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

Ariane L. Herrick, Sebastien Peytrignet, Mark Lunt, Xiaoyan Pan, Roger Hesselstrand, Luc Mouthon, Alan Silman, Graham Dinsdale, Edith Brown, Laszlo Czirjak, et al.

OPEN ACCESS

EXTENDED REPORT

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

Ariane L Herrick,[1,2] Sebastien Peytrignet,[3] Mark Lunt,[3] Xiaoyan Pan,[3] Roger Hesselstrand,[4] Luc Mouthon,[5] Alan J Silman,[6] Graham Dinsdale,[1] 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 C 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] Paresh Jobanputra,[21] Alison C Jordan,[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 S 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] Susanna M 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-2017-211912).

For numbered affiliations see end of article.

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

Received 8 June 2017
Revised 13 November 2017
Accepted 19 November 2017
Published Online First 6 January 2018

Check for updates

To cite: Herrick AL, Peytrignet S, Lunt M, *et al*. *Ann Rheum Dis* 2018;**77**:563–570.

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 shorter 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 (Pol3+) 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 Pol3 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,[5 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,[13 14] with 22 as a proposed level.[13] 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)[15] 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 prediction 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[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,[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'), anticentromere ('ACA')) and presence of visceral organ involvement were recorded for all patients.[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,[17] thus enabling model comparisons.[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.[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.[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 (49.0–76.5)	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.

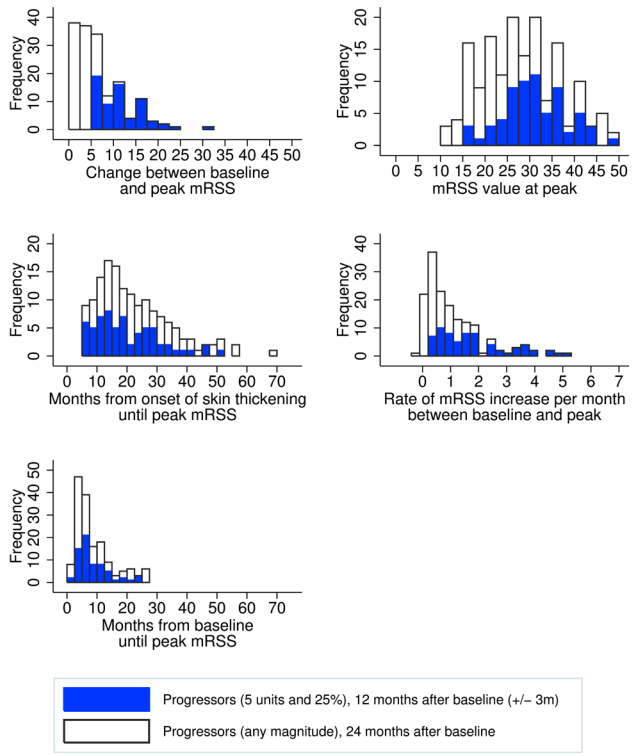

Figure 1 Characteristics of mRSS progression. The five histograms describe modified Rodnan skin score (mRSS) progression for all patients whose skin score during the study ever increases beyond their baseline level (n=160) and for those whose progression satisfies the 5-unit and 25% increase rule during the first 12 months (±3 months) (n=66). Here, histograms summarise the distribution of changes between baseline and peak mRSS, the mRSS value at its peak, the time elapsed between the onset of skin thickening and the recorded peak, the rate of mRSS increase per month between baseline and peak, and the time elapsed between baseline and the recorded peak. The rate of mRSS progression (in units/month) was computed by specifying individual simple linear regressions of mRSS according to time, between baseline and peak.

If a patient tested positive for an autoantibody, we assumed they did not have the other two if those data were missing. Patients with more than one autoantibody were excluded from our models.

Predictive models of mRSS progression

Logistic regressions were fitted to predict progression using baseline characteristics. Associations with progression (including those in table 1) and the predictive performance of single predictors were assessed to select potential covariates, resulting in different models.

Those models were then compared on the basis of the area under curve (AUC), sensitivity, specificity, positive predictive value (PPV) and accuracy at each curve's optimal point—but also according to their simplicity and interpretability. Predictive ability can be optimistic when assessed using its own model-generating data. An additional optimism-adjusted bootstrapped AUC was therefore also computed and reported in online supplementary table S2, suggesting modest corrections.[20] Calibration plots for the retained models were also assessed.[21]

When including autoantibodies in predictive models, certain specifications produced predicted progression probabilities that were too low for certain subgroups and were thus avoided because they were considered too restrictive to apply in practice. 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).

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

Table 2 Characteristics of mRSS progression according to autoantibody status

Autoantibody make-up	Anti-TOPO-isomerase (anti-Scl70) (TOPO) (TOPO+) (Pol3– or N/A) (ACA– or N/A) n=124 (47.3%)	Anti-RNA polymerase III (Pol3) (TOPO– or N/A) (Pol3+) (ACA– or N/A) n=50 (19.1%)	Anticentromere (ACA) (TOPO– or N/A) (Pol3– or N/A) (ACA+) n=20 (7.6%)	None (TOPO–) (Pol3–) (ACA–) n=68 (26.0%)	Total n=262 (100%)	P	Missing at baseline, n (%)
mRSS at baseline (0–51)	19 (15–25.5)	24 (19-31)	20 (17–24.5)	20 (16–24)	20 (16–26)	0.003	0 (0)
mRSS peak*	26 (19.5–33.5)	35 (26–40)	29 (26–35)	24.5 (17.5–29)	27 (21–34.5)	0.001	0 (0)
Difference in mRSS between baseline and peak*	5 (3–10.5)	7 (3–10)	4 (4–11)	3 (1.5–7)	5 (3–10)	0.059	0 (0)
Months since onset of skin thickening (at baseline)	12.6 (6.2–21.6)	11.2 (7.8–17.9)	14.9 (5.4–24.0)	12.6 (9.2–21.9)	12.6 (7.3–21.5)	0.593	10 (3.8)
Months until peak since onset of skin thickening*	21.0 (12.9–31.6)	16.3 (12.9–21.4)	29.3 (15.5–35.7)	20.1 (13.2–32)	19.0 (12.9–30.0)	0.199	5 (3.9)
Months until peak since baseline*	6.4 (4.0–14.4)	5.8 (2.9–12.0)	6.5 (2.9–9.2)	6.0 (3.1–11.6)	6.2 (3.2–12.1)	0.329	0 (0)
Progressor (5 points and 25% according to baseline)	29 (25.9)	14 (29.2)	4 (23.5)	7 (11.9)	54 (22.9)	0.105	26 (9.9)

Median (IQR) unless otherwise indicated.
P values refer to the Kruskal-Wallis (for continuous variables) or Fisher's test (for categorical variables).
*For these comparisons, an unrestricted definition of progression was used, meaning that all 160 patients in the cohort with mRSS progression of any magnitude were initially considered but only 128 of those could be included because of patients with missing autoantibody data.
This table includes comparisons of patient characteristics at baseline between different autoantibody groups, using the subset of 262 patients for whom the autoantibody status could be assessed.
mRSS, modified Rodnan skin score; N/A, not available.

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.

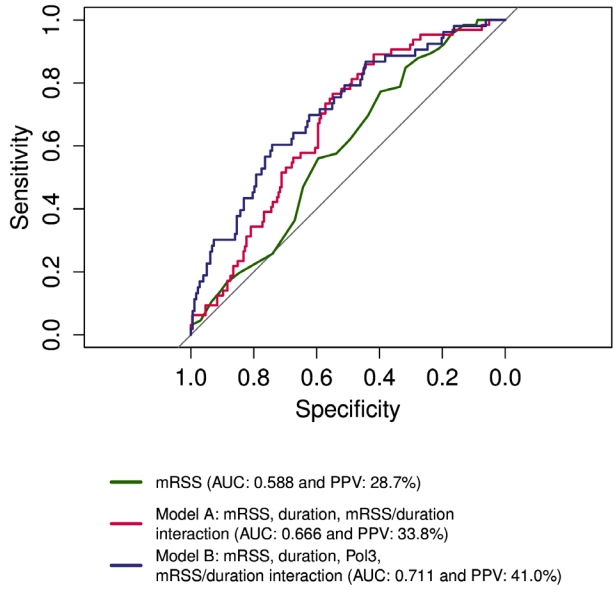

Figure 2 ROC of three selected models. Three ROC curves summarise the predictive power of three different models by plotting sensitivity with respect to 100-specificity. For each model/ROC curve, there is an optimal point (the one closest to the top-left corner) that corresponds to a threshold of predicted 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.

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 3

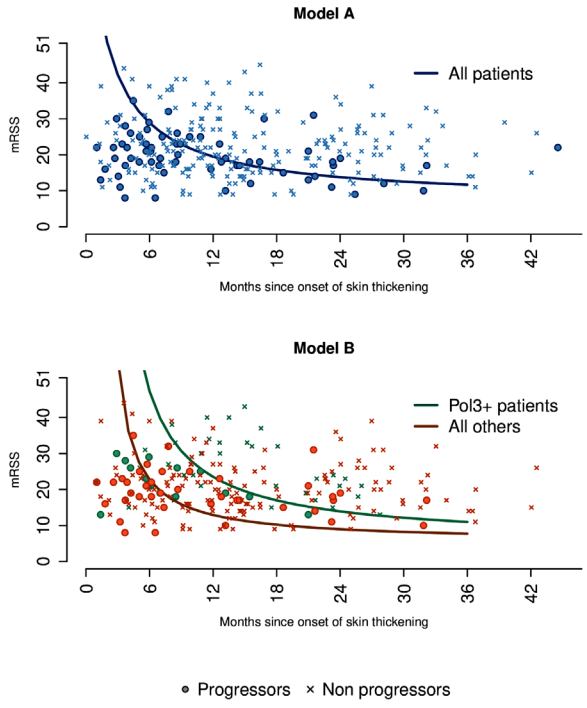

Figure 3 Rules for selecting progressive patients according to two selected models. According to each model, in order to select progressive patients, they should be selected from the area under each relevant curve. These curves are superposed over a plot of the baseline mRSS of patients with respect 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 $Y_i = \text{progression}$ and X are a selection of covariates. Using ROC curve analysis, each model has an optimal p^* for which, if $\widehat{Pr}(Y_i = 1|X) > p^*$, the patient is predicted to progress. Each frontier in the graphs above corresponds to the combination of mRSS and disease duration points, for which $\widehat{Pr}(Y_i = 1|X) = p^*$ 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, anticentromere; AUC, area under curve; mRSS, modified Rodnan skin score; Pol3, anti-RNA polymerase III; ROC, receiver operating characteristic; TOPO, topoisomerase.

and figure 3 for the two models). For instance, using the selection rule produced by model A, a patient recruited at 9 months of skin thickening would be predicted to progress if mRSS was between 0 and 23 units. However, if a patient presented at 6 months, the mRSS would be allowed to go as high as 29.

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

	Model A	Model B	
	AUC: 0.666 Sensitivity: 73.4% Specificity: 57.2% PPV: 33.8% NPV: 87.9% Accuracy: 60.9%	AUC: 0.711 Sensitivity: 60.4% Specificity: 74.2% PPV: 41.0% NPV: 86.3% Accuracy: 71.0%	
If duration of skin thickening is (months)	All patients mRSS should be (units) or less	Pol3+ patients mRSS should be (units) or less	All others mRSS should be (units) or less
1	51	51	51
2	51	51	51
3	43	51	51
4	37	51	37
5	33	51	28
6	29	48	23
7	27	40	20
8	25	35	18
9	23	31	16
10	22	28	15
11	21	26	14
12	20	24	13
13	19	22	13
14	18	21	12
15	18	20	12
16	17	19	11
17	17	18	11
18	16	17	11
19	16	17	10
20	15	16	10
21	15	16	10
22	15	15	10
23	15	15	10
24	14	14	9
25	14	14	9
26	14	14	9
27	14	13	9
28	13	13	9
29	13	13	9
30	13	13	9
31	13	12	9
32	13	12	9
33	13	12	8
34	12	12	8
35	12	12	8
36	12	11	8

For each duration, the required mRSS level is rounded above to the nearest integer to reflect real mRSS values.
AUC, area under curve; mRSS, modified Rodnan skin score; PPV, positive predictive value.

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.[12 22] 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*[13] 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*[13] 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'.[13]

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*[13] 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*[14] 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,[13] 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. Two 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
[1]Centre for Musculoskeletal Research, Salford Royal NHS Foundation Trust, Manchester Academic Health Science Centre, University of Manchester, Manchester, UK
[2]NIHR Manchester Musculoskeletal Biomedical Research Centre, 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 Medicine Interne, Hôpital Cochin, Centre de Référence pour les Vascularites Nécrosantes et la Sclérodermie Systémique, Université Paris Descartes, Paris, France
[6]Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK
[7]University of Manchester, Manchester, Greater Manchester, UK
[8]Department of Rheumatology and Immunology, Medical Center, University of Pecs, Pecs, Hungary
[9]Department of Internal Medicine III, 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, Iasi, Romania
[15]UCL Division of Medicine, Centre for Rheumatology and Connective Tissue Diseases, London, UK
[16]Unite Clinique de Medicine 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, Paris, France
[17]Jewish General Hospital, Lady Davis Institute and McGill University, Montreal, Canada
[18]Department of Experimental and Clinical Medicine, Division of Rheumatology AOUC, University of Florence, Florence, Italy
[19]Shine Rheumatology Unit, Rambam Health 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, Victoria, Australia
[23]Department for Dermatology, University of Cologne Kerpener Str, Cologne, 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, Tours, France
[28]Centre for Rheumatic Diseases, 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 Centre, Leeds Teaching Hospitals NHS Trust, Leeds, 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, Limoges, France
[46]Centre National de Référence Maladies Systémiques et Auto-immunes Rares, Département de Médecine Interne et Immunologie Clinique, Université de Lille, Lille, France
[47]Gateshead Hospitals Foundation Trust, Gateshead, UK
[48]Department of Internal Medicine, Division of Rheumatology, Istanbul University, Istanbul Medical Faculty, 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, Victoria, Australia
[53]Department of Rheumatology and Inflammation Research, The Sahlgrenska Academy at Gothenburg University, Gothenburg, Sweden
[54]Monash Health and Department Medicine, Monash Centre for Inflammatory Diseases, Monash University, Melbourne, Victoria, 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, Lyon, France
[60]Internal 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 ALH, ML, RH, LM, AJS, EB, LaC, JHWD, OD, KF, WJG, RO, MCV and CPD were members of the European Scleroderma 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, ACJ, WS, PM, FCH, ChA, MEA, ED, RM, MA, MHB, LoC, NSD, HG, PL, YA, KC, SJ, AJM, NM, UM-L, GR, MB, JR, PEC, A-LF, EH, JH, MI, JSM, JMvL, SaP, SuP, AR, JS, BC, CS, TS, DJV, CG, G-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 Scleroderma 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 Apricus, received research funding and speaker's fees from Actelion, and speaker's fees from GSK. JHWD has consultancy 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, Biogenidec, BMS, Boehringer Ingelheim, EpiPharm, Ergonex, espeRare Foundation, Genentech/Roche, GSK, Inventiva, Lilly, Medac, Medimmune, Pharmacyclics, Pfizer, Serodapharm, Sinoxa and UCB, 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. WJG 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, BMS and Novartis. FCH has received research funding from Actelion. MEA has undertaken advisory board work and received honoraria from Actelion, and received speaker's fees from Bristol-Myers Squibb. NSD has done consultancy for AbbVie, Pfizer, Roche and MSD, and received speaker's fees from AbbVie, Boehringer-Ingelheim, Pfizer, Richter Gedeon, Roche and MSD. HG has done consultancy work and received honoraria from Actelion. UM-L is funded in part by EUSTAR, EULAR and the European Community (Desscipher programme). JMvL has received honoraria from Eli Lilly, Pfizer, Roche, 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 advice to Merck-Serono.

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

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 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.

2 Clements PJ, Hurwitz EL, Wong WK, *et al*. Skin thickness score as a predictor and correlate of outcome in systemic sclerosis: high-dose versus low-dose penicillamine trial. *Arthritis Rheum* 2000;43:2445–54.
3 Amjadi S, Maranian P, Furst DE, *et al*. Course of modified Rodnan skin score in systemic sclerosis clinical trials: analysis of 3 large multicenter, randomized clinical trials. *Arthritis Rheum* 2009;60:2490–8.
4 Steen VD, Medsger TA, Rodnan GP. D-Penicillamine therapy in progressive systemic sclerosis (scleroderma): a retrospective analysis. *Ann Intern Med* 1982;97:652–9.
5 Clements PJ, Furst DE, Wong WK, *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.
6 Stevens W, Vancheeswaran R, Black CM, *et al*. Alpha interferon-2a (Roferon-A) in the treatment of diffuse cutaneous systemic sclerosis: a pilot study. UK Systemic Sclerosis Study Group. *Br J Rheumatol* 1992;31:683–9.
7 Black CM, Silman AJ, Herrick AI, *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.
8 Denton CP, Merkel PA, Furst DE, *et al*. Recombinant human anti-transforming growth factor beta1 antibody therapy in systemic sclerosis: a multicenter, randomized, placebo-controlled phase I/II trial of CAT-192. *Arthritis Rheum* 2007;56:323–33.
9 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.
10 Khanna D, Clements PJ, Furst DE, *et al*. Recombinant human relaxin in the treatment of systemic sclerosis with diffuse cutaneous involvement: a randomized, double-blind, placebo-controlled trial. *Arthritis Rheum* 2009;60:1102–11.
11 van Laar JM, Farge D, Sont JK, *et al*. Autologous hematopoietic stem cell transplantation vs intravenous pulse cyclophosphamide in diffuse cutaneous systemic sclerosis: a randomized clinical trial. *JAMA* 2014;311:2490–8.
12 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.
13 Maurer B, Graf N, Michel BA, *et al*. Prediction of worsening of skin fibrosis in patients with diffuse cutaneous systemic sclerosis using the EUSTAR database. *Ann Rheum Dis* 2015;74:1124–31.
14 Dobrota R, Maurer B, Graf N, *et al*. Prediction of improvement in skin fibrosis in diffuse cutaneous systemic sclerosis: a EUSTAR analysis. *Ann Rheum Dis* 2016;75:1743–8.
15 Herrick AL, Pan X, Peytrignet S, *et al*. Treatment outcome in early diffuse cutaneous systemic sclerosis: the European Scleroderma Observational Study (ESOS). *Ann Rheum Dis* 2017;76:1207–18.
16 LeRoy EC, Black C, Fleischmajer R, *et al*. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. *J Rheumatol* 1988;15:202–5.
17 Khanna D, Furst DE, Hays RD, *et al*. Minimally important difference in diffuse systemic sclerosis: results from the D-penicillamine study. *Ann Rheum Dis* 2006;65:1325–9.
18 Domsic RT, Lucas M, Steen VD, *et al*. A clinical risk prediction model for skin thickness progression in early diffuse scleroderma. *Arthritis Rheum* 2016;68.
19 Wiese AB, Berrocal VJ, Furst DE, *et al*. Correlates and responsiveness to change of measures of skin and musculoskeletal disease in early diffuse systemic sclerosis. *Arthritis Care Res* 2014;66:1731–9.
20 Smith GC, Seaman SR, Wood AM, *et al*. Correcting for optimistic prediction in small data sets. *Am J Epidemiol* 2014;180:318–24.
21 Steyerberg EW, Vickers AJ, Cook NR, *et al*. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;21:128–38.
22 Khanna D, Denton CP, Jagerschmidt A, *et al*. SAR100842, an antagonist of lysophosphatidic acid receptor 1, as a potential treatment for patients with systemic sclerosis: results from a phase 2a study. *Arthritis Rheumatol* 2014;66:S387.