British Society for Rheumatology

RHEUMATOLOGY
ADVANCES IN PRACTICE

OXFORD

Clinical science

Clinical phenotyping in patients with anti-synthetase antibodies using cluster analysis

Shintaro Yamamoto [1], Akira Yoshida [1], Yuka Okazaki[1], Takahisa Gono [1,2], Masataka Kuwana [1,2],*

[1]Department of Allergy and Rheumatology, Nippon Medical School Graduate School of Medicine, Tokyo, Japan
[2]Scleroderma/Myositis Centre of Excellence, Nippon Medical School Hospital, Tokyo, Japan
*Correspondence to: Masataka Kuwana, Department of Allergy and Rheumatology, Nippon Medical School Graduate School of Medicine, 1-1-5, Sendagi, Bunkyo-ku, Tokyo 113-8602, Japan. E-mail: kuwanam@nms.ac.jp

Abstract

Objectives: To characterize clinically distinct subgroups among unselected patients with anti-synthetase antibodies using cluster analysis.

Methods: This study evaluated patients with anti-synthetase antibodies registered to two independent cohorts; 106 consecutive patients from a prospective, single-centre cohort of the Scleroderma/Myositis Centre of Excellence (SMCE) were used as a derivation cohort and 125 patients from the Multicentre Retrospective Cohort of Japanese Patients with Myositis-Associated Interstitial Lung Disease (JAMI) were used as a validation cohort. Anti-synthetase antibodies were identified by RNA immunoprecipitation. A multiple correspondence analysis followed by hierarchical clustering was performed to aggregate the patients into homogeneous subgroups. Subsequently, a simple-to-use classification tree was generated using classification and regression tree analysis.

Results: Three clusters were identified in the SMCE cohort: cluster 1 ($n = 48$), the interstitial pneumonia with autoimmune features/amyopathic dermatomyositis cluster, associated with older age at diagnosis and a higher frequency of malignancy; cluster 2 ($n = 46$), the DM cluster, corresponded to a younger age at diagnosis with a higher prevalence of myositis, arthritis, DM pathognomonic rashes, mechanic's hands and fever; and cluster 3 ($n = 12$), the SSc cluster, characterized by chronic interstitial lung disease. There was no significant difference in overall survival or progression-free survival between the clusters. A simple classification tree using myositis and RP was created in the SMCE cohort. Clusters 1 and 2 were successfully reproduced and the classification tree demonstrated favourable performance in the JAMI cohort.

Conclusion: Patients with anti-synthetase antibodies were classified into three distinct phenotypes, indicating substantial heterogeneity within this patient group.

Lay Summary

What does this mean for patients?

Anti-synthetase antibodies are part of the immune system. They mistakenly target an enzyme important in the process of making proteins in the body. Individuals with anti-synthetase antibodies can develop a wide range of symptoms, including muscle weakness, joint pain and/or swelling, interstitial lung disease, fever, Raynaud's phenomenon (RP) and skin rashes. This is referred to as anti-synthetase syndrome. To better understand this disease, we performed a cluster analysis, a statistical method used to group objects based on their shared characteristics. Using data from 106 people with anti-synthetase antibodies, three distinct subgroups (i.e. clusters) were identified, corresponding to individuals characterized by interstitial lung disease with or without skin rashes; muscle weakness, arthritis, fever and skin rashes; and RP and skin thickening. The presence of muscle involvement and RP was found to be useful for clustering. Our findings not only help doctors better understand the disease, but also provide insights that could help develop shared classification criteria, which is critical for conducting clinical trials and developing new treatments.

Graphical abstract

Graphical abstract

Key messages
- Cluster analysis identified three clinically distinct phenotypes in unselected patients with anti-synthetase antibodies.
- A classification tree comprising myositis and RP efficiently divided anti-synthetase-positive patients into three clusters.

Introduction

Autoantibodies to aminoacyl transfer RNA (tRNA) synthetases or anti-synthetase antibodies are detected in 20–30% of patients with idiopathic inflammatory myopathies (IIMs) [1]. To date, at least six anti-synthetase antibodies are well characterized, including anti-Jo-1, anti-PL-7, anti-EJ, anti-PL-12, anti-OJ and anti-KS antibodies [1]. Patients with anti-synthetase antibodies present with a different combination of clinical features, including myositis, interstitial lung disease (ILD), arthritis, RP, fever and mechanic's hands [2, 3]. Anti-synthetase syndrome (ASyS) has been proposed as a distinct clinical entity representing this unique disease spectrum, and international classification criteria are currently under development [4]. Some patients with anti-synthetase antibodies suffer from ILD but lack clinical features characteristic of IIMs and/or other connective tissue diseases; these patients are classified as having interstitial pneumonia with autoimmune features (IPAF) [5]. Importantly, survival was not found to differ between IPAF patients with myositis-specific autoantibodies (MSAs), including anti-synthetase antibodies, and patients with established IIMs and ILD [6]. In addition, a subset of patients with anti-synthetase antibodies also exhibit

an SSc phenotype in the absence of SSc-specific autoantibodies [7, 8]. Therefore, patients with anti-synthetase antibodies constitute a heterogeneous disease population in the context of IIMs and non-IIMs.

Cluster analysis is a multivariable method that has the ability to classify a heterogeneous population into distinct subsets with more homogeneous characteristics [9]. Disease clustering facilitates appropriate management decisions through more precise disease phenotyping. In recent years, cluster analysis has been applied to IIMs [10] and various subclassifications of IIMs, including anti-MDA5 antibody-positive DM [11] and ASyS with anti-Jo-1, anti-PL-7 or anti-PL-12 antibodies [12]. In a situation where prior research is scarce, patients with anti-synthetase antibodies are especially good candidates for unsupervised clustering considering their significant heterogeneity, encompassing the IIM and non-IIM disease spectrum. In this study we performed cluster analysis using a single-centre cohort of unselected patients with anti-synthetase antibodies regardless of clinical diagnosis as a derivation cohort. Furthermore, the validity of the cluster classification was evaluated using a multicentre cohort of patients with myositis-associated ILD and anti-synthetase

antibodies, aiming to characterize clinically distinct sub-groups in this substantially heterogeneous disease entity.

Methods

Cohorts

This study used two independent cohorts, including a single-centre prospective cohort of patients undergoing full autoantibody panel screening at the Scleroderma/Myositis Centre of Excellence (SMCE), Nippon Medical School Hospital, as a derivation cohort and the Multicentre Retrospective Cohort of Japanese Patients with Myositis-Associated ILD (JAMI) [13] as a validation cohort. For the SMCE autoantibody cohort, all patients who underwent full autoantibody panel screening since August 2014, regardless of their clinical diagnosis, were consecutively enrolled. For the present study, we selected patients with anti-synthetase antibodies from the SMCE autoantibody cohort, while patients who had received treatment at other centres were excluded because we were unable to obtain the detailed clinical features at diagnosis. The patients enrolled were recruited from the rheumatology clinic or respiratory clinic. For the JAMI cohort, patients with incident cases of myositis-associated ILD were recruited from 21 rheumatology, 17 pulmonology and 5 dermatology departments across Japan between October 2011 and October 2015. The detailed inclusion criteria were described in a previous publication [13]. Briefly, patients who were >16 years of age at disease onset; had definite or probable PM/DM according to the classification criteria of Bohan and Peter [14] or clinically amyopathic DM (CADM) according to the criteria proposed by Sontheimer [15]; had the presence of ILD; and had serum samples at diagnosis for comprehensive MSA analysis were eligible for inclusion. All patients confirmed to be positive for anti-synthetase antibodies in the JAMI cohort were selected for the present study. There was no overlap between patients enrolled in the SMCE autoantibody cohort and patients enrolled in the JAMI cohort.

Written informed consent was obtained from every patient at cohort enrolment. The present study was approved by the Ethics Committee of Nippon Medical School Hospital (B-2020-127) and conducted according to the Declaration of Helsinki. The JAMI cohort was registered in the University Hospitals Medical Information Network Clinical Trial Registry (UMIN000018663).

Clinical data

Clinical data derived from the database as well as their definitions are detailed in Supplementary Data S1, available *Rheumatology Advances in Practice* online. The classification of PM/DM and SSc was based on the 2017 EULAR/ACR classification criteria for IIMs [16] and the 2013 ACR/EULAR classification criteria for SSc [17], respectively. PM/DM–SSc overlap was defined as fulfilling both classification criteria. The proposed criteria for IPAF [5] were applied to patients unclassified by the above criteria.

Autoantibody detection

Serum samples were obtained from all patients referred to our institution at their initial visit and were subjected to comprehensive autoantibody testing. Anti-synthetase antibodies were identified by RNA immunoprecipitation (RNA-IP) assay, as described previously [18]. The antigenic specificity of individual anti-synthetase antibodies was determined based on comparison with the precipitated RNA components by the prototype sera positive for anti-Jo-1, anti-PL-7, anti-PL-12, anti-EJ, anti-OJ or anti-KS antibodies. Anti-Ro52 antibodies were measured by a commercial ELISA kit according to the manufacturer's instructions (ORG652, ORGENTEC, Mainz, Germany).

Outcome measures

Overall survival (OS) and progression-free survival (PFS) were the outcome measures. PFS was defined as the time from diagnosis to the first incidence of disease progression, all-cause mortality or the most recent visit. The definition of disease progression is detailed in Supplementary Data S1, available *Rheumatology Advances in Practice* online. The time of the last observation was 30 November 2022 for the SMCE cohort and 1 January 2016 for the JAMI cohort. The information on PFS was unavailable for the JAMI cohort.

Statistical analysis

Continuous variables are presented as median and interquartile range (IQR) and comparisons were made using the Kruskal–Wallis test or Mann–Whitney U test, as appropriate. Fisher's exact test was employed for comparisons of categorical variables. We did not perform statistical analysis for variables with >10% missing data.

First, an unsupervised machine learning approach, namely multiple correspondence analysis (MCA) followed by hierarchical clustering, was applied to aggregate patients in the SMCE cohort into homogeneous subgroups [10, 11] using 13 clinical variables: demographics (age and sex), clinical features associated with ASyS [myositis, ILD as well as high-resolution CT (HRCT) pattern, arthritis, mechanic's hand, RP and unexplained fever], IIMs (DM pathognomonic rash and malignancy) or SSc (sclerodactyly). Anti-synthetase antibody specificities and anti-Ro52 antibodies were also included as variables [19, 20]. The detailed definition of each variable is provided in Supplementary Table S1, available *Rheumatology Advances in Practice* online. MCA was performed using the corresp function and the result was depicted in a two-dimensional factor map. The first and second principal components (PC1 and PC2) of the factor map were considered new variables and were subjected to hierarchical clustering. The clustering was performed by the hclust function using the squared Euclidean distance and the Ward agglomerative method. The clustering process was plotted as a dendrogram. To determine the optimal number of clusters, we used the following two distinct approaches: a visual distance criterion by cutting the dendrogram horizontally at the level of highest dissimilarity, i.e. where vertical branches were the longest [9]; and the NbClust function, which proposes the best number of clusters derived from multiple indices [21]. Cluster stability was evaluated by the Jaccard coefficient, a measure of similarity between datasets obtained using the clusterboot function (low, ≤0.6; medium, >0.6–≤0.75; high, >0.75).

Next, we applied supervised machine learning approaches to select the clinical or serological features specific to the clusters identified in the SMCE cohort and quantified the weight of each variable for clustering. We implemented the random forest algorithm using the randomForest function, where the weight of each variable was calculated by the mean decrease in the Gini impurity index. To construct a simple-to-use classification tree that can easily position each patient into the

clusters, classification and regression tree (CART) analysis was performed using the rpart function. Finally, the reproducibility of clusters and the utility of the classification tree generated in the SMCE cohort were investigated in the external validation JAMI cohort.

Survival analyses were performed using Kaplan–Meier curves and the difference between groups was tested by the logrank test. Statistical significance was defined as a two-sided *P*-value <0.05. We adjusted *P*-values by multiplying the number of comparisons made where multiple testing was applied. All statistical analyses were performed with R version 4.3.1 (R Foundation for Statistical Computing, Vienna, Austria) and EZR version 1.61 (Saitama Medical Centre, Jichi Medical University, Shimotsuke, Japan) [22].

Results

Patients in the SMCE cohort

Supplementary Fig. S1, available *Rheumatology Advances in Practice* online, displays the study flowchart. From August 2014 to March 2022, RNA-IP was performed for 736 patients who had visited the SMCE owing to suspected myositis or SSc. Of these, 136 patients were identified as having anti-synthetase antibodies. Patients who had already received immunomodulatory treatment in other hospitals ($n = 29$) and those with missing variables necessary for MCA ($n = 1$) were excluded; ultimately, 106 patients were subjected to cluster analysis. Of these, 67 (63.2%) were referred from pulmonologists. The median age at diagnosis was 64 years (IQR 53–71) and 73 (68.9%) were female. The antigenic specificity of anti-synthetase antibodies was as follows: Jo-1 in 24 (22.6%), PL-7 in 18 (17.0%), PL-12 in 12 (11.3%), EJ in 25 (23.6%), OJ in 6 (5.7%) and KS in 19 (17.9%). Two patients were positive for two anti-synthetase antibodies simultaneously: one for anti-Jo-1 and anti-EJ and the other for anti-EJ and anti-KS antibodies. Forty-three (40.6%) patients were positive for anti-Ro52 antibodies. All 106 patients were classified into any one of DM [$n = 15$ (14.2%)], ADM [$n = 25$ (23.6%)], PM/IMNM [$n = 14$ (13.2%)], SSc [$n = 4$ (3.8%)], IIM-SSc overlap [$n = 9$ (8.5%)] or IPAF [$n = 39$ (36.8%)]. Most patients classified as having IPAF [$n = 33$ (84.6%)] were referred from pulmonologists, in whom the classification was made with multidisciplinary discussion involving pulmonologists and rheumatologists.

Cluster description

The result of MCA performed in the SMCE cohort was depicted on a factor map (Fig. 1A). PC1 and PC2 on the factor map were considered two new variables and subjected to hierarchical clustering and a dendrogram was generated (Fig. 1B). The NbClust function proposed the optimal number of clusters as 3 or 13 (Supplementary Fig. S2, available *Rheumatology Advances in Practice* online). Combined with the visual distance criterion, we chose three clusters as an appropriate subclassification. Clusters 1–3 demonstrated average Jaccard coefficients of 0.89, 0.67 and 0.68, respectively, indicating moderate to high stability.

The clinical characteristics of the patients included in each cluster are presented in Table 1. Patients in cluster 2 were younger at diagnosis than those in the other clusters ($P = 0.033$). There was no difference in the distribution of anti-synthetase antibody specificities, but there was a trend towards a higher prevalence of anti-Jo-1 and anti-PL-7 and a

lower prevalence of anti-KS in cluster 2. The prevalence of anti-Ro52 antibody positivity was comparable between the clusters.

Cluster 1 [$n = 48$ (45.3%)] mostly consisted of patients classified as having IPAF [$n = 32$ (66.7%)] or ADM [$n = 10$ (20.8%)]. No patients presented with myositis or RP, and only four patients (8.3%) had either arthritis or fever. Of note, malignancy was observed most frequently in this cluster ($P = 0.042$). Cluster 2 [$n = 46$ (43.4%)] was dominated by those classified as having DM [$n = 15$ (32.6%)] or ADM [$n = 15$ (32.6%)], characterized by a higher prevalence of muscle weakness [$n = 25$ (54.3%)], arthritis [$n = 21$ (45.7%)], mechanic's hands [$n = 34$ (73.9%)], fever [$n = 17$ (37.0%)], nailfold capillary abnormalities [$n = 24$ (52.2%)] and skin ulcers [$n = 5$ (10.9%)]. Cluster 3 [$n = 12$ (11.3%)] consisted mostly of those classified as having SSc [$n = 3$ (25.0%)] or IIM–SSc overlap [$n = 3$ (25.0%)]. None of the patients in this cluster had myositis, and the prevalence of ILD (83.3%) was lower than that in the other clusters ($P = 0.018$). In addition, SSc-related manifestations were the most remarkable, including RP [$n = 9$ (75.0%)], puffy fingers [$n = 9$ (75.0%)], sclerodactyly [$n = 8$ (66.7%)] and proximal scleroderma [$n = 4$ (33.3%)]. Of note, two patients in cluster 3 (16.7%) were positive for SSc-specific autoantibodies, including anticentromere or anti-RNA polymerase III antibodies.

There was no significant difference in the mode of ILD onset or the frequency of rapidly progressive ILD (RP-ILD) between the clusters, while the prevalence of chronic ILD was numerically higher in cluster 3. Regarding the ILD morphologic pattern on HRCT, there was a trend towards more usual interstitial pneumonia (UIP) patterns rather than non-specific interstitial pneumonia (NSIP) and/or organizing pneumonia (OP) patterns in cluster 3 ($P = 0.013$). In terms of serum biomarkers, the median level of Krebs von den Lungen-6 was higher in cluster 1 ($P = 0.008$). The distribution of initial treatment regimens was comparable between clusters 1 and 2, while patients in cluster 3 were treated with glucocorticoids or immunosuppressive agents less frequently than those in clusters 1 or 2 (41.7% *vs* 79.2% or 97.8%, adjusted $P < 0.001$ for both comparisons).

Next, we applied the random forest algorithm to quantify the weight of each clinical or serological variable for clustering. Following model optimization, the random forest model demonstrated a favourable performance with an out-of-bag error rate of 1.89%. The weight of each variable for clustering, as calculated by the mean decrease in the Gini impurity index, was the highest for myositis, followed by RP (Fig. 2A). Mirroring this result, the CART analysis generated a classification tree based on myositis (muscle weakness and/or muscle enzyme elevation) and RP (Fig. 2B). Using this simple classification tree, 90 patients (84.9%) were successfully positioned into correct clusters (Supplementary Fig. S3A, available *Rheumatology Advances in Practice* online).

Validation in the JAMI cohort

Among 497 patients in the JAMI cohort, 163 were positive for anti-synthetase antibodies by RNA-IP. Excluding 38 patients due to missing variables for applying the MCA, 125 patients were eligible for the validation (Supplementary Fig. S1, available *Rheumatology Advances in Practice* online). The result of MCA in the JAMI cohort was depicted on a factor map (Supplementary Fig. S4A, available *Rheumatology*
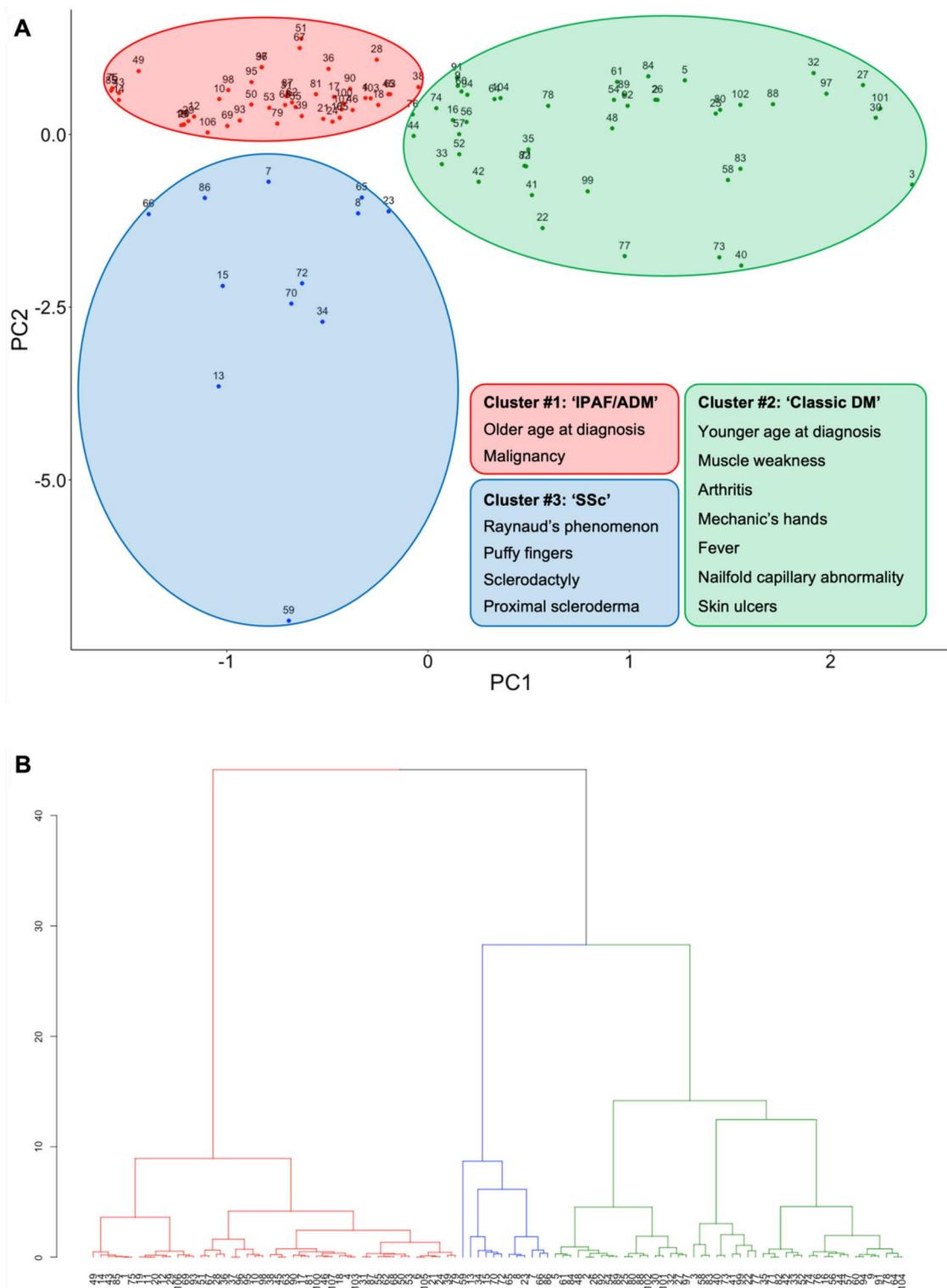
Figure 1. Unsupervised clustering using the SMCE cohort. **(A)** Factor map depicting the result of MCA. **(B)** Dendrogram generated by hierarchical clustering. The colours indicate the clusters to which each case belongs

Advances in Practice online) and a dendrogram was generated through hierarchical clustering (Supplementary Fig. S4B, available *Rheumatology Advances in Practice* online). The visual distance criterion and the NbClust function (Supplementary Fig. S5, available *Rheumatology Advances in Practice* online) proposed the optimal number of clusters as two and three, respectively. We stratified the patients into two clusters, considering the third cluster would include only five patients.

Notably, clinical characteristics of patients included in the two clusters of the JAMI cohort mirrored the features of clusters 1 and 2 in the SMCE cohort (Table 2). Specifically, clusters 1 ($n = 60$) and 2 ($n = 65$) in the JAMI cohort were dominated by ADM [$n = 35$ (58.3%)] and DM [$n = 34$

Table 1. Clinical characteristics of 106 patients with anti-synthetase antibodies in the SMCE cohort, stratified by cluster

Characteristics	Cluster 1 ($n = 48$)	Cluster 2 ($n = 46$)	Cluster 3 ($n = 12$)	P-value
Demographics				
Age at diagnosis, years, median (IQR)	67 (55–72)	60 (50–68)	65 (62–69)	**0.033**
Female	30 (62.5)	36 (78.3)	7 (58.3)	0.180
Clinical diagnosis				**<0.001**
DM	0	15 (32.6)	0	
ADM	10 (20.8)	15 (32.6)	0	
PM/IMNM	5 (10.4)	8 (17.4)	1 (8.3)	
SSc	1 (2.1)	0	3 (25.0)	
IIM-SSc overlap	0	6 (13.0)	3 (25.0)	
IPAF	32 (66.7)	2 (4.3)	5 (41.7)	
Autoantibodies				
Anti-synthetase antibodies				**0.021**
Anti-Jo-1	7 (14.6)	15 (32.6)	2 (16.7)	
Anti-PL-7	4 (8.3)	10 (21.7)	4 (33.3)	
Anti-PL-12	6 (12.5)	6 (13.0)	0	
Anti-EJ	14 (29.2)	10 (21.7)	1 (8.3)	
Anti-OJ	4 (8.3)	2 (4.3)	0	
Anti-KS	12 (25.0)	3 (6.5)	4 (33.3)	
Double-positive	1[a] (2.1)	0	1[b] (8.3)	
Anti-Ro52 antibody	22 (45.8)	17 (37.0)	4 (33.3)	0.638
Clinical features at initial presentation				
Muscle weakness	0	25 (54.3)	0	**<0.001**
ILD	48 (100)	42 (91.3)	10 (83.3)	**0.017**
DM pathognomonic rash	10 (20.8)	33 (71.7)	2 (16.7)	**<0.001**
Clinical features				
Arthritis	4 (8.3)	21 (45.7)	1 (8.3)	**<0.001**
Muscle involvement				**<0.001**
No myositis	48 (100)	13 (28.3)	12 (100)	
Subclinical myositis	0	8 (17.4)	0	
Clinical myositis	0	25 (54.3)	0	
ILD	48 (100)	44 (95.7)	10 (83.3)	**0.018**
Onset of ILD ($n = 102$)				0.391
Acute	17 (35.4)	15 (34.1)	2 (20.0)	
Subacute	9 (18.8)	7 (15.9)	1 (10.0)	
Chronic	17 (35.4)	16 (36.4)	6 (60.0)	
Asymptomatic/unclassified	5 (10.4)	6 (13.6)	1 (10.0)	
RP-ILD	12 (25.0)	8 (17.4)	2 (20.0)	0.699
$SpO_2:FiO_2$ ratio ($n = 99$), median (IQR)	457.1 (447.6–461.9)	457.1 (452.4–466.7)	461.9 (438.1–471.4)	0.289
$PaO_2:FiO_2$ ratio ($n = 80$), median (IQR)	370.0 (350.0–417.1)	415.0 (359.6–457.0)	405.7 (371.0–440.0)	NA
Alveolar-arterial oxygen gradient, mmHg ($n = 80$), median (IQR)	23.6 (8.8–32.8)	14.7 (7.1–26.8)	12.9 (5.1–29.6)	NA
%FVC ($n = 82$), median (IQR)	82.6 (67.1–94.2)	75.7 (62.7–87.8)	81.2 (75.8–97.6)	NA
%DLCO ($n = 74$), median (IQR)	66.9 (58.3–78.6)	63.8 (51.2–77.1)	84.6 (62.2–102.4)	NA
HRCT pattern of ILD				**0.013**
UIP	3 (6.2)	2 (4.5)	2 (20.0)	
NSIP and/or OP	40 (83.3)	40 (90.9)	6 (60.0)	
DAD	4 (8.3)	1 (2.3)	0	
Unclassified/unknown	1 (2.1)	1 (2.3)	2 (20.0)	
DM pathognomonic rash	14 (29.2)	35 (76.1)	2 (16.7)	**<0.001**
Gottron's papules	3 (6.2)	4 (8.7)	0	0.750
Gottron's sign	14 (29.2)	31 (67.4)	2 (16.7)	**<0.001**
Heliotrope rash	3 (6.2)	12 (26.1)	0	**0.008**
V sign	4 (8.3)	3 (6.5)	1 (8.3)	1.000
Shawl sign	1 (2.1)	6 (13.0)	1 (8.3)	0.127
Scratch dermatitis	1 (2.1)	3 (6.5)	0	0.605
Periungual erythema	11 (22.9)	18 (39.1)	3 (25.0)	0.245
Mechanic's hands	20 (41.7)	34 (73.9)	3 (25.0)	**0.001**
Fever	4 (8.3)	17 (37.0)	0	**0.001**
Raynaud's phenomenon	0	16 (34.8)	9 (75.0)	**<0.001**
Nailfold capillary abnormality	9 (18.8)	24 (52.2)	4 (33.3)	**0.002**
Puffy fingers	1 (2.1)	12 (26.1)	9 (75.0)	**<0.001**
Sclerodactyly	0	3 (6.5)	8 (66.7)	**<0.001**
Proximal scleroderma	0	2 (4.3)	4 (33.3)	**<0.001**
Skin ulcer	0	5 (10.9)	0	**0.045**
Malignancy diagnosed within 5 years before or after disease diagnosis	10 (20.8)	2 (4.3)	1 (8.3)	**0.042**
Serum biomarkers at initial presentation, median (IQR)				
CK, U/l	107 (63–144)	451 (92–1185)	64 (51–101)	**<0.001**
Aldolase, U/l ($n = 65$)	5.8 (4.3–7.4)	12.2 (6.4–24.7)	3.8 (3.2–4.5)	NA
CRP, mg/dl	0.32 (0.07–1.35)	0.56 (0.19–2.14)	0.28 (0.07–0.72)	0.371
ESR, mm/h ($n = 48$)	25 (1–41)	20 (15–37)	20 (8–34)	NA

(continued)

Table 1. (continued)

Characteristics	Cluster 1 ($n = 48$)	Cluster 2 ($n = 46$)	Cluster 3 ($n = 12$)	P-value
Ferritin, ng/ml ($n = 66$)	184.6 (144.7–248.2)	182.4 (103.3–487.8)	103.5 (100.8–116.4)	NA
KL-6, U/ml	1214.4 (763.0–2147.5)	983.9 (547.9–1335.3)	643.9 (290.3–903.6)	**0.008**
SP-D, ng/ml ($n = 94$)	267.0 (151.6–429.0)	166.1 (103.8–264.2)	163.9 (104.6–235.4)	NA
Initial treatment regimen				
Glucocorticoids	38 (79.2)	45 (97.8)	5 (41.7)	**<0.001**
Pulse methylprednisolone	17 (35.4)	20 (43.5)	4 (33.3)	0.719
Initial glucocorticoid dose (mg/day[c])	30 (30–40)	40 (30–45)	30 (22–38)	0.336
TAC	26 (54.2)	30 (65.2)	0	**<0.001**
CYA	2 (4.2)	2 (4.3)	1 (8.3)	0.654
MTX	0	2 (4.3)	0	0.401
AZA	0	2 (4.3)	0	0.401
MMF	0	1 (2.2)	0	0.547
IVCY	8 (16.7)	14 (30.4)	0	**0.046**
IVIG	1 (2.1)	0	0	1.000

AZA: azathioprine; CK: creatinine phosphokinase; CYA: ciclosporin A; DLCO: diffusing capacity of lung for carbon monoxide; FVC: forced vital capacity; IMNM: immune-mediated necrotizing myopathy; IVCY: intravenous cyclophosphamide; KL-6: Krebs von den Lungen 6; NSIP: non-specific interstitial pneumonia; OP: organizing pneumonia; SP-D: surfactant protein D; TAC: tacrolimus.
Values are presented as n (%) unless stated otherwise.
Statistically significant values are in bold.
Statistical analysis was not performed for the PaO$_2$:FiO$_2$ ratio, alveolar-arterial oxygen gradient, %FVC, %DLCO, aldolase, ESR, ferritin or SP-D due to the high frequency of missing data.
  [a] Positive for both anti-Jo-1 and anti-EJ antibodies.
  [b] Positive for both anti-EJ and anti-KS antibodies.
  [c] Prednisolone/prednisone-equivalent dose.
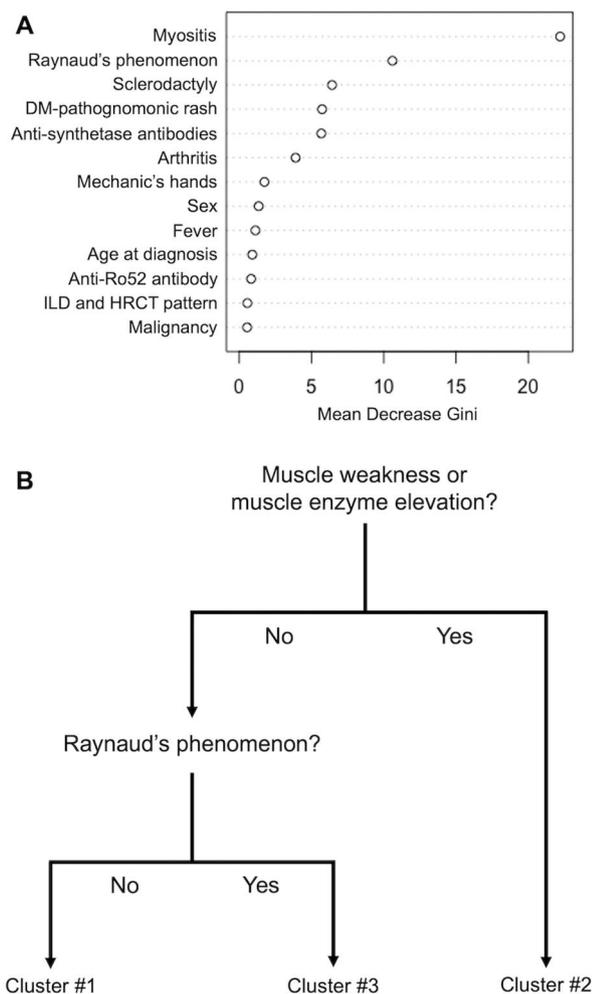
Figure 2. Generation of a classification tree using the SMCE cohort. **(A)** The weight of each clinical or serological variable for clustering assessed by the random forest algorithm. **(B)** Pruned model of prediction generated by the classification and regression tree analysis

(52.3%)], respectively. Arthritis, fever, RP and sclerodactyly were all more prevalent in cluster 2 ($P = 0.02$, 0.001, <0.001 and 0.001, respectively). Interestingly, the prevalence of each anti-synthase antibody was different between the two clusters ($P < 0.001$), with a higher prevalence of anti-PL-12, anti-EJ, anti-OJ and anti-KS antibodies in cluster 1 and anti-Jo-1 and anti-PL-7 antibodies in cluster 2.

The classification tree generated in the SMCE cohort performed well in the JAMI cohort, which positioned 88/125 patients (70.4%) into correct clusters (Supplementary Fig. S3B, available *Rheumatology Advances in Practice* online).

Outcomes

Of 106 patients with anti-synthetase antibodies in the SMCE cohort, 14 (13.2%) patients died and 36 (34.0%) patients experienced disease progression during the median follow-up period of 40 months (IQR 18–77). There was no significant difference in either OS or PFS between the three clusters (Fig. 3A and B), while cluster 2 showed a trend towards worse PFS outcomes than the other clusters.

Of 125 patients with PM/DM-ILD and anti-synthetase antibodies registered in the JAMI cohort, 7 (5.6%) patients died during the median follow-up period of 23 months (IQR 7–42). There was no difference in OS between the two clusters (Supplementary Fig. S6, available *Rheumatology Advances in Practice* online).

Discussion

The use of an unsupervised clustering method to evaluate a cohort of unselected patients with anti-synthetase antibodies from the SMCE cohort identified three clinically distinct subgroups. Cluster 1 was composed mainly of patients with IPAF or ADM ('IPAF/ADM cluster') and was characterized by older age at diagnosis and a higher frequency of malignancy. Cluster 2 is regarded as the 'DM cluster', which included younger patients with higher prevalence of myositis, arthritis, DM pathognomonic rash, mechanic's hands and

Table 2. Clinical characteristics of 125 patients with anti-synthetase antibodies in the JAMI cohort, stratified by cluster

Characteristics	Cluster 1 ($n = 60$)	Cluster 2 ($n = 65$)	P-value
Demographics			
Age at diagnosis, years, median (IQR)	57 (47–68)	58 (45–66)	0.718
Female	44 (73.3)	48 (73.8)	1.000
Clinical diagnosis			**<0.001**
DM	11 (18.3)	34 (52.3)	
ADM	35 (58.3)	10 (15.4)	
PM/IMNM	14 (23.3)	21 (32.3)	
Anti-synthetase antibodies			**<0.001**
Anti-Jo-1	4 (6.7)	30 (46.2)	
Anti-PL-7	3 (5.0)	24 (36.9)	
Anti-PL-12	11 (18.3)	5 (7.7)	
Anti-EJ	27 (45.0)	5 (7.7)	
Anti-OJ	5 (8.3)	1 (1.5)	
Anti-KS	10 (16.7)	0	
Clinical features			
Arthritis	24 (40.0)	40 (61.5)	**0.02**
Muscle involvement			**<0.001**
No myositis	35 (58.3)	8 (12.3)	
Subclinical myositis	8 (13.3)	7 (10.8)	
Clinical myositis	17 (28.3)	50 (76.9)	
DM pathognomonic rash	26 (43.3)	36 (55.4)	0.212
Gottron's papules	21 (35.0)	29 (44.6)	0.361
Gottron's sign	10 (16.7)	13 (20.3)	0.650
Heliotrope rash	9 (15.0)	13 (20.6)	0.484
Mechanic's hands	33 (55.0)	30 (46.2)	0.373
Fever	15 (25.0)	35 (53.8)	**0.001**
Malignancy	0	5 (7.7)	0.059
Raynaud's phenomenon	5 (8.3)	24 (36.9)	**<0.001**
Sclerodactyly	0	10 (15.4)	**0.001**
Skin ulcers	0	4 (6.2)	0.120
Serum biomarkers, median (IQR)			
CK, U/l	124 (63–651)	1078 (410–2443)	**<0.001**
Aldolase, U/l ($n = 107$)	9.6 (6.5–25.7)	25.3 (16.5–51.3)	NA
CRP, mg/dl	0.69 (0.20–2.14)	1.01 (0.44–2.23)	0.198
ESR, mm/h ($n = 89$)	26.0 (13.0–50.0)	29.5 (15.8–61.0)	NA
Ferritin, ng/ml ($n = 81$)	165.2 (67.4–275.7)	241.4 (107.0–412.6)	NA
KL-6, U/ml ($n = 123$)	1093.0 (548.0–1834.0)	855.0 (612.0–1451.3)	0.304
SP-D, pg/ml ($n = 99$)	179.5 (118.0–291.1)	155.0 (101.0–232.0)	NA
Initial treatment regimen			
Glucocorticoids	58 (96.7)	65 (100)	0.228
Pulse methylprednisolone	27 (45.0)	32 (49.2)	0.721
Initial glucocorticoid dose, mg/day[a], median (IQR)	40 (39–55)	45 (40–50)	0.420
TAC	27 (45.0)	36 (55.4)	0.285
CYA	18 (30.0)	25 (38.5)	0.351
MTX	1 (1.7)	5 (7.7)	0.210
AZA	1 (1.7)	5 (7.7)	0.210
MMF	0	1 (1.5)	1.000
IVCY	12 (20.0)	14 (21.5)	1.000
IVIG	1 (1.7)	8 (12.3)	**0.034**
RTX	0	1 (1.5)	1.000

AZA: azathioprine; CK: creatinine phosphokinase; CYA, ciclosporin A; IMNM: immune-mediated necrotizing myopathy; IVCY: intravenous cyclophosphamide; KL-6: Krebs von den Lungen 6; RTX: rituximab; SP-D: surfactant protein-D; TAC: tacrolimus.
Values are presented as n (%) unless stated otherwise.
Statistically significant values are in bold.
Statistical analysis was not performed for aldolase, ESR, ferritin or SP-D due to the high frequency of missing data.
  [a] Prednisolone/prednisone-equivalent dose.

fever. Interestingly, none of the patients in cluster 3 had myositis, and this cluster was regarded as the 'SSc cluster' because of a high prevalence of SSc-related manifestations. This cluster was associated with chronic ILD with a predominant UIP pattern on chest HRCT, and patients in this cluster might correspond to the previously reported subgroup of patients with anti-synthetase antibodies presenting with the SSc phenotype [7, 8]. Despite clinical heterogeneity among clusters, there was no difference in prognosis. A simple-to-use

classification tree using myositis and RP was generated with the SMCE cohort. Clusters 1 and 2 were successfully reproduced and the utility of the classification tree was validated in the JAMI cohort. Our results have implications for better understanding the heterogeneity of clinical manifestations in patients with anti-synthetase antibodies.

Hervier *et al.* [12] conducted cluster analysis in 233 patients with ASyS in a French multicentre cohort and found that two of the three clusters were characterized mainly by
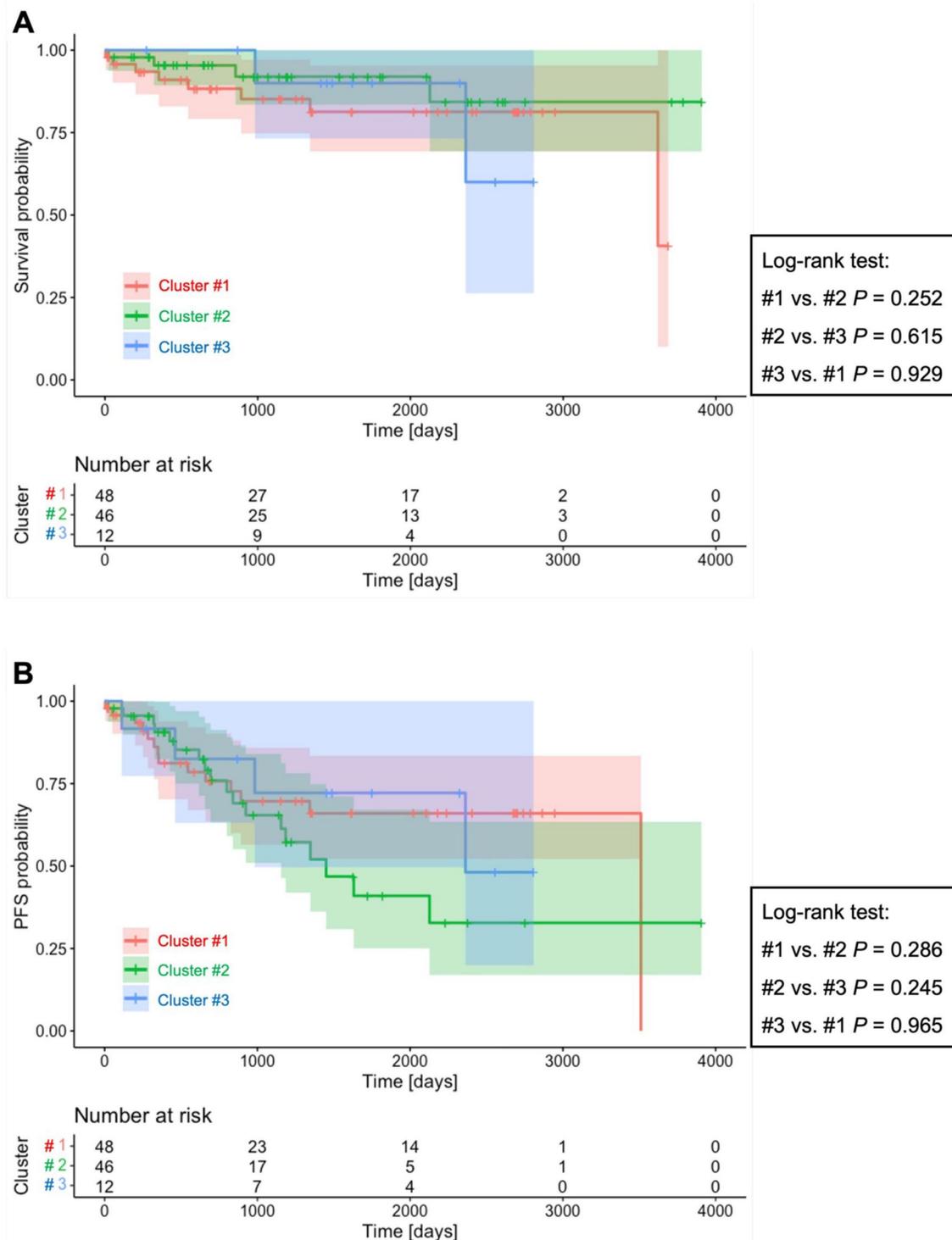
Figure 3. Kaplan–Meier survival curves for each cluster in the SMCE cohort. **(A)** Overall survival. **(B)** Progression-free survival. [a]Adjusted *P*-value

specificities of the anti-synthetase antibodies, including anti-Jo-1 and anti-PL-7/PL-12. In contrast, in the SMCE cohort, myositis and RP obtained the highest weight for clustering. The inconsistent clustering between the studies could be explained by the difference in the inclusion criteria. The study by Hervier *et al.* [12] enrolled patients with restricted profiles, including those who were positive for anti-Jo-1, PL-7 or PL-12 antibodies alone and patients who met arbitrary criteria (i.e. anti-synthetase antibody positivity plus clinical manifestations compatible with ASyS). On the other hand, our

study included unselected patients with anti-synthetase antibodies identified by RNA-IP, regardless of their clinical diagnosis. This indicates that phenotypes identified by cluster analysis depend on the patients enrolled due to the heterogeneity of patients with anti-synthetase antibodies.

Clusters 1 and 2 in the SMCE cohort were successfully reproduced in the JAMI cohort. Meanwhile, a cluster corresponding to cluster 3 in the SMCE cohort was not identified in the JAMI cohort. This could also be explained by the inconsistency in the inclusion criteria for the two cohorts;

patients with SSc alone were not eligible for the JAMI cohort. Interestingly, the prevalence of each anti-synthase antibody was different between the two clusters in the JAMI cohort, while in the SMCE cohort, anti-synthetase antibodies did not obtain a high weight for clustering, as demonstrated by the random forest model. However, there was a trend towards a higher prevalence of anti-Jo-1 and PL-7 and a lower prevalence of anti-KS in cluster 2 in the SMCE cohort, consistent with the results in the JAMI cohort. Importantly, patients classified as IPAF were not involved in the JAMI cohort, which could also lead to significant differences in the clinical and serological profile of patients included in cluster 1 between the SMCE and JAMI cohorts.

Our clustering failed to predict future outcomes in terms of OS and PFS. This suggests that patients with anti-synthetase antibodies have a consistent treatment response and prognosis as a group, but we need to consider that treatment regimens were decided by attending physicians based on clinical presentation, disease activity and prognostic factors in individual patients. The prevalence of RP-ILD and respiratory parameters, such as the peripheral capillary oxygen saturation/fraction of inspiratory oxygen ratio and pulmonary function parameters, were similar across the clusters, although the mode of ILD onset (acute/subacute *vs* chronic) and ILD patterns on HRCT (NSIP and/or OP *vs* UIP) were somewhat different. In fact, tacrolimus or intravenous cyclophosphamide were more commonly used in clusters 1 and 2 than in cluster 3. In addition, the median follow-up periods for the SMCE and JAMI cohorts were 40 and 23 months, respectively, which might not be sufficient to evaluate prognosis in patients with anti-synthetase antibodies, who have favourable short-term survival [13, 23, 24].

The strength of our study is the use of two independent cohorts for derivation and validation. In addition, patients were enrolled in these cohorts not only by rheumatologists, but also by pulmonologists, and the cohorts potentially captured the wide disease spectrum of patients with anti-synthetase antibodies in the real-world setting. On the other hand, we should acknowledge several limitations. First, the inclusion criteria for the SMCE and JAMI cohorts were somewhat different, which limited the assessment of cluster reproducibility. The clusters identified in this study should be assessed for reproducibility using additional multicentre validation cohorts. Second, in the SMCE cohort, pulmonary function tests and chest HRCT were not performed at the timing of the pre-defined protocol, hindering the evaluation of PFS. In addition, the SMCE is unique in its dual speciality in both IIMs and SSc, resulting in the overrepresentation of SSc patients with anti-synthetase antibodies.

In summary, we identified three phenotypes with distinct clinical characteristics using cluster analysis in a cohort of unselected patients with anti-synthetase antibodies. Our results will help to improve our understanding of the substantial heterogeneity in this disease entity and provide insights into the development of shared classification criteria for ASyS, which is currently in progress.

Supplementary material

Supplementary material is available at *Rheumatology Advances in Practice* online.

Masaki Terada (Saiseikai Niigata Daini Hospital) and Yoshie Kawahara (Keiyu Hospital).

References

1. McHugh NJ, Tansley SL. Autoantibodies in myositis. Nat Rev Rheumatol 2018;14:290–302.
2. Huang K, Aggarwal R. Antisynthetase syndrome: a distinct disease spectrum. J Scleroderma Relat Disord 2020;5:178–91.
3. Kuwana M, Gil-Vila A, Selva-O'Callaghan A. Role of autoantibodies in the diagnosis and prognosis of interstitial lung disease in autoimmune rheumatic disorders. Ther Adv Musculoskelet Dis 2021;13:1759720X211032457.
4. Zanframundo G, Faghihi-Kashani S, Scirè CA et al. Defining antisynthetase syndrome: a systematic literature review. Clin Exp Rheumatol 2022;40:309–19.
5. Fischer A, Antoniou KM, Brown KK et al. An official European Respiratory Society/American Thoracic Society research statement: interstitial pneumonia with autoimmune features. Eur Respir J 2015;46:976–87.
6. Graham J, Bauer Ventura I, Newton CA et al. Myositis-specific antibodies identify a distinct interstitial pneumonia with autoimmune features phenotype. Eur Respir J 2020;56:2001205.
7. Pauling JD, Salazar G, Lu H et al. Presence of anti-eukaryotic initiation factor-2B, anti-RuvBL1/2 and anti-synthetase antibodies in patients with anti-nuclear antibody negative systemic sclerosis. Rheumatology (Oxford) 2018;57:712–7.
8. Yoshida A, Gono T, Okazaki Y et al. Severe digital ischemia as an unrecognized manifestation in patients with antisynthetase autoantibodies: case series and systematic literature review. J Scleroderma Relat Disord 2022;7:204–16.
9. Han L, Benseler SM, Tyrrell PN. Cluster and multiple correspondence analyses in rheumatology: paths to uncovering relationships in a sea of data. Rheum Dis Clin North Am 2018;44:349–60.e29.
10. Mariampillai K, Granger B, Amelin D et al. Development of a new classification system for idiopathic inflammatory myopathies based on clinical manifestations and myositis-specific autoantibodies. JAMA Neurol 2018;75:1528–37.
11. Allenbach Y, Uzunhan Y, Toquet S et al. Different phenotypes in dermatomyositis associated with anti-MDA5 antibody: study of 121 cases. Neurology 2020;95:e70–8.
12. Hervier B, Devilliers H, Stanciu R et al. Hierarchical cluster and survival analyses of antisynthetase syndrome: phenotype and outcome are correlated with anti-tRNA synthetase antibody specificity. Autoimmun Rev 2012;12:210–7.
13. Sato S, Masui K, Nishina N et al. Initial predictors of poor survival in myositis-associated interstitial lung disease: a multicentre cohort of 497 patients. Rheumatology (Oxford) 2018; 57:1212–21.
14. Bohan A, Peter JB. Polymyositis and dermatomyositis (first of two parts). N Engl J Med 1975;292:344–7.
15. Sontheimer RD. Would a new name hasten the acceptance of amyopathic dermatomyositis (dermatomyositis siné myositis) as a distinctive subset within the idiopathic inflammatory dermatomyopathies spectrum of clinical illness? J Am Acad Dermatol 2002;46:626–36.
16. Lundberg IE, Tjärnlund A, Bottai M et al. 2017 European League Against Rheumatism/American College of Rheumatology classification criteria for adult and juvenile idiopathic inflammatory myopathies and their major subgroups. Ann Rheum Dis 2017; 76:1955–64.
17. 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;65:2737–47.
18. Hirakata M, Suwa A, Nagai S et al. Anti-KS: identification of autoantibodies to asparaginyl-transfer RNA synthetase associated with interstitial lung disease. J Immunol 1999;162:2315–20.
19. Shi J, Li S, Yang H et al. Clinical profiles and prognosis of patients with distinct antisynthetase autoantibodies. J Rheumatol 2017; 44:1051–7.
20. Bozzalla-Cassione E, Zanframundo G, Biglia A et al. Anti-Ro52 antibodies positivity in antisynthetase syndrome: a single centre cohort study. Clin Exp Rheumatol 2022;40(Suppl 134):27–31.
21. Charrad M, Ghazzali N, Boiteau V, Niknafs A. NbClust: an R package for determining the relevant number of clusters in a data set. J Stat Softw 2014;61:1–36.
22. Kanda Y. Investigation of the freely available easy-to-use software 'EZR' for medical statistics. Bone Marrow Transplant 2013; 48:452–8.
23. Gono T, Masui K, Nishina N et al. Risk prediction modeling based on a combination of initial serum biomarker levels in polymyositis/dermatomyositis-associated interstitial lung disease. Arthritis Rheumatol 2021;73:677–86.
24. Karampitsakos T, Tzilas V, Papaioannou O et al. Clinical features and outcomes of patients with myositis associated-interstitial lung disease. Front Med (Lausanne) 2022;9:1096203.