Original Research Article

Exploratory clinical subgroup clustering in systemic sclerosis: Results from the Indian Progressive Systemic Sclerosis Registry

Shery Susan Philip[1] (iD), Ramya Janardana[1], Padmanabha Shenoy[2] (iD),
Chengappa Kavadichanda[3], Devender Bairwa[3], Geetabali Sircar[4],
Parasar Ghosh[4], Anupam Wakhlu[5], Sumithra Selvam[6],
Dinesh Khanna[7] (iD), and Vineeta Shobha[1] (iD); for Indian Progressive
Systemic Sclerosis Registry (IPSSR)

Abstract

Objective: To conduct an exploratory cluster analysis of systemic sclerosis patients from the baseline data of the Indian systemic sclerosis registry.

Methods: Patients satisfying American College of Rheumatology-European League Against Rheumatism classification criteria for systemic sclerosis were included. The clusters formed using clinical and immunological parameters were compared.

Results: Of the 564 systemic sclerosis registry participants, 404 patients were included. We derived four clusters of which three were anti-topoisomerase I predominant and one was anti-centromere antibody 2 dominant. **Cluster 1** (*n*-82 (20.3%)) had diffuse cutaneous systemic sclerosis patients with the most severe skin disease, anti-topoisomerase I positivity, males, younger age of onset and high prevalence of musculoskeletal, vasculopathic and gastrointestinal features. **Cluster 2** (*n*-141 (34.9%)) was also diffuse cutaneous systemic sclerosis and anti-topoisomerase I predominant but with less severe skin phenotype than cluster 1 and a lesser prevalence of musculoskeletal, vasculopathic and gastrointestinal features. **Cluster 3** (*n*-119 (29.5%)) had limited cutaneous systemic sclerosis patients with anti-topoisomerase I positivity along with other antibodies. The proximal muscle weakness was higher and digital pitting scars were lower, while other organ involvement was similar between clusters 2 and 3. **Cluster 4** (*n*-62 (15.30%)) was the least severe group with limited cutaneous systemic sclerosis and anti-centromere antibody predominance. Age of onset was higher with low musculoskeletal disease and a higher presence of upper gastrointestinal features. The prevalence of interstitial lung disease was similar in the three anti-topoisomerase I predominant clusters.

Conclusion: With exploratory cluster analysis, we confirmed the possibility of subclassification of systemic sclerosis along a spectrum based on clinical and immunological characteristics. We also corroborated the presence of anti-topoisomerase I in limited cutaneous systemic sclerosis and the association of interstitial lung disease with anti-topoisomerase I.

[1]Department of Clinical Immunology and Rheumatology, St. John's Medical College Hospital, Bengaluru, India
[2]Centre for Arthritis and Rheumatism Excellence (CARE), Cochin, India
[3]Department of Clinical Immunology, Jawaharlal Institute of Postgraduate Medical Education and Research, Puducherry, India
[4]Department of Rheumatology and Clinical Immunology, Institute of Post-Graduate Medical Education & Research and S. S. K. M. Hospital, Kolkata, India
[5]Clinical Immunology and Rheumatology, Apollomedics Super Speciality Hospitals, Lucknow, India

[6]Division of Epidemiology and Biostatistics, St. John's Research Institute, Bengaluru, India
[7]Division of Rheumatology, Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA

Corresponding author:
Vineeta Shobha, Department of Clinical Immunology and Rheumatology, St. John's Medical College Hospital, 3rd Floor, Unit of Hope, Bengaluru 560034, Karnataka, India.
Email: vineeta.s@stjohns.in

Key message

1. The exploratory cluster analysis revealed four clusters, demonstrating the heterogeneity of systemic sclerosis.
2. Anti-topoisomerase I occurs in a significant proportion of limited cutaneous systemic sclerosis.
3. Anti-topoisomerase I is strongly associated with interstitial lung disease.

Introduction

Systemic sclerosis (SSc) is a systemic autoimmune disease characterized by small vessel vasculopathy, autoantibody production and fibroblast dysfunction leading to increased deposition of extracellular matrix. It is chronic, progressive and heterogenous in its course. Recently, the European League Against Rheumatism (EULAR) and the American College of Rheumatology (ACR) classification criteria for SSc have expanded the clinical phenotype and the autoantibody spectrum.[1] However, the subclassification of SSc remains a challenge owing to its heterogeneity. Even today, the most accepted and prevailing system of subclassification remains to be the 1988 LeRoy's dichotomous subsets revised in 2001, based on the extent of cutaneous involvement. It defines two groups: diffuse cutaneous systemic sclerosis (dcSSc) associated with skin changes affecting the trunk and proximal limbs, and limited cutaneous systemic sclerosis (lcSSc) where skin fibrosis is limited to the hands, face, feet and forearms. Moreover, these two groups were demonstrated to differ in terms of severity and frequency of organ involvement, autoantibody profiles and mortality.[2–4] However, an 'inverted phenotype' is well described wherein the disease process does not fall into these specified categories ushering to dissolve the binary system and give way to a continuous or expanded spectrum of categories.[5] Several intermediate subtypes such as patients with skin involvement extending proximally but excluding the trunk with intermediate risks of interstitial lung disease (ILD), unique autoantibody profile and survival, or lcSSc patients with anti-topo I antibodies and severe ILD have been described.[6–10] Large registries have called attention to this heterogeneity within the classical dcSSc and lcSSc subtypes.[11–14] Cluster analysis can provide valuable insight in this regard and is being utilized to group similar patients to predict the course and outcome of SSc, thereby guiding clinical management.[11,15]

Geographical differences in the clinical phenotype and the autoantibody profile have been well described, further highlighting the diversity of this complex disease.[16,17] There is sparse data from the Indian subcontinent in this respect.[18–20] We report an exploratory cluster analysis utilizing all relevant clinical parameters and the autoantibody profile from the baseline data of the prospective SSc registry in India.

Methods

The Indian Progressive Systemic Sclerosis Registry (IPSSR) was established in 2018 involving five specialist rheumatology referral centres across India with the intention of creating a nationwide cohort of SSc patients. This is a prospective registry in which all consenting prevalent and incident SSc patients are included. All patients undergo a detailed clinical assessment and appropriate laboratory evaluation as per clinician judgement at least every 6 months. The information is recorded into a prespecified article case record form and transcribed into an electronic database (Supplemental material 1). From this registry, all patients satisfying the 2013 ACR-EULAR classification criteria at recruitment were considered for the study and those with incomplete assessments were excluded.[1] Patients were classified into lcSSc and dcSSc as per LeRoy's classification.[2,3] SSc overlap cases were excluded. This study complies with the Declaration of Helsinki. The research protocol was approved by the locally appointed ethics committee, and written informed consent was obtained from all patients included in this registry.

This study is cross-sectional in nature, since for the current analysis, we used baseline data from the registry, which is the data recorded at the time of patient recruitment into the registry. Most clinical variables (including forced vital capacity (FVC)) are recorded as 'ever-present' at the time of recruitment. Some exceptions include the modified Rodnan skin score (mRSS), physician global assessment (PGA), patient global assessment (PtGA) and body mass index (BMI) which were recorded as examined by the clinician at the time of recruitment. For variables that change serially (FVC and mRSS), only the value recorded at the time of recruitment (i.e. baseline) was used. Prior serial changes for these values have not been used in the current analysis. Overall, 34 clinical variables were used for the cluster analysis. Variable selection was based on their relevance to the understanding of the disease spectrum; redundant variables and variables with a higher proportion of missing values were removed. Skin thickening was assessed by the mRSS method which was homogenized across the participant centres.[21] The autoantibody profile was assessed by the Euroimmun antinuclear antibodies (ANA) EUROLINE assay (EUROLineScan) and included anti ribonucleoprotein (anti-RNP/Sm), anti smith (anti-Sm), anti-Sjogren's syndrome A (anti-SSA), anti-Ro52, anti-Sjogren's syndrome B (anti-SSB), anti-topoisomerase I (anti-topo I/anti-Scl-70), anti-polymyositis/scleroderma (anti-PM/Scl-100), anti-Jo-1, anti-centromere B(ACA), anti-proliferative cell nuclear antigen (PCNA), anti-dsDNA, antinucleosome, antihistones, anti-ribosomal-P-protein and anti-mitochondrial antibody M2 (AMA-M2) autoantibodies.[22] A value of 1 + or more was considered positive. Suspected pulmonary hypertension (PH) was diagnosed based on transthoracic echocardiographic

evidence of right ventricular systolic pressure (RVSP) >40 mm Hg. None of the patients had PH confirmed by right heart catheterizations.[23] Pulmonary function tests (PFT) were performed according to standard protocol, and the FVC alone was considered for analysis.[24] ILD was diagnosed based on radiographic evidence (chest radiograph and/or high-resolution computed tomography (HRCT) chest).[25] Cardiac conduction defects were recorded as ever-present as per electrocardiograms and large joint contractures (defined as a restricted range of movement in shoulder joints, elbow joints and/or knee joints) were recorded as ever-present by the recruiting clinician at baseline. PtGA and PGA were assessed on a numerical rating scale (NRS) of 1 to 10 at the time of recruitment.

The unsupervised two-step cluster analysis method was performed to identify subgroups in patients with SSc based on all relevant immunological and clinical characteristics. In the first step, the algorithm produced preclusters in which the patients were sequentially clustered by constructing a cluster feature tree. In the second step, these pregroups were grouped into the final number of clusters using the hierarchical clustering method where the preclusters were recursively merged. The number of clusters was fixed as four taking into account the sample size, the average silhouette value which determines the quality of the cluster and the clinical relevance of the clusters, with the objective to explore possible clusters based on the available characteristics. Both continuous and categorical variables were analysed using this method. The clusters formed were then compared between the clinical characteristics and analysed using chi-square, analysis of variance (ANOVA) or Kruskal–Wallis test as appropriate. The internal consistency of the clusters was examined using Spearman's rank correlation or Pearson's correlation test for clinical parameters. A silhouette index was used for cluster validations. All analyses were performed using statistical package for the social sciences (SPSS) version 25.0 (SPSS Inc., Chicago, IL, USA). Statistical significance was considered at $p < 0.05$.

Results

Of the 564 SSc registry participants till November 2021, we included 404 patients with complete assessments in this analysis (Figure 1). The cohort consisted of predominantly prevalent SSc cases (97%). A case was defined as an incident case if the year of diagnosis and year of recruitment were the same; otherwise, it was considered a prevalent case. The cohort was female predominant (93.8%), and the mean age was 42.6 ± 11.4 years. The median duration of disease defined as the time since the onset of the first non-Raynaud's symptom to cohort entry was 6 (4,11) years, and the median age at disease onset was 35 (26, 43.7) years. A diffuse pattern of skin involvement was noted at recruitment in 214 (53.2%) patients, and the median mRSS was 9 (4,18).
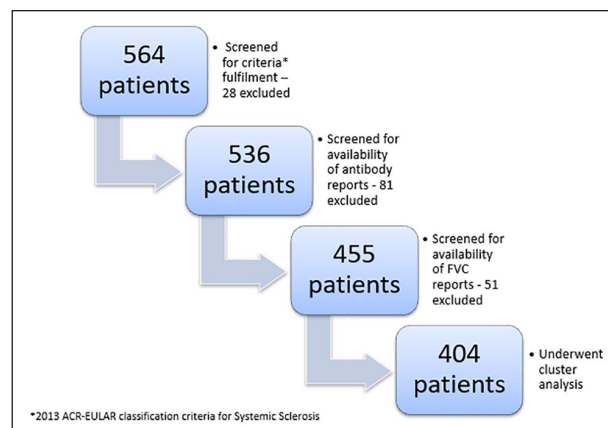

Figure 1. Schematic representation of screening process.

The predominant antibody in our cohort was anti-topo I (49.0%), followed by ACA (17.8%), anti-SSA/anti-SSB (12.6%) and anti-PM/Scl-100 (6.2%). The prevalence of scleroderma renal crisis was 0.7%. The baseline clinical characteristics of the cohort are reported in Table 1.

For cluster analysis, we used 34 clinically relevant variables (data as recorded at baseline at the time of recruitment into the registry were used). Preliminary analysis clustered patients into two groups, which partially corresponded to the dichotomous classification of limited and diffuse SSc. Cluster A was composed of lcSSc, multiple autoantibodies with a higher prevalence of ACA while Cluster B consisted of dcSSc with a higher prevalence of anti-topo I. These two clusters were driven by the severity and the extent of skin involvement, and the two autoantibodies, namely ACA and anti-topo I. There were no significant differences in the pattern of organ involvement or vasculopathy.

On exploratory analysis, we derived four clusters. The proportion of patients distributed into clusters 1, 2, 3 and 4 was 20.3%, 34.9%, 29.5% and 15.3%, respectively. The associations between the clusters and the various clinical and immunological characteristics are presented in Table 2 and Figure 2. Cluster 1 and cluster 2 presented diffuse skin fibrosis, whereas cluster 3 and cluster 4 had a limited pattern of cutaneous involvement. While clusters 1, 2 and 3 were found to have anti-topo I as the predominant antibody (84.1%, 56.0% and 37.0%, respectively), cluster 4 had a predominance of ACA (51.6%). One hundred and twelve patients (27.5%) tested negative for the autoantibodies studied. These patients were distributed throughout the four clusters, and there was no significant difference in their prevalence between the clusters.

The quality of the cluster was described to be fair by the Silhouette measure of cohesion and separation. The overall importance of predictors in cluster formation is presented in Figure 3. Of the 34 variables considered, dcSSc, lcSSc, tendon friction rubs, FVC <70%, proximal muscle weakness,

Table 1. Baseline clinical characteristics (*n* = 404).

Characteristics	*n* (%)/*mdn* (IQR)	Characteristics	*n* (%)/*mdn* (IQR)
Gender – Females	379 (93.8)	Digital pitting scars	267 (66.1)
Age at disease onset (years)	35 (26, 43.7)	Digital Gangrene	15 (3.7)
Disease duration (years)	6 (4, 11)	Tendon friction rubs	35 (8.6)
PtGA (NRS of 1–10)	4 (3, 5)	Large joint contractures	26 (6.4)
PGA (NRS of 1–10)	4 (2, 5)	Arthritis	95 (23.5)
BMI (kg/m^2)	21.9 (19.5, 25.2)	Proximal muscle weakness	101 (25.0)
lcSSc	163 (40.3)	Hypertension	79 (19.6)
dcSSc	214 (53.2)	Pericardial effusion	41 (10.1)
mRSS	9 (4, 18)	Suspected PH	44 (10.9)
ANA	354 (86.8)	ILD	257 (63.6)
Anti-topo I	198 (49.0)	SpO$_2$ >94%	319 (79.0)
ACA	72 (17.8)	FVC predicted >70%	164 (40.6)
Anti-SSA/SSB	51 (12.6)	Anal incontinence	19 (4.7)
Anti-PM/Scl-100	25 (6.2)	Diarrhoea	35 (8.7)
Negative for antibodies studied	112 (27.5)	Heartburn	209 (51.7)
History of Raynaud's phenomenon	340 (84.2)	Dysphagia	125 (30.9)
Digital ulcers	51 (12.6)	Constipation	46 (11.4)

IQR: interquartile range; PtGA: patient global assessment; PGA: physician global assessment; NRS: numerical rating scale; BMI: body mass index; SSc: systemic sclerosis; mRSS: modified Rodnan skin score; anti-SSA/SSB: anti-Sjogren's syndrome A/Sjogren's syndrome B; anti-PM/Scl-100: anti-polymyositis/scleroderma; PH: pulmonary hypertension; ILD: interstitial lung disease; SpO$_2$: saturation of peripheral oxygen; FVC: forced vital capacity; lcSSc: limited cutaneous systemic sclerosis; dcSSc: diffuse cutaneous systemic sclerosis; ANA: anti-nuclear antibodies; anti-topo I: anti-topoisomerase I; ACA: anti-centromere antibodies;

Table 2. Association of various characteristics with different clusters and their respective proportions.

Characteristics	Cluster 1	Cluster 2	Cluster 3	Cluster 4	*p*
	n-82 (20.3%)	*n*-141 (34.9%)	*n*-119 (29.5%)	*n*-62 (15.30%)	
	n (%)				
Gender – Females[b]	71 (86.6)	137 (97.2)	111 (93.3)	60 (96.8)	0.011
lcSSc[a]	0	0	110 (92.4)	53 (85.5)	<0.0001
dcSSc[a]	79 (96.3)	135 (95.7)	0	0	<0.0001
Anti-topo I[b,e]	69 (84.1)	79 (56.0)	44 (37.0)	6 (9.7)	<0.0001
ACA[b]	2 (2.4)	13 (9.2)	25 (21.0)	32 (51.6)	<0.0001
Anti-SSA/SSB[d,e]	6 (7.3)	14 (9.9)	21 (17.6)	10 (16.1)	0.091
Anti-PM/Scl-100	6 (7.3)	7 (5.0)	11 (9.2)	1 (1.6)	0.196
History of RP[d,e]	77 (93.9)	126 (89.4)	91 (76.5)	46 (74.2)	<0.0001
Digital ulcers[b]	24 (29.3)	6 (4.3)	17 (14.3)	4 (6.5)	<0.001
Digital pitting scars[d,e]	72 (87.8)	93 (66.0)	67 (56.3)	35 (56.5)	<0.0001
Digital Gangrene	7 (8.5)	0	6 (5.0)	2 (3.2)	0.01
Tendon friction rubs[b]	32 (39.0)	0	2 (1.7)	1 (1.6)	<0.0001
Large joint contractures[b]	22 (26.8)	2 (1.4)	1 (0.8)	1 (1.6)	<0.0001
Arthritis[e]	32 (39.0)	26 (18.4)	31 (26.1)	6 (9.7)	<0.0001
Proximal muscle weakness[a]	48 (58.5)	14 (9.9)	39 (32.8)	0	<0.001
Hypertension[b]	12 (14.6)	20 (14.2)	26 (21.8)	21 (33.9)	0.001
Cardiac conduction defects[b,d]	0	1	1	0	0.768
Pericardial effusion[d,e]	6 (7.3)	8 (5.7)	19 (16.0)	8 (12.9)	0.033
Suspected PH[b]	2 (2.4)	16 (11.3)	17 (14.3)	9 (14.5)	0.04
ILD[e]	68 (82.9)	94 (66.7)	87 (73.1)	8 (12.9)	<0.0001
SpO$_2$ >94%[b]	55 (67.1)	116 (82.3)	93 (78.2)	55 (88.7)	<0.0001
FVC predicted >70%[a]	25 (30.5)	60 (42.6)	20 (16.8)	59 (95.2)	<0.01
Renal crisis	1 (1.2)	0	2 (1.7)	0	0.363
Anal incontinence	7 (8.5)	3 (2.1)	2 (1.7)	7 (11.3)	0.004

(Continued)

Table 2. (Continued)

Characteristics	Cluster 1	Cluster 2	Cluster 3	Cluster 4	*p*
	n-82 (20.3%)	*n*-141 (34.9%)	*n*-119 (29.5%)	*n*-62 (15.30%)	
	n (%)				
Diarrhoea[b]	16 (20.0)	4 (2.8)	7 (5.9)	8 (12.9)	<0.0001
Heartburn[a]	60 (73.2)	60 (42.6)	48 (40.3)	41 (51.7)	<0.0001
Dysphagia[b,e]	48 (58.5)	29 (20.6)	27 (22.7)	21 (33.9)	<0.0001
Constipation[e]	20 (24.4)	9 (6.4)	14 (11.8)	3 (4.8)	<0.001
Continuous variables*					
Age at disease onset (years)[b,e]	32.6 ± 10.8	32.9 ± 11.6	37.3 ± 9.88	37.2 ± 10.8	<0.0001
Disease duration (years)[d]	7 (4, 11)	5 (3, 10)	5 (3, 9)	10 (6, 14)	0.09
mRSS[a]	16 (7, 25.0)	14 (6, 23.0)	7 (2, 11.0)	4 (2, 8.0)	<0.001
PtGA (NRS)[e]	5 (4, 7)	3 (2, 5)	4 (2, 5)	2 (1, 3.25)	<0.0001
PGA (NRS)[e]	6 (5, 7)	4 (3,5)	4 (3, 5)	3 (2, 4)	<0.0001
BMI (kg/m^2)[e]	21.2 ± 4.75	21.5 ± 4.07	22.6 ± 3.86	23.3 ± 3.74	0.018

[a]Each cluster is significantly different from other clusters; [b]cluster 1 is significantly different from other groups; [c]cluster 2 is significantly different from other groups; [d]cluster 3 is significantly different from other groups; [e]cluster 4 is significantly different from other groups. Statistical significance was considered at $p < 0.05$. IQR: interquartile range; lcSSc: limited cutaneous systemic sclerosis; dcSSc: diffuse cutaneous systemic sclerosis; anti-topo I: anti-topoisomerase I; ACA: anti-centromere antibody; anti-SSA/SSB: anti-Sjogren's syndrome A/Sjogren's syndrome B; anti-PM/Scl-100: anti-polymyositis/scleroderma; RP: Raynaud's phenomenon; PH: pulmonary hypertension; ILD: interstitial lung disease; SpO$_2$: saturation of peripheral oxygen; PtGA: patient global assessment; PGA: physician global assessment; BMI: body mass index; FVC: forced vital capacity; mRSS: modified Rodnan skin score; NRS: numerical rating scale.[1–10]
Continuous variables marked by * has been represented by median (interquartile range)/ mean (standard deviation) as applicable.
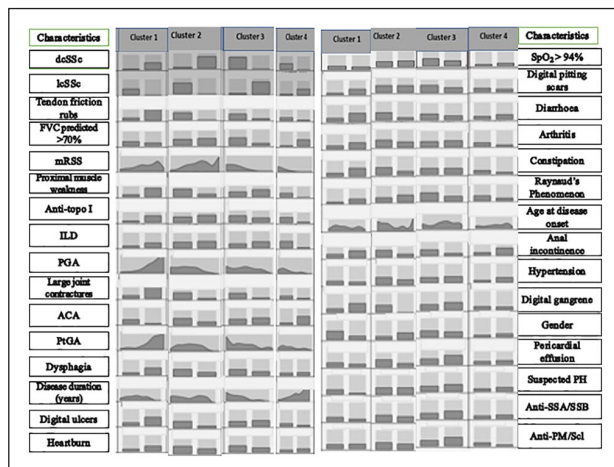

Figure 2. Distribution of characteristics within each cluster in descending order of importance in clustering.
The left bar indicates the distribution of the absence of the characteristic (darker segment) relative to that in the overall cohort (lighter segment).
dcSSc: diffuse cutaneous systemic sclerosis; lcSSc: limited cutaneous systemic sclerosis; FVC: forced vital capacity; mRSS: modified Rodnan skin score; ILD: interstitial lung disease; anti-topo I: anti-topoisomerase I; PGA: physician global assessment; ACA: anti-centromere antibody; PtGA: patient global assessment; PH: pulmonary hypertension; anti-SSA/SSB: anti-Sjogren's syndrome A/Sjogren's syndrome B; anti-PM/Scl: anti-polymyositis/scleroderma.

Figure 3. Overall importance of predictors in cluster formation.
dcSSc: diffuse cutaneous systemic sclerosis; lcSSc: limited cutaneous systemic sclerosis; FVC: forced vital capacity; mRSS: modified Rodnan skin score; ILD: interstitial lung disease; anti-topo I: anti-topoisomerase I; PGA: physician global assessment; ACA: anti-cetromere antibody; PtGA: patient global assessment.

mRSS, ILD, anti-topo I, PGA and large joint contractures appeared to be the most important contributors to the formation of the four clusters. The characteristics of the individual clusters are detailed below.
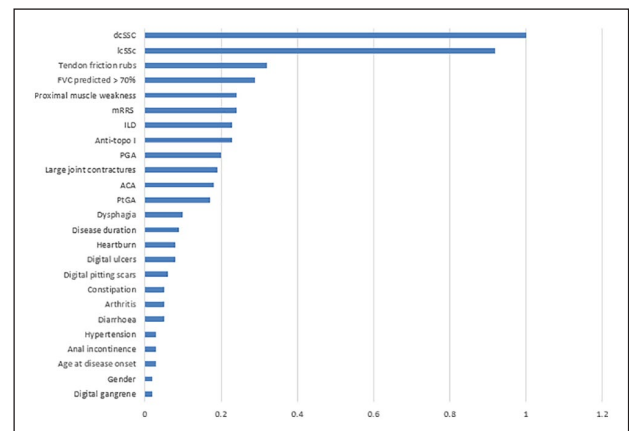
Cluster 1

This cluster consisted of patients with diffuse skin involvement, a younger age of onset and a significantly higher prevalence of male patients (13.4%) compared with the other subtypes. The predominant antibody was anti-topo I, which was significantly higher in proportion compared with the other clusters. With respect to clinical features, this cluster seemed to represent severe skin involvement

with the highest mRSS and a significantly higher prevalence of large joint contractures and tendon friction rubs. The prevalence of musculoskeletal features such as arthritis and proximal muscle weakness; vasculopathic features such as digital ulcers and pitting scars; and gastrointestinal (GI) features such as diarrhoea, heartburn, dysphagia and constipation were significantly higher. ILD was observed in 82.9%, and about two-thirds (69.5%) of the patients had FVC <70% indicating a significant proportion of lung involvement. Both PtGA and PGA were higher, reflecting a severe phenotype.

Cluster 2

This cluster also consisted of patients with a diffuse cutaneous pattern having predominantly anti-topo I and a younger age of onset. However, unlike cluster 1, cluster 2 seemed to have a less severe skin phenotype, reflected by a comparatively lower mRSS score and lower prevalence of large joint contractures and tendon friction rubs. ILD was noted in 66.7%, and a relatively lower proportion (57.4%) of patients had FVC <70%. Overall, the prevalence of musculoskeletal, GI and vasculopathic features was lesser compared with cluster 1.

Cluster 3

Cluster 3 was of particular interest, as it showed a limited pattern of skin involvement with anti-topo I as the most prevalent antibody. The mRSS was significantly higher than that in cluster 4 but lesser than that in clusters 1 and 2. In addition, cluster 3 had a unique antibody profile with a notable prevalence of other antibodies such as ACA (21%), anti-SSA/SSB (17.6%) and anti-PM/Scl-100 (9.2%). There was a significantly higher prevalence of proximal muscle weakness compared with cluster 2. ILD was seen in a notable proportion (73.1%), and the highest proportion of patients with FVC <70% (83.2%) was observed here. Digital pitting scars were found to be significantly lesser than the dcSSc predominant clusters

Clusters 2 and 3 did not differ significantly with respect to most clinical features, except for the extent of skin involvement, presence of autoantibodies other than anti-topo I, digital pitting scars and proximal muscle weakness.

Cluster 4

This cluster seemed to represent the group with the least severe skin involvement with a limited skin pattern and the lowest mRSS. The age of onset was significantly higher, and the predominant antibody was ACA (51.6%). It had a significantly low prevalence of musculoskeletal features (i.e. arthritis and proximal muscle weakness). Of particular note are the significantly low frequencies of ILD (12.9%) and patients with FVC <70% (4.8%). There was a notable

prevalence of upper GI features such as heartburn and dysphagia. Reflecting the less severe phenotype, this cluster had the least PtGA and PGA.

In addition, the clusters with limited skin involvement (i.e. clusters 3 and 4) had a significantly lower prevalence of digital pitting scars compared with the clusters with diffuse skin involvement (i.e. clusters 1 and 2).

Autoantibody and ILD association

The prevalence of ILD among the clusters with anti-topo I (clusters 1, 2 and 3) was similar. The proportion of patients with FVC <70% was highest in cluster 3 (83.2%) followed by clusters 1, 2 and 4 (69.5%, 57.4% and 4.8%, respectively) with significant differences between all the clusters. However, when the FVC values were studied as a continuous variable, the mean FVCs were similar in the three anti-topo I predominant clusters (59.7 ± 19.5, 66.5 ± 16.8 and 59.3 ± 14, respectively), while the ACA predominant cluster 4 had a significantly higher mean FVC (83.8 ± 13.2) and the lowest proportion of ILD patients (12.9%). In cluster 3, even though there were several autoantibodies in addition to anti-topo I, no independent association between these autoantibodies and ILD was established.

Discussion

We present an exploratory cluster analysis of 404 SSc patients fulfilling the 2013 ACR-EULAR classification criteria from the baseline data (recorded at the time of recruitment) of the IPSSR registry. Unsupervised two-step cluster analysis was performed to identify clusters of homogeneous patients based on 34 clinically relevant variables. Overall, the cohort consisted of prevalent cases and younger patients and was female predominant, which corroborates the findings of several studies that have shown that Asian SSc patients have a relatively younger onset of disease compared with their Caucasian counterparts.[18,26–28] Our cohort had a greater proportion of patients with a diffuse pattern of cutaneous involvement, similar to others demonstrating higher prevalence and regional differences in South Asia.[18,27–32] The predominant antibody in our cohort was anti-topo I, followed by ACA, anti-SSA/SSB and anti-PM/Scl-100. Around one-fourth of the patients were negative for all the antibodies studied, probably related to the utilization of ANA blot for antibody assays without an SSc-specific antibody panel (as the extended scleroderma blot was not available in all centres); hence, antibodies such as anti-RNA polymerase III and anti-Ku would not be captured. Like other Indian studies, our cohort was also found to have a very low prevalence of scleroderma renal crisis.[18,33]

The exploratory two-step cluster analysis provided four clusters, of which three were anti-topo I predominant and one was ACA dominant. Among these clusters, two

were dcSSc and another two were lcSSc predominant. Sobanski et al. performed exploratory cluster analysis using unsupervised machine learning on the EUSTAR cohort consisting of approximately 7000 SSc patients. Their primary analysis also revealed two clusters similar to our initial analysis that partially overlapped with the classical lcSSc and dcSSc, while exploratory analysis yielded six distinct clusters with unique features.[11]

In our study, cluster 1 seemed to have the most severe phenotype characterized by the highest proportion of dcSSc patients, the highest median mRSS, males, a younger age of onset, the highest proportion of anti-topo I positivity and a higher proportion of musculoskeletal, vasculopathic, GI and lung disease. Irrespective of the initial course, cluster 1 with a high prevalence of end-organ damage remained the cluster with a high mRSS at the time of data collection for this study. This cluster was similar to cluster 6 in the EUSTAR analysis which also exhibited a high proportion of dcSSc, anti-topo I, males, the highest mean peak mRSS and a high frequency of musculoskeletal, GI, lung, renal and cardiac disease.[11] Cluster 2 in our analysis also consisted predominantly of dcSSc, however with a lesser proportion of anti-topo I positivity. The median mRSS and the prevalence of musculoskeletal, GI, vasculopathic and lung disease were lower than those in cluster 1. This cluster resembled cluster 5 in the EUSTAR study, which consisted of mainly dcSSc and a similar prevalence of anti-topo I positivity.[11]

Cluster 3 in our cohort consisted predominantly of lcSSc with notable co-prevalence of ACA, anti-SSA/SSB and anti-PM/Scl-100 in addition to anti-topo I. Interestingly, it had the highest prevalence of severe lung involvement in terms of FVC <70%, which was noted in about four-fifths. This cluster was similar to cluster 2 of the EUSTAR cohort which was composed predominantly of lcSSc, with a higher frequency of lung involvement and co-prevalence of anti-topo I and ACA.[11] Cluster 4 seemed to represent the least severe group characterized by a predominance of lcSSc, least median mRSS, ACA positivity, older age at onset, least prevalence of musculoskeletal and lung involvement and a notable frequency of some GI features. This cluster resembled cluster 1 of the EUSTAR analysis, which had a majority of lcSSc, older age of onset, low proportion of ILD, but a high prevalence of GI features.[11]

Our clusters 2 and 3 did not seem to differ significantly with respect to organ involvement, except for the extent of skin involvement, presence of multiple autoantibodies, lower frequency of digital pitting scars and a higher frequency of proximal muscle weakness in cluster 3. The greater prevalence of proximal muscle weakness in cluster 3 may be related to the presence of anti-PM/Scl-100 associated with overlap syndromes. Digital pitting scars were found to be more common in the dcSSc clusters compared with the lcSSc clusters.

Traditionally, antibodies have been thought to play a major role in determining the extent of cutaneous involvement, such as the strong association of ACA with lcSSc and anti-topo I with dcSSc. However, several observational studies have shown that anti-topo I can also occur in lcSSc, with around 13%–15% of lcSSc patients having anti-topo I.[8,9,34] Nihtyanova et al.[8] described a novel classification by combining several autoantibody specificities and the extent of skin involvement in a cohort of 1325 SSc patients, which provided a more precise outcome-based stratification than the classical lcSSc and dcSSc. In our analysis, patients with lcSSc who had anti-topo I positivity were classified into cluster 3, hence reconfirming that anti-topo I is not exclusively associated with dcSSc.

The association of anti-topo I with ILD has been well established. In our study, we found that all three clusters with anti-topo I predominance have a higher proportion of ILD, irrespective of the extent of skin involvement. Although the proportion of patients with FVC <70% was highest in cluster 3, mean FVC values and the prevalence of ILD were similar in all three. The moderate–severe FVC restriction in clusters 1, 2 and 3 could be due to the large proportion of prevalent cases (97%) in the cohort with long-standing disease (median disease duration of 6 (4,11) years). Cluster 4 characterized by lcSSc and ACA predominance had a significantly low proportion of patients with ILD as well as FVC <70% and the highest mean FVC. Therefore, our study agrees with other studies that have shown that the frequency of ILD is similar in anti-topo I-positive dcSSc and lcSSc patients.[8–10] These studies have also shown the survival advantage of anti-topo I-positive lcSSc patients over anti-topo I-positive dcSSc patients. This may be explained by the lower incidence of other internal organ involvement such as PH and cardiac involvement in anti-topo I-positive lcSSc compared with anti-topo I-positive dcSSc patients. Although we could confirm the similar frequency of ILD in anti-topo I-positive lcSSc and dcSSc, we could not demonstrate differences in (suspected) PH and cardiac involvement probably due to the low capture of these parameters in our cohort. Zanatta et al.,[35] obtained similar results in their study on a large EUSTAR cohort and, in addition, reported that the risk of ILD progression was significantly lower in the anti-topo I-positive lcSSc patients than in anti-topo I-positive dcSSc but similar across the other lcSSc subtypes[35]. This aspect can be studied in the future follow-up of our cohort as well.

Our study corroborates the results from previous studies demonstrating the heterogeneity of SSc and yet the possibility of subgrouping beyond the classical dcSSc and lcSSc. We used real-life baseline data from a multicentric registry, which is likely to be fairly representative of the Indian cohort of SSc patients. We reconfirm that the scleroderma renal crisis is rather infrequent in the SSc population in the Indian subcontinent. However, our analysis had limitations as well. Our data were captured in a cross-sectional

manner, for both the prevalent and incident cases. Therefore, the early course of the disease and those with early mortality may be underrepresented, especially as prevalent cases constituted the majority in the cohort. In addition, for prevalent cases, serial mRSS and FVC values prior to recruitment were not captured; hence, the temporal course of these variables was unavailable. Our cohort had a fairly large number of patients who tested negative for all the antibodies studied, possibly due to the use of the Euroimmun ANA EUROLINE assay for autoantibody profiling in accordance with prevalent practice, which does not include several SSc-specific antibodies including the anti-RNA polymerase III.

Despite these limitations, we strongly feel that this study provides insight into the heterogeneous nature of SSc and can be further explored. We are working on a longitudinal follow-up of these patients to observe the disease course and outcomes of the clusters in terms of organ involvement (including progression of ILD and skin involvement) and mortality so as to discern their utility in clinical practice with respect to therapeutics, prognosis and survival. Our findings have to be validated in other regional cohorts to establish the stability and reproducibility of these clusters. Once validated, these clusters may be helpful in the early prognostication and risk stratification of SSc patients at the time of initial evaluation.

Conclusion

The IPSSR serves as a valuable nationwide source of data on Indian SSc patients. In this cohort, the exploratory cluster analysis revealed four scleroderma clusters, reflecting the heterogeneous nature of the disease. We reconfirm that lcSSc can also be associated with anti-topo I, organ involvement and more severe disease. We also report that clusters with a higher prevalence of anti-topo I have a higher proportion of ILD, irrespective of the extent of skin fibrosis. Prospectively, we intend to follow-up with these patients to study the disease course and the outcomes in each cluster and determine their relevance in terms of prognosis and survival.

Author contributions

1. Conception and design of the study – Dr.Vineeta Shobha, Dr. Ramya Janardana.
2. Analysis and interpretation of data – Dr. Vineeta Shobha, Dr. Ramya Janardana, Dr. Shery Susan Philip, Sumithra Selvam.
3. Data acquisition – Dr. Vineeta Shobha, Dr.Ramya Janardana, Dr.Shery Susan Philip, Dr. Padmanabha Shenoy, Dr. Chengappa Kavadichanda, Dr. Devender Bairwa, Dr. Geetabali Sircar, Dr. Parasar Ghosh, Dr. Anupam Wakhlu.
4. Drafted the work and revised it critically for important intellectual content – Dr.Shery Susan Dr. Ramya Janardana, Dr.Vineeta Shobha, Dr. Dinesh Khanna.
5. Approved the version to be published – All authors.
6. Agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved – All authors.

ORCID iDs

Shery Susan Philip (iD) https://orcid.org/0000-0001-7301-4970

Padmanabha Shenoy (iD) https://orcid.org/0000-0002-7666-1361

Dinesh Khanna (iD) https://orcid.org/0000-0001-6822-3401

Vineeta Shobha (iD) https://orcid.org/0000-0001-5271-0766

Supplemental material

Supplemental material for this article is available online.

References

1. van den Hoogen F, Khanna D, Fransen J, et al. 2013 classification criteria for systemic sclerosis: an American College of Rheumatology/European League against Rheumatism collaborative initiative. *Arthritis Rheum* 2013; 65(11): 2737–2747.
2. LeRoy EC, Black C, Fleischmajer R, et al. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. *J Rheumatol* 1988; 15(2): 202–205.
3. LeRoy EC and Medsger TA. Criteria for the classification of early systemic sclerosis. *J Rheumatol* 2001; 28(7): 1573–1576.
4. Giordano M, Valentini G, Migliaresi S, et al. Different antibody patterns and different prognoses in patients with scleroderma with various extent of skin sclerosis. *J Rheumatol* 1986; 13(5): 911–916.
5. Tieu A, Chaigne B, Dunogué B, et al. Autoantibodies versus skin fibrosis extent in systemic sclerosis: a case-control study of inverted phenotypes. *Diagnostics* 2022; 12(5): 1067.

6. Cottrell TR, Wise RA, Wigley FM, et al. The degree of skin involvement identifies distinct lung disease outcomes and survival in systemic sclerosis. *Ann Rheum Dis* 2014; 73(6): 1060–1066.

7. Barnett AJ, Miller MH and Littlejohn GO. A survival study of patients with scleroderma diagnosed over 30 years (1953-1983): the value of a simple cutaneous classification in the early stages of the disease. *J Rheumatol* 1988; 15(2): 276–283.

8. Nihtyanova SI, Sari A, Harvey JC, et al. Using autoantibodies and cutaneous subset to develop outcome-based disease classification in systemic sclerosis. *Arthritis Rheumatol* 2020; 72(3): 465–476.

9. Kranenburg P, van den Hombergh WMT, Knaapen-Hans HKA, et al. Survival and organ involvement in patients with limited cutaneous systemic sclerosis and anti-topoisomerase-I antibodies: determined by skin subtype or auto-antibody subtype? A long-term follow-up study. *Rheumatology* 2016; 55(11): 2001–2008.

10. Srivastava N, Hudson M, Tatibouet S, et al. Thinking outside the box – the associations with cutaneous involvement and autoantibody status in systemic sclerosis are not always what we expect. *Semin Arthritis Rheum* 2015; 45(2): 184–189.

11. Sobanski V, Giovannelli J, Allanore Y, et al. Phenotypes determined by cluster analysis and their survival in the prospective European scleroderma trials and research cohort of patients with systemic sclerosis. *Arthritis Rheumatol* 2019; 71(9): 1553–1570.

12. Frantz C, Huscher D, Avouac J, et al. Outcomes of limited cutaneous systemic sclerosis patients: results on more than 12,000 patients from the EUSTAR database. *Autoimmun Rev* 2020; 19(2): 102452.

13. Leclair V, Hudson M, Proudman SM, et al. Subsets in systemic sclerosis: one size does not fit all. *J Scleroderma Relat Disord* 2016; 1(3): 298–306.

14. Ingegnoli F, Ardoino I, Boracchi P, et al. Nailfold capillaroscopy in systemic sclerosis: data from the EULAR scleroderma trials and research (EUSTAR) database. *Microvasc Res* 2013; 89: 122–128.

15. Stafford IS, Kellermann M, Mossotto E, et al. A systematic review of the applications of artificial intelligence and machine learning in autoimmune diseases. *NPJ Digit Med* 2020; 3(1): 1–11.

16. Patterson KA, Roberts-Thomson PJ, Lester S, et al. Interpretation of an extended autoantibody profile in a well-characterized Australian systemic sclerosis (scleroderma) cohort using principal components analysis. *Arthritis Rheumatol* 2015; 67(12): 3234–3244.

17. Allanore Y, Gharibdoost F, Jamshidi AR, et al. Comparison of the clinical phenotype of systemic sclerosis patients in Iran and France in two university centers. *J Scleroderma Relat Disord* 2019; 4(2): 149–159.

18. Janardana R, Nair AM, Surin AK, et al. Unique clinical and autoantibody profile of a large Asian Indian cohort of scleroderma – do South Asians have a more aggressive disease? *Clin Rheumatol* 2019; 38(11): 3179–3187.

19. Krishnamurthy V, Porkodi R, Ramakrishnan S, et al. Progressive systemic sclerosis in South India – PubMed. *J Assoc Phys Ind* 1991; 39(3): 254–257.

20. Pradhan V, Rajadhyaksha A, Nadkar M, et al. Clinical and autoimmune profile of scleroderma patients from Western India. *Int J Rheumatol* 2014; 2014(983781): 1–6.

21. Khanna D, Furst DE, Clements PJ, et al. Standardization of the modified Rodnan skin score for use in clinical trials of systemic sclerosis. *J Scleroderma Relat Disord* 2017; 2(1): 11–18.

22. Immunoblot: EUROIMMUN immunoblots, https://www.euroimmun.com/products/techniques/immunoblot/

23. Hsu VM, Moreyra AE, Wilson AC, et al. Assessment of pulmonary arterial hypertension in patients with systemic sclerosis: comparison of noninvasive tests with results of right-heart catheterization. *J Rheumatol* 2008; 35(3): 458–465.

24. Graham BL, Steenbruggen I, Barjaktarevic IZ, et al. Standardization of spirometry 2019 update. an official American Thoracic Society and European Respiratory Society technical statement. *Am J Respir Crit Care Med* 2019; 200(8): E70–E88.

25. Raghu G, Remy-Jardin M, Myers JL, et al. Diagnosis of idiopathic pulmonary fibrosis. An official ATS/ERS/JRS/ALAT clinical practice guideline. *Am J Respir Crit Care Med* 2018; 198(5): e44–e68.

26. Ng SA and Low AHL. Systemic sclerosis in Asians: are there racial differences? *J Scleroderma Relat Disord* 2022; 7(2): 98–109.

27. Chung MP, Dontsi M, Postlethwaite D, et al. Increased mortality in Asians with systemic sclerosis in Northern California. *ACR Open Rheumatol* 2020; 2(4): 197–206.

28. Jaeger VK, Tikly M, Xu D, et al. Racial differences in systemic sclerosis disease presentation: a European Scleroderma Trials and Research group study. *Rheumatology* 2020; 59(7): 1684–1694.

29. Proudman SM, Huq M, Stevens W, et al. What have multicentre registries across the world taught us about the disease features of systemic sclerosis? *J Scleroderma Relat Disord* 2017; 2(3): 169–182.

30. Li X, Qian YQ, Liu N, et al. Survival rate, causes of death, and risk factors in systemic sclerosis: a large cohort study. *Clin Rheumatol* 2018; 37(11): 3051–3056.

31. Foocharoen C, Peansukwech U, Mahakkanukrauh A, et al. Clinical characteristics and outcomes of 566 Thais with systemic sclerosis: a cohort study. *Int J Rheum Dis* 2020; 23(7): 945–957.

32. Santosa A, Tan CS, Teng GG, et al. Lung and gastrointestinal complications are leading causes of death in SCORE, a multi-ethnic Singapore systemic sclerosis cohort. *Scand J Rheumatol* 2016; 45(6): 499–506.

33. Gupta R, Bammigatti C, Dinda AK, et al. Prevalence of renal involvement in Indian patients with systemic sclerosis. *Indian J Med Sci* 2007; 61(2): 91–96.

34. Steen VD. Autoantibodies in systemic sclerosis. *Semin Arthritis Rheum* 2005; 35(1): 35–42.

35. Zanatta E, Huscher D, Ortolan A, et al. Phenotype of limited cutaneous systemic sclerosis patients with positive anti-topoisomerase I antibodies: data from the EUSTAR cohort. *Rheumatology* 2022; 61(12): 4786–4796.