disease progression was defined in terms of mRSS worsening, in line with most recent RCTs. For the univariate analysis and predictive models, patients with progressive disease (‘progressors’) are defined as those with a 5-unit and 25% increase in their mRSS between baseline and their highest subsequent score. This threshold is generally considered to reflect meaningful change in mRSS progression,\textsuperscript{17} thus enabling model comparisons.\textsuperscript{13,14,18} We considered only peaks occurring during the first 12±3 months after baseline, using all 3-monthly observations. The time window was chosen because it is considered an appropriate period to detect clinically meaningful changes in the skin score.\textsuperscript{19} Most cases of progression occurred early: extending the time period to 24 months would have added only four additional ‘progressors’ and would have lost comparability with other published models of progression which examined a 12-month window.\textsuperscript{13,14,18}

To distinguish between non-progressors and patients with insufficient data to describe their status, data requirements were set up as detailed in table 1 footnote (*).

Univariate analysis
The univariate analysis compared progressors and non-progressors according to patient characteristics using the Kruskal-Wallis (for continuous variables) or Fisher’s test (for categorical variables). To characterise the progression of skin thickening according to autoantibody status, those same tests assessed differences in distribution for certain features (such as disease duration and mRSS peak) between autoantibody groups.

| Table 1 Characteristics of progressors and non-progressors according to clinical features and autoantibody status (at baseline) |
|---------------------------------|------------------|-------------------|-----------------|-----------------|----------------|
| Characteristics                | Progressor, n=66 (22.5%) | Non-progressor, n=227 (77.5%) | P               | Total, n=293 (100%) | Missing at baseline, n (%) |
| mRSS (0–51)                    | 19 (16–23)        | 21 (16–27)        | 0.030           | 21 (16–26)       | 0 (0)          |
| Months since onset of skin thickening | 8.1 (4.7–16.0) | 12.6 (8.1–22.0) | 0.001           | 12.0 (7.0–21.0) | 14 (4.8)       |
| Pulmonary fibrosis, n (%)      | 9 (13.6)          | 31 (13.7)         | 1               | 40 (13.7)        | 0 (0)          |
| FVC (% predicted)              | 87.5 (72.0–101.0) | 91.0 (75.0–102.0) | 0.129           | 90.0 (75.0–102.0) | 16(5.5)        |
| DLCO (% predicted)             | 62.8 (49.0–78.5)  | 66.0 (52.0–79.0)  | 0.455           | 65.0 (50.0–79.0) | 31 (10.6)      |
| Pulmonary hypertension, n (%)  | 1 (1.5)           | 19 (8.4)          | 0.054           | 20 (6.8)         | 1 (0.3)        |
| Antitopoisomerase (anti-Scl70 (TOPO), n (%) | 30 (46.2) | 84 (38)          | 0.252           | 114 (39.9)       | 7 (2.4)        |
| Anti-RNA polymerase III (Pol3), n (%) | 14 (25.9) | 34 (18.5)       | 0.249           | 48 (20.2)        | 55 (18.8)      |
| Anticentromere (ACA), n (%)    | 5 (7.7)           | 14 (6.4)          | 0.777           | 19 (6.7)         | 8 (2.7)        |
| No autoantibodies (TOPO, Pol3 or ACA), n (%) | 7 (12.7) | 52 (28.4)       | 0.020           | 59 (24.8)        | 55 (18.8)      |

Median (IQR) unless otherwise indicated.

P values refer to the Kruskal-Wallis (for continuous variables) or Fisher’s test (for categorical variables).

This table compares the distribution of patient characteristics at baseline between progressors and non-progressors, using the subset of 293 for whom the progression status is known. To distinguish between non-progressors and patients with insufficient data to describe their status, data requirements were set up. If progression was not detected using all data from the first >12±3 months, patients needed at least two data points to be considered non-progressors: one at baseline and another at least 5 months after baseline. Otherwise, we considered there were not enough data to ascertain their status. The 5-month limit was chosen so that all visits in the vicinity of the 6-month study mark could be counted.

The presence of pulmonary hypertension was not included as a variable in prediction models for progression. Only one patient had pulmonary hypertension and progressed.

Thus, a prediction model using mRSS, duration of skin thickening, an mRSS/duration interaction and the presence of pulmonary hypertension was too restrictive: no combinations of mRSS and duration of skin thickening enabled patients with pulmonary hypertension to progress.

DLCO, diffusing capacity for carbon monoxide; FVC, forced vital capacity; mRSS, modified Rodnan skin score.
Consequently, patients were only classed according to their Pol3 positivity rather than having indicator variables for each autoantibody (see note (1) in figure 3 and online supplementary table S2).

RESULTS

Univariate analysis: associates of mRSS progression and autoantibody status

The characteristics of mRSS progression are summarised in figure 1, including the increase in mRSS and the peak reached. During the study, the median number of skin scores recorded for each patient was 7 over a median follow-up of 23.4 months. There were 160 patients who had an increase in mRSS (of any magnitude) during the study (149 during the first 12 (±3) months).

Characteristics of progressors versus non-progressors

Out of 326 patients recruited at baseline, based on the retained progression criterion, 66 (22.5%) progressed and 227 (77.5%) did not (table 1). Progression status could not be assessed in 33 patients: 16 had no postbaseline skin scores and 17 did not fulfil the data requirements to ascertain progression status (see footnote (*) of table 1). Among those 33 patients with unknown status, 12 (36.4%) died during the analysis period.

At the time of recruitment, progressors had shorter disease duration than those who did not progress (median 8.1 vs 12.6 months (P=0.001)).

In addition, progressors tended to start with lower skin scores, median mRSS of 19 units, compared with 21 for non-progressors (P=0.030). Nevertheless, 30.3% of progressors started with mRSS >22 units and 15.2% with mRSS >25 units (online supplementary figure S1).% of progressors started with mRSS >22 units and 15.2% with mRSS >25 units (online supplementary figure S1).

Characteristics of mRSS progression according to autoantibody status

Out of the 326 patients, 124 were TOPO+, 50 were Pol3+, 20 were ACA+, 2 were TOPO+/ACA+, 68 were autoantibody-negative and 62 could not have their status determined: in 51 cases, this was because the Pol3 test was not done (unavailable in some centres) and the patient had neither TOPO nor ACA antibodies (table 2).

At baseline, Pol3+ patients had higher mRSS than patients in the other autoantibody groups (P=0.003) despite similar disease durations (P=0.593) (table 2).

There was a trend for Pol3+ patients to be more likely to progress than the other subgroups: 29.2% were progressors compared with 11.9% for the ‘no autoantibody’ group (P=0.105) (table 2). Pol3+ patients experienced higher increases in mRSS between baseline and peak: median increase of 7 units, compared with 3 for the ‘no autoantibody’ group (P=0.059) (table 2). Combined with their higher mRSS starting point, this results in Pol3+ patients having the highest peaks of all autoantibody groups with a median peak of 35 units (P=0.001) (table 2).

In terms of the speed of progression following onset, Pol3+ patients had the lowest observed median time to peak at 16.3 months (P=0.199) (table 2).

Predictive models of mRSS progression in first year of follow-up

Univariate and multivariate predictive models

Online supplementary table S2 and figure 2 show the values associated with the ROC curves for the multiple models tested, and online supplementary table S3 displays the details of different models.
selected logistic models to predict progression and the regression outputs. As a single predictor for progression, mRSS performed poorly with an AUC of 0.588 (95% CI 0.515 to 0.661). Duration of skin thickening performed better on its own, with an AUC of 0.634 (95% CI 0.553 to 0.715). A model combining mRSS, disease duration and an interaction between the two improved those univariate performances, with an AUC of 0.666 (95% CI 0.597 to 0.736). In addition, that model had a high 73.4% sensitivity, alongside its 57.2% specificity, and accurately predicted 60.9% of cases.

The interaction between mRSS and disease duration indicated that future progressors presented at their first visit with earlier disease and lower skin scores, and that higher skin score usually had to be compensated by lower disease duration for progression to occur (figure 3). Graphically, this could be identified by noting that, in figure 3 (model A), the points indicating progressors were mostly contained within the triangular lower half of a rectangle.

Adding an indicator variable for Pol3 positivity induced further gains in the model (already including mRSS, duration and their interaction), yielding an AUC of 0.711 (95% CI 0.633 to 0.790), 60.4% sensitivity, 74.2% specificity and accurately predicting 71% of cases (online supplementary table S2).

By graphical assessment, model A appeared to be better calibrated than the one including only mRSS, and model B appeared to improve on model A (online supplementary figure S2). The predicted probabilities for models A and B are summarised in online supplementary figures S3 and S4.

Properties of predictive models and application in practice
Two models described above were retained: model A and model B, which also includes Pol3+ status (figure 3, online supplementary table S2). Their ROC curves are shown in figure 2, and each curve yielded an optimal point nearest to the top-left corner, representing a threshold probability of progression. If a patient’s predicted probability was above this threshold, it was predicted that she/he would progress. Thus, for each level of disease duration at baseline, there corresponded an entry mRSS under which a patient met that threshold (summarised and plotted in table S2).

![Figure 2](https://example.com/figure2.png)

**Figure 2** ROC of three selected models. Three ROC curves summarise the predictive power of three different models by plotting sensitivity with respect to 100-specificity. For each model/ROC curve, there is an optimal point (the one closest to the top-left corner) that corresponds to a threshold of predicted probability of progression. For each model, patients with a predicted probability above that threshold are predicted to progress. AUC, area under curve; mRSS, modified Rodnan skin score; Pol3, anti-RNA polymerase III; PPV, positive predictive value; ROC, receiver operating characteristic.
Clinical and epidemiological research

Figure 3  Rules for selecting progressive patients according to two selected models. According to each model, in order to select progressive patients, they should be selected from the area under each relevant curve. These curves are superposed over a plot of the baseline mRSS of patients with respect to their duration of skin thickening, with progressors (of at least 5 units and 25%) being highlighted. Notes to the figure: (1) Analysing patients separately according to all autoantibody groups (TOPO, Pol3, ACA, ‘no autoantibodies’) was avoided because of the small number of ACA+ patients who could be included (n=16). Another possible approach was the inclusion of two indicator variables: one for Pol3+ and another for TOPO+, meaning that ACA+ and ‘no autoantibody’ patients formed the reference group, for which the resulting model proved too restrictive: only 2 out of 75 patients in the reference group were predicted to progress. Another reason for considering Pol3 patients separately was that they were suspected from preliminary analysis to be the most clinically different group, and stratifying by Pol3 status produced a higher AUC than doing so by TOPO status. (2) Each prediction model is based on a logistic regression model, where the outcome for patient i is **Y** _i_ = progression and _X_ are a selection of covariates. Using ROC curve analysis, each model has an optimal **p** _*_ for which, if **P** _r_ (**Y** _i_ = 1|_X_) > **p** _*_ , the patient is predicted to progress. Each frontier in the graphs above corresponds to the combination of mRSS and disease duration points, for which **P** _r_ (**Y** _i_ = 1|_X_) = **p** _*_  in the domains where both predictors are defined. Therefore, if a patient is in the area under the relevant curve, she/he is predicted to progress according to the model ACA, anticitromere; AUC, area under curve; mRSS, modified Rodnan skin score; Pol3, anti-RNA polymerase III; ROC, receiver operating characteristic; TOPO, topoiso merase.

<table>
<thead>
<tr>
<th>If duration of skin thickening is (months)</th>
<th>mRSS should be</th>
<th>Model A</th>
<th>Model B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(units) or less</td>
<td>AUC: 0.666</td>
<td>AUC: 0.711</td>
</tr>
<tr>
<td></td>
<td>mRSS should be</td>
<td>Sensitivity: 73.4%</td>
<td>Sensitivity: 60.4%</td>
</tr>
<tr>
<td></td>
<td>(units) or less</td>
<td>Specificity: 57.2%</td>
<td>Specificity: 74.2%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PPV: 33.8%</td>
<td>PPV: 41.0%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NPV: 87.9%</td>
<td>NPV: 86.3%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Accuracy: 60.9%</td>
<td>Accuracy: 71.0%</td>
</tr>
</tbody>
</table>

Table 3 Rules for selecting progressive patients according to two selected models

If applying this selection rule (model A) to the ESOS cohort, 139 patients (49.8% of the 279 patients included in the model) would be predicted to progress, of whom 47 actually did in the year following baseline (PPV: 33.8%). Conversely, 140 were predicted not to progress, of whom 123 did not (negative predictive value (NPV): 87.9%), whereas 17 (12.1%) did. Model B is used in the same way as model A, but accounting for Pol3 status. The curves in figure 3 (summarising selection criteria)
shift across the diagonal axis to reflect that Pol3+ patients have a higher propensity to progress during the first year compared with Pol3− patients.

Model B had a higher accuracy than model A (71.0%–60.9%). Model B, which was more specific, was also more restrictive: only 78 patients were predicted to progress, of whom 32 actually did (PPV: 41.0%). Therefore this model identified a ‘high risk’ subset of patients with a proportion of progressors 1.8 times higher than the overall cohort. In model B, 153 patients were predicted not to progress, of whom 132 did not (NPV: 86.3%), whereas 21 did (13.7%).

The predictive power of model B was particularly strong for Pol3+ patients, for whom the sensitivity was 100% and the specificity was 70.6%.

**DISCUSSION**

The major strength of this study compared with previous recent analyses of mRSS is that this was a well-defined cohort with prospective assessment of mRSS by experienced assessors. Assessments every 3 months provide detailed insight into disease trajectory and (burden) for the practising clinician. For the clinical trialist, the time frames examined were comparable to those of recent and current RCTs, which include assessments at 24 weeks (and less) as well as at 12 months. In addition, as the data set was derived from an observational study of standard current treatments for skin, we expect that our findings are generalisable to current or future clinical trials of skin therapy in dcSSc. This is especially relevant since current trials often permit standard background therapy, as used in ESOS, to which a novel agent may be added. The key finding here was the development of a predictive model for mRSS (disease) progression which had an accuracy of 60.9% (model A), achieved by recognising that the initial skin score is a poor predictor of progression on its own and that prediction is improved by simultaneously accounting for disease duration. By including autoantibodies in this analysis, the model improved and reached an accuracy of 71.0% (model B).

When recruiting patients into clinical trials of rare diseases, any algorithm should not be too restrictive. Higher sensitivity was favoured because it was considered more appropriate to have more inclusive models at the risk of mischaracterising non-progressors as progressors. We believe that model A will be the more useful for studies aiming for cohort enrichment, while model B will help to identify patients at higher risk for mRSS progression in a clinical setting. The use of the second model to inform patient selection into RCTs could risk over-representing Pol3+ patients, for whom the criteria to predict progression are less strict, thus yielding a sample not reflecting the overall dcSSc population.

Other ‘take home messages’ were that skin score progression did occur in some patients who presented with high baseline mRSS (25 or higher), although this tended to be compensated by shorter disease durations, that Pol3+ patients tended to reach their peak mRSS earlier than other patients, and that this peak was much higher than for patients with other (or no) autoantibodies. Patients without TOPO, Pol3 or ACA autoantibodies had smaller increases in mRSS and lower peak skin scores. Our 3-monthly data allowed us to capture peaks in mRSS, which would have been ‘smoothed over’ in other studies because of less frequent data. Had we only recorded baseline and 12-month data (two observations), 53% of our cases of progression would have been missed.

Taking into account peak mRSS in defining progression (as opposed to considering only baseline and 12-month data) was therefore a major difference between ESOS and the study by Maurer et al, who also looked at prediction of extent of skin thickening in patients with systemic sclerosis in a study of 637 patients from the EULAR Scleroderma Trials and Research group (EUSTAR) cohort and an average follow-up time between visits of 12 months (compared with 3-monthly in ESOS). Disease duration was 42 months (therefore substantially longer than in the ESOS cohort) and baseline mRSS was 17 units (compared with a mean of 22.1 units for ESOS). ESOS had 22.5% of progressors compared with EUSTAR’s 9.7%, possibly because ESOS was an earlier cohort and the 3-monthly follow-ups made any disease progression more likely to be detected. Maurer et al established that lower mRSS and shorter disease duration were associated with more progressive cases, as confirmed here, although we accept that the two studies are not strictly comparable given the differing time frames of defining ‘progressors’.

However, if we do apply a 22-unit mRSS cut-off point to the ESOS cohort, its size would decrease from 326 to 189, and the share of progressors (among those with known status) would only increase from 22.5% to 26.4%. In contrast, that share (PPV) rises to 33.8% with model A and 41.0% with model B. Like Maurer et al we found that skin score alone was a poor predictor for progression and that other factors including disease duration should also be considered.

Dobrota et al also looked at patterns of mRSS changes but focused on regression rather than progression, validating that a low baseline mRSS predicts progression.

Our study has certain limitations. It can be very difficult to gauge onset of skin thickening (in 18 (5.5%) patients we had no data on duration of skin thickening at baseline, other than that this was under 3 years). It is likely that in some patients (especially those who steadily improve after baseline), peak mRSS occurred prior to study entry. Also, unlike the EUSTAR study, we have not externally validated the model, and this will be an important step before using the models widely. Among the patients with unknown progression status, 36.4% died, thus potentially inducing bias (it is likely they had progressive disease) but also mirroring the attrition occurring in clinical trials. In model B, missing data in autoantibodies (19.6%) reduce the predictive power.

In conclusion, among patients with early dcSSc, those with shorter disease duration and lower mRSS are most likely to be ‘progressors’ with a trade-off between the two factors, and patients who are Pol3+ have the highest mRSS peaks and tend to reach peak mRSS earliest, providing a valuable message for clinicians that patients with short disease duration and Pol3+ must be especially closely monitored. Other prediction models for progressive skin thickening were derived. The model incorporating Pol3 (model B) more accurately identifies high-risk patients, but risks being too restrictive for patient selection into trials and over-representing Pol3+ patients. Both models were more flexible (for a given skin score) and more accurate than a ‘22 mRSS’ cut-off model and may offer advantages for cohort enrichment in clinical trials to ensure that the most informative patients are included.

**Author affiliations**

Centre for Musculoskeletal Research, Salford Royal NHS Foundation Trust, Manchester Academic Health Science Centre, University of Manchester, Manchester, UK

NIHR Manchester Musculoskeletal Biomedical Research Centre, Central Manchester NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK
Clinical and epidemiological research

3Centre for Musculoskeletal Research, The University of Manchester, Manchester Academic Health Science Centre, Manchester, UK
4Department of Rheumatology, Lund University, Lund, Sweden
5Service de Medicine Interne, Hôpital Cochin, Centre de Référence pour les Vasculaires Nécrosantes et la Sclérodémie Systémique, Université Paris Descartes, Paris, France
6Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK
7University of Manchester, Manchester, Greater Manchester, UK
8Department of Rheumatology and Immunology, Medical Center, University of Pecs, Pecs, Hungary
9Department of Internal Medicine III, University of Erlangen-Nuremberg, Erlangen, Germany
10Department of Rheumatology, University of Zurich, Zurich, Switzerland
11Royal Free London NHS Foundation Trust, London, UK
12Rehabilitation Services, Salford Royal NHS Foundation Trust, Salford, UK
13Department of the Rheumatic Diseases, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands
14Rheumatology and 2 Department, 'Grigore T. Popa’ University of Medicine and Pharmacy, Clinical Rehabilitation Hospital, Iasi, Romania
15UCL Division of Medicine, Centre for Rheumatology and Connective Tissue Diseases, London, UK
16Unité Clinique de Medicine Interne, Maladies Auto-immunes et Pathologie Vasculaire, U130, Hôpital Saint-Louis, AP-HP Assistance Publique des Hôpitaux de Paris, INSERM UMRS 1160, Paris Denis Diderot University, Paris, France
17Jewish General Hospital, Lady Davis Institute and McGill University, Montreal, Canada
18Department of Experimental and Clinical Medicine, Division of Rheumatology ADUC, University of Florence, Florence, Italy
19Shine Rheumatology Unit, Rambam Health Care Campus, Rappaport Faculty of Medicine, Haifa, Israel
20Rheumatology Unit, Oslo University Hospital Rikshospitalet, Oslo, Norway
21Queen Elizabeth Hospital Birmingham, UHB Foundation Trust, Birmingham, UK
22St Vincent’s Hospital, Melbourne, Victoria, Australia
23Department for Dermatology, University of Cologne Kerpener Str, Cologne, Germany
24Cambridge University NHS Hospital Foundation Trust, Cambridge, UK
25Department of Internal Medicine, Hôtel-Dieu Hospital, University of Nantes, Nantes, France
26University of Liverpool, Aintree University Hospital, Liverpool, UK
27Service de Médecine Interne, Hôpital Bretonneau, Tours, France
28Centre for Rheumatic Diseases, Royal Infirmary, Glasgow, UK
29Sheffield Teaching Hospitals, Sheffield, UK
30Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds and NIHR Leeds Musculoskeletal Biomedical Research Centre, Leeds Teaching Hospitals NHS Trust, Leeds, UK
31Stanford University, Stanford, California, USA
32University of Belgrade School of Medicine, Institute of Rheumatology, Belgrade, Serbia
33Clinical and Academic Rheumatology, North Bristol NHS Trust, Bristol, UK
34Nottingham University Hospitals NHS Trust and Nottingham NHS Treatment Centre, Nottingham, UK
35Peter Maddison Rheumatology Centre, Llandudno, UK
36Queens Hospital, Romford, UK
37University of Copenhagen, Copenhagen Lupus and Vasculitis Clinic, Center for Rheumatology and Immunology, Rigshospitalet, Copenhagen, Denmark
38Norwich Medical School, University of East Anglia, Norwich, UK
39Royal National Hospital for Rheumatic Diseases, Bath, UK
40Department of Rheumatology and Clinical Immunology, Justus-Liebig University Giessen, Bad Nauheim, Germany
41Department of Rheumatology, University of Lübeck, Lübeck, Germany
42Department of Rheumatology and Clinical Immunology, University Hospital Charité Berlin, Berlin, Germany
43Department of Rheumatology, Royal Perth Hospital, Perth, Australia
44Servicio de Reumatologia, Hospital Universitario 12 de Octubre, Madrid, Spain
45Internal Medicine Unit, Limoges University Hospital, Limoges, France
46Centre National de Référence Maladies Systémiques et Auto-immunes Rares, Département de Médecine Interne et Immunologie Clinique, Université de Lille, Lille, France
47Gateshead Hospitals Foundation Trust, Gateshead, UK
48Department of Internal Medicine, Division of Rheumatology, Istanbul University, Istanbul Medical Faculty, Istanbul, Turkey
49Hille Rheumatic Diseases Unit, Whytman’s Brae Hospital, Kirkcaldy, UK
50Department of Rheumatology and Clinical Immunology, UMC Utrecht, Utrecht, The Netherlands
51James Cook University Hospital, Middlesbrough, UK
52Rheumatology Unit, Royal Adelaide Hospital, and Discipline of Medicine, University of Adelaide, Adelaide, Victoria, Australia
53Department of Rheumatology and Inflammation Research, The Sahlgrenska Academy at Gothenburg University, Gothenburg, Sweden
54Monash Health and Department Medicine, Monash Centre for Inflammatory Diseases, Monash University, Melbourne, Victoria, Australia
55Department of Internal Medicine, Hôpital Edouard Herriot, Lyon, France
56Department of Internal Medicine, Foundation Hospital Saint Joseph, Marseille, France
57Cannock Chase Hospital, Cannock, UK
58St Vincent’s University Hospital, Dublin, Ireland
59Department of Internal Medicine, Centre Hospitalier Lyon Sud, Lyon, France
60Internal Medicine, Ambroise Paré Hospital, Boulogne-Billancourt, France

Acknowledgements We are grateful to Dr Holly Ennis for study set-up and for project coordination during the earlier phases of the study, and also to the members of the independent oversight board: Stephen Cole, Dinesh Khanna and Frank Wollheim.

Contributors AL, ML, RH, LM, AJS, EB, LaC, JHWD, OD, KF, WIG, RO, MCV and CPD were members of the European Sclerodema Observational Study (ESOS) Steering Committee and designed the ESOS study. SeP and ML were responsible for the statistical analysis. ALH, RH, LaC, JHWD, OD, MV, CoA, VHO, DF, MH, MM-C, AB-G, OM, PJ, ACI, WS, PM, FCH, Cha, MEA, ED, RM, MA, MHB, LoC, NSD, HG, PL, YA, KC, SL, AIM, NM, UM-L, GR, MB, JR, PEC, A-LF, EH, JH, IM, JS, JMvL, SaP, SiP, AR, JS, BC, CS, TDJ, GV, CG, ST and CPD were principal investigators at the different sites and recruited patients. XP and GD were study coordinators. ALH, SeP, ML, RH, LM, AJS and CPD wrote the draft report, and all authors reviewed the report and provided comments.

Funding ESOS was funded by a grant from the EULAR (European League Against Rheumatism) Orphan Disease Programme. Additional funding from Sclerodema and Raynaud’s UK allowed a 1-year extension of the study.

Competing interests ALH has done consultancy work for Actelion, served on a Data Safety Monitoring Board for Apricis, received funding research and speaker’s fees from Actelion, and speaker’s fees from GSK. JHWD has consultation relationships and/or has received research funding from Actelion, BMS, Celgene, Bayer Pharma, Boehringer Ingelheim, JB Therapeutics, Sanofi-Aventis, Novartis, UCB, GSK, Array BioPharma, Active Biotech, Galapagos, Inventiva, Medac, Pfizer, Anamar and RuiYi, and is stock owner of 4D Science. OD has received consultancy fees from 4D Science, Actelion, Active Biotech, Bayer, Biogenidesc, BMS, Boehringer Ingelheim, EpiPhann, Ergonex, espeRare Foundation, Genentech/Roche, GSK, Inventiva, Lilly, Medac, Medimmune, Pharmyciclis, Pfizer, Serodaphax, Sinouxa and UCBB, and received research grants from Actelion, Bayer, Boehringer Ingelheim, Ergonex, Pfizer and Sanofi, and has a patent mir-29 for the treatment of systemic sclerosis licensed. WIG has received teaching fees from Pfizer. CA has served as a consultant for AbbVie, Pfizer, Roche, UCB, MSD, BMS and Novartis, and has received research funding and speaker fees from AbbVie, Pfizer, Roche, UCB, MSD, and received research funding and speaker fees from AbbVie, Pfizer, Roche, UCB, MSD and BMS. SP has received research grants from Actelion Pharmaceuticals Australia, Bayer, GlaxoSmithKline Australia and Pfizer, and speaker fees from Actelion. AR receives funding from AstraZeneca. CPD has done consultancy for GSK, Actelion, Bayer, Inventiva and Merck-Serono, received research grant funding from GSK, Actelion, CSL Behring and Inventiva, received speaker’s fees from Bayer and given trial related funding to Merck-Serono.

Ethics approval The ethics committee of each participating centre approved the study.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/ © Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES
Clinical and epidemiological research


