



HAL
open science

**Identification of distinct subgroups of Sjögren's disease
by cluster analysis based on clinical and biological
manifestations: data from the cross-sectional
Paris-Saclay and the prospective ASSESS cohorts**

Yann Nguyen, Gaëtane Nocturne, Julien Henry, Wan-Fai Ng, Rakiba Belkhir,
Frédéric Desmoulins, Elisabeth Bergé, Jacques Morel, Aleth Perdriger,
Emmanuelle Dernis, et al.

► **To cite this version:**

Yann Nguyen, Gaëtane Nocturne, Julien Henry, Wan-Fai Ng, Rakiba Belkhir, et al.. Identification of distinct subgroups of Sjögren's disease by cluster analysis based on clinical and biological manifestations: data from the cross-sectional Paris-Saclay and the prospective ASSESS cohorts. *The Lancet Rheumatology*, 2024, 10.1016/S2665-9913(23)00340-5 . hal-04489543

HAL Id: hal-04489543

<https://hal.science/hal-04489543>

Submitted on 8 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Identification of distinct subgroups of Sjögren's disease by cluster analysis based on clinical and biological manifestations: data from the cross-sectional Paris-Saclay and the prospective ASSESS cohorts

Yann Nguyen ¹, Gaëtane Nocturne ¹, Julien Henry ², Wan-Fai Ng ³, Rakiba Belkhir ², Frédéric Desmoulins ², Elisabeth Bergé ², Jacques Morel ⁴, Aleth Perdriger ⁵, Emmanuelle Dernis ⁶, Valérie Devauchelle-Pensec ⁷, Damien Sène ⁸, Philippe Dieudé ⁹, Marion Couderc ¹⁰, Anne-Laure Fauchais ¹¹, Claire Larroche ¹², Olivier Vittecoq ¹³, Carine Salliot ¹⁴, Eric Hachulla ¹⁵, Véronique Le Guern ¹⁶, Jacques-Eric Gottenberg ¹⁷, Xavier Mariette ¹, Raphaële Seror ¹⁸

¹ Department of Rheumatology, Hôpital Bicêtre, Assistance Publique - Hôpitaux de Paris, Université Paris-Saclay, Paris, France; Center for Immunology of Viral Infections and Auto-immune Diseases (IMVA), Institut pour la Santé et la Recherche Médicale (INSERM) UMR 1184, Université Paris-Saclay, Paris, France.

² Department of Rheumatology, Hôpital Bicêtre, Assistance Publique - Hôpitaux de Paris, Université Paris-Saclay, Paris, France.

³ Faculty of Medical Sciences, Clinical and Translational Research Institute, Newcastle University, NIHR Newcastle Biomedical Research Centre and NIHR Newcastle Clinical Research Facility, Newcastle upon Tyne NHS Hospitals Foundation Trust, Newcastle upon Tyne, UK.

⁴ Rheumatology Department, CHU de Montpellier, PhyMedExp, Université de Montpellier, INSERM, CNRS, Montpellier, France.

⁵ Rheumatology Department, CHU Rennes, Université Rennes, Rennes, France.

⁶ Department of Rheumatology and Clinical Immunology, General Hospital, Le Mans, France.

⁷ Department of Rheumatology, CHU de Brest, INSERM 1227, LBAI, Université de Bretagne Occidentale, Centre de Référence des Maladies Auto-Immunes Rares de l'Adulte, Brest, France.

⁸ Department of Internal Medicine, Hôpital Lariboisière, Assistance Publique - Hôpitaux de Paris, Paris, France.

⁹ Department of Rheumatology, Hôpital Bichat-Claude Bernard, Assistance Publique - Hôpitaux de Paris, INSERM UMR1152, Paris-Cité University, Paris, France.

¹⁰ Department of Rheumatology, CHU de Clermont-Ferrand, INSERM UMR 1240, Clermont Auvergne University, Clermont-Ferrand, France.

¹¹ Department of Internal Medicine, University Hospital of Limoges, Limoges, France.

¹² Department of Internal Medicine, Assistance Publique - Hôpitaux de Paris, Hôpital Avicenne, Bobigny, France.

¹³ Department of Rheumatology, Rouen University Hospital, Rouen, France.

¹⁴ Department of Rheumatology, Centre Hospitalier Universitaire d'Orléans, Orléans, France.

¹⁵ Department of Internal Medicine and Clinical Immunology, Hôpital Claude Huriez, University of Lille, Lille, France.

¹⁶ National Referral Centre for Rare Autoimmune and Systemic Diseases, Department of Internal Medicine, Hôpital Cochin, Assistance Publique - Hôpitaux de Paris Centre, Université Paris Cité, Paris, France.

¹⁷ Rheumatology Department, EA 3432, Hôpitaux Universitaires de Strasbourg, Université de Strasbourg, France.

¹⁸ Department of Rheumatology, Hôpital Bicêtre, Assistance Publique - Hôpitaux de Paris, Université Paris-Saclay, Paris, France; Center for Immunology of Viral Infections and Auto-immune Diseases (IMVA), Institut pour la Santé et la Recherche Médicale (INSERM) UMR 1184, Université Paris-Saclay, Paris, France. Electronic address: raphaelle.seror@aphp.fr.

Abstract

Background:

Sjögren's disease is a heterogeneous autoimmune disease with a wide range of symptoms including dryness, fatigue, and pain in addition to systemic manifestations and an increased risk of lymphoma. We aimed to identify distinct subgroups of the disease, using cluster analysis based on subjective symptoms and clinical and biological manifestations, and to compare the prognoses of patients in these subgroups.

Methods:

This study included patients with Sjögren's disease from two independent cohorts in France: the cross-sectional Paris-Saclay cohort and the prospective Assessment of Systemic Signs and Evolution of Sjögren's Syndrome (ASSESS) cohort. We first used an unsupervised multiple correspondence analysis to identify clusters within the Paris-Saclay cohort using 26 variables comprising patient-reported symptoms and clinical and biological manifestations. Next, we validated these clusters using patients from the ASSESS cohort. Changes in disease activity (measured by the European Alliance of Associations for Rheumatology [EULAR] Sjögren's Syndrome Disease Activity Index [ESSDAI]), patient-acceptable symptom state (measured by the EULAR Sjögren's Syndrome Patient Reported Index [ESSPRI]), and lymphoma incidence during follow-up were compared between clusters. Finally, we compared our clusters with the symptom-based subgroups previously described by Tarn and colleagues.

Findings:

534 patients from the Paris-Saclay cohort (502 [94%] women, 32 [6%] men, median age 54 years [IQR 43-64]), recruited between 1999 and 2022, and 395 patients from the ASSESS cohort (370 [94%] women, 25 [6%] men, median age 53 years [43-63]), recruited between 2006 and 2009, were included in this study. In both cohorts, hierarchical cluster analysis revealed three distinct subgroups of patients: those with B-cell active disease and low symptom burden (BALS), those with high systemic disease activity (HSA), and those with low systemic disease activity and high symptom burden (LSAHS). During follow-up in the ASSESS cohort, disease activity and symptom states worsened for patients in the BALS cluster (67 [36%] of 186 patients with ESSPRI score <5 at month 60 vs 92 [49%] of 186 at inclusion; $p < 0.0001$). Lymphomas occurred in patients in the BALS cluster (five [3%] of 186 patients; diagnosed a median of 70 months [IQR 42-104] after inclusion) and the HSA cluster (six [4%] of 158 patients; diagnosed 23 months [13-83] after inclusion). All patients from the Paris-Saclay cohort with a history of lymphoma were in the BALS and HSA clusters. This unsupervised clustering classification based on symptoms and clinical and biological manifestations did not correlate with a previous classification based on symptoms only.

Interpretation:

On the basis of symptoms and clinical and biological manifestations, we identified three distinct subgroups of patients with Sjögren's disease with different prognoses. Our results suggest that these subgroups represent different heterogeneous pathophysiological disease mechanisms, stages of disease, or both. These findings could be of interest when stratifying patients in future therapeutic trials.

Funding:

Fondation pour la Recherche Médicale, French Ministry of Health, French Society of Rheumatology, Innovative Medicines Initiative 2 Joint Undertaking, Medical Research Council UK, and Foundation for Research in Rheumatology.

Introduction

Sjögren's disease is a common systemic autoimmune exocrinopathy with a female-to-male predominance of 9:1 and a peak incidence at approximately 50 years of age.^{1,2} Symptoms including oral or ocular dryness, fatigue, and joint pain are present in almost all patients and have a major effect on quality of life. In addition, systemic manifestations occur in approximately 30–60% of patients.¹ Such manifestations can be related to the lymphocytic infiltration of the epithelia of organs, leading to interstitial nephritis, autoimmune primary biliary cholangitis, or obstructive bronchiolitis, and to vasculitis or immune complex deposition owing to ongoing B-cell hyper-reactivity, leading to purpura, cryoglobulinemia-associated glomerulonephritis, interstitial pneumonitis, or peripheral neuropathy. A major complication of Sjögren's disease is B-cell lymphoma, of which the risk is approximately 14–15 times higher than in the general population. Patients with Sjögren's disease are therefore heterogeneous in terms of clinical symptoms, systemic manifestations, and risks.

Hierarchical cluster analysis is an unsupervised statistical method of data partitioning whereby individuals are categorised into homogeneous groups on the basis of similarity.^{3–6} Previous attempts to use this technique to identify distinct clusters of patients with Sjögren's disease have yielded useful results; however, those classifications were based only on patient-reported symptoms and did not consider all aspects of the disease.^{7,8} Previous studies have shown weak correlations between subjective symptoms and objective dryness measures or systemic complications, highlighting the need to consider all aspects of Sjögren's disease in such analyses.^{9,10}

This study aimed to identify distinct subgroups of Sjögren's disease using cluster analysis on two independent patient cohorts—including not only symptoms but also demographic characteristics, clinical parameters, and biological data—and to prospectively compare disease evolution and the incidence of lymphoma in the different clusters.

Methods

Study design and participants

This study included patients with Sjögren's disease from two independent cohorts in France: the Paris-Saclay cohort and the Assessment of Systemic Signs and Evolution of Sjögren's Syndrome (ASSESS) cohort. The Paris-Saclay cohort is a prospectively collected cohort that includes all patients who participated in a multidisciplinary diagnostic session, in a French National Referral Center for Rare Systemic Autoimmune Diseases, for those suspected of having Sjögren's disease. All patients were recruited between 1999 and 2022 and gave oral informed consent for the collection of their data. Data collection was approved by the Bicêtre Hospital ethics committee.

The ASSESS cohort was created in 2006 as part of a prospective, multicentre, national cohort study. Its primary objective is to identify factors predicting systemic complications and lymphoma in Sjögren's disease during a prospective 20-year follow-up.^{11–13} Patients who met the 2002 American-European Consensus Group (AECG) criteria for Sjögren's disease¹⁴ were recruited from 15 French centres for tertiary autoimmune diseases between 2006 and 2009.¹⁵ The study, promoted first by the Assistance Publique – Hôpitaux de Paris and then

by the Société Française de Rheumatology for the 20-year follow-up, was approved by the Ethics Committee of Bichat Hospital and the National Commission for Computing and Liberties in 2006. All patients gave their written informed consent for participation in the study.

For the current study, we included only patients fulfilling the 2002 AECG criteria for Sjögren's disease.¹⁴ For patients enrolled in the Paris-Saclay cohort before 2002, the 2002 AECG criteria were retrospectively verified. Patients enrolled in both cohorts were analysed only in the ASSESS cohort and were therefore excluded from the Paris-Saclay cohort. Also excluded from the Paris-Saclay cohort were patients who had another defined autoimmune disease associated with Sjögren's disease or those for whom scores on the European Alliance of Associations for Rheumatology (EULAR) Sjögren's Syndrome

Disease Activity Index (ESSDAI)¹⁶ or the EULAR Sjögren's Syndrome Patient Reported Index (ESSPRI)¹⁷ were missing at inclusion.

Procedures

In both cohorts, the following data were collected at enrolment: age at diagnosis, sex (male or female), race (Black, White, or Asian; data available only for the Paris-Saclay cohort¹⁸), and smoking status (ever or never smokers). Subjective symptoms were collected on visual analogue scales (VAS, scored 0–100, for the Paris-Saclay cohort) or visual numeric scales (VNS, scored 0–10, for the ASSESS cohort) for pain, fatigue, and dryness, thus enabling the calculation of the ESSPRI.

Clinical systemic manifestations (yes or no) were defined by an activity level for each domain of the ESSDAI, which was assessed on entry to the cohort (ie, not cumulative). These manifestations were constitutional, lymphadenopathy, glandular, articular, pulmonary, cutaneous, renal, muscular, haematological, peripheral nervous system, and central nervous system involvement. In the ASSESS cohort, the activity level (none, mild, moderate, or severe) of each ESSDAI domain was collected on inclusion into the cohort, whereas in the Paris-Saclay cohort, such levels were retrospectively evaluated through chart reviews conducted by expert rheumatologists.

Biological data on autoantibodies (anti-Sjögren's syndrome A [anti-SSA], anti-Sjögren's syndrome B [anti-SSB], anti-ribonucleoprotein [anti-RNP], anti-centromere, and anti-DNA antibodies), cryoglobulinaemia, rheumatoid factor, γ globulins or IgG concentration, and presence of a monoclonal gammopathy were collected. All data were measured when patients were enrolled into the cohort.

Patients in the prospective ASSESS cohort attended an annual standardised clinical visit for the first 5 years after enrolment, during which a detailed case report form was completed to enable the yearly assessment of ESSDAI and ESSPRI. After the initial 5 years, the case report form was simplified so that only a proportion of the variables were recorded; however, to date, only cases of incident lymphoma have been analysed. Information regarding ongoing and previous treatments were also collected for patients in this cohort.

Disease activity states were defined according to previously described ESSDAI thresholds: 19 low activity (ESSDAI score <5), moderate activity (ESSDAI score 5–13), and high activity (ESSDAI score \geq 14). We also defined a state of no systemic activity (ESSDAI score=0).

The patient-acceptable symptom state was defined as an ESSPRI score of less than 5, as previously described.¹⁹

Statistical analysis

To identify clusters of patients with Sjögren's disease, we first used an unsupervised multiple correspondence analysis in the Paris-Saclay cohort, selecting 26 variables to broadly cover the manifestations of Sjögren's disease: age at diagnosis (years); sex (male or female); race (Black, White, or Asian); VNS (0–10) or VAS (0–100) for pain, fatigue, and dryness; systemic manifestations as defined by ESSDAI domains (yes or no); rheumatoid factor (present or absent); autoantibodies (anti-SSA, anti-SSB, anti-RNP, anti-centromere, or anti-DNA; present or absent); high γ globulin or IgG concentrations (>15 g/L); monoclonal gammopathy (present or absent); cryoglobulinaemia (present or absent); and low C4 concentrations (<0.15 μ mol/L). Because multiple correspondence analysis was conducted only on categorical variables, we divided age (0 to 30, >30 to 45, >45 to 60, and >60 years) as well as VNS (0 to 2, >2 to 4, >4 to 6, >6 to 8, and >8 to 10) and VAS (0 to 20, >20 to 40, >40 to 60, >60 to 80, and >80 to 100) for pain, fatigue, and dryness into several groups. When VAS for overall dryness was missing, it was calculated as follows: $(2 \times \text{oral dryness} + \text{ocular dryness}) / 3$, as previously described in the EULAR Sicca score derivation, in which only oral and ocular dryness were correlated with overall dryness in multivariate analysis.¹⁷

We then considered the coordinates of the observations of the factorial axes obtained by multiple correspondence analysis as new variables for the cluster analysis. The first k-axes, explaining at least 90% of the total variability, were used to conduct hierarchical clustering based on the Ward method, followed by consolidation (k-means algorithms) to build homogeneous clusters of patients with Sjögren's disease. The number of clusters was identified visually on the plotted dendrogram and by the gain in inertia, and the clusters were described and named according to their most prominent characteristics.

Next, to validate our obtained clusters, we applied the same method in the ASSESS cohort as a replication cohort. More detailed information on the methods is available in the appendix (pp 2–3).

For descriptive analyses, categorical variables were presented as counts and percentages and continuous variables as medians (IQR). We compared subgroups using the Kruskal–Wallis rank sum test for continuous variables; for categorical variables, we used Pearson's χ^2 test if the expected cell counts were greater than 5 and Fisher's exact tests otherwise.

To describe the evolution of disease activity and patient-reported symptoms in the obtained clusters, we calculated the proportion of patients with each disease activity state and patient-acceptable symptom state at each annual follow-up visit during the first 5 years of follow-up in the ASSESS cohort. If ESSPRI or ESSDAI scores were missing during an annual follow-up visit, we imputed values using the scores from the previous visit. Changes in disease activity states and patient-acceptable symptom states were illustrated with alluvial diagrams.

To assess changes across the follow-up visits, we used ANOVAs for repeated measures within each cluster.

In addition, we assessed the associations between the different clusters and the risk of incident lymphoma (ie, occurring after inclusion in the cohort). This risk was assessed and compared for each cluster using survival analysis and Cox proportional hazards regression models to calculate hazard ratios (HR) and their 95% CIs. Patients contributed patient-years from the date of enrolment in the ASSESS cohort until the date of diagnosis of lymphoma, death, end of follow-up, or loss to follow-up, whichever occurred first.

Finally, we compared our clusters with the symptom-based subgroups previously described by Tarn and colleagues.⁷ Their stratification, based on five patient-reported outcomes (dryness, fatigue, pain, anxiety, and depression), was derived from a cluster analysis of data from the UK Primary Sjögren's Syndrome Registry (UKPSSR) and gave rise to four subgroups: low symptom burden, high symptom burden, dryness dominant with fatigue, and pain dominant with fatigue. The study led to the construction of a simple algorithm, the Newcastle Sjögren's Stratification Tool (NSST), which was applied on the ASSESS cohort. To compare the two tools, we described the proportions of patients in each of the symptom-based subgroups, according to the NSST, within each of our obtained clusters.

Two-tailed p values of 0.05 or less were considered statistically significant. All analyses were conducted with R version 4.3.1.

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

For the Paris-Saclay cohort, of the 1793 patients referred to the French National Referral Center for Rare Systemic Autoimmune Diseases, 748 (42%) met the 2002 AECG criteria. Among them, 534 (71%) met our inclusion criteria (ie, they had ESSDAI and ESSPRI scores available at inclusion and no other defined associated autoimmune disease) and were enrolled into the study. 502 (94%) were women, 32 (6%) were men, and the median age was 54 years (IQR 43–64; table 1). For the ASSESS cohort, all 395 patients recruited to the cohort were included in this analysis. 370 (94%) were women, 25 (6%) were men, with a median age of 53 years (44–60). ESSPRI scores were available for 361 (91%) and ESSDAI scores were available for all 395 (100%) patients in the ASSESS cohort. The median follow-up time for patients in this cohort was 147 months (IQR 93–167).

Multiple correspondence analysis followed by hierarchical clustering analysis in the Paris-Saclay cohort yielded three subgroups of patients (table 1), within which homogeneity was apparent both visually on the dendrogram (figure 1A) and in terms of the gain in inertia (appendix p 8). Patients in cluster one (n=205) were diagnosed slightly younger on average (median age 48 years [IQR 35–61]) than those in the other clusters, and had low symptom burden (median ESSPRI score 5 [IQR 3–7]) and low systemic disease activity (median ESSDAI score 2 [1–4]), but a high proportion had B-cell activation hallmarks: an active

biological ESSDAI domain (124 [60%] of 205 patients), high IgG concentrations (108 [53%]), presence of rheumatoid factor (113 [55%]), low C4 concentrations (53 [26%]), and a higher frequency of presence of anti-SSA (171 [83%]) and anti-SSB (95 [47%] of 202 patients) antibodies. This cluster was named B-cell active disease with low symptom burden (BALS).

	Paris-Saclay cohort				pvalue	ASSESS cohort				pvalue
	Total cohort (n=534)	BALS (n=205)	HSA (n=160)	LSAHS (n=169)		Total cohort (n=395)	BALS (n=186)	HSA (n=158)	LSAHS (n=51)	
Age at diagnosis, years*	54 (43-64)	48 (35-61)	53 (43-63)	57 (50-65)	<0.0001†	53 (44-60)	51 (38-59)	54 (46-61)	53 (48-58)	0.035†
Sex*					<0.0001†					0.033‡
Female	502 (94%)	205 (100%)	128 (80%)	169 (100%)		370 (94%)	169 (91%)	150 (95%)	51 (100%)	
Male	32 (6%)	0	32 (20%)	0		25 (6%)	17 (9%)	8 (5%)	0	
Race*					<0.0001†					-
Black	33 (6%)	0	33 (21%)	0		-	-	-	-	
Asian	34 (6%)	30 (15%)	4 (3%)	0		-	-	-	-	
White	467 (87%)	175 (85%)	123 (77%)	169 (100%)		-	-	-	-	
Ever smoker	104 (19%)	36 (18%)	32 (20%)	36 (21%)	0.60‡	105 (27%)	52 (28%)	41 (26%)	12 (24%)	0.90‡
Patient-reported outcomes										
Pain (VAS or VNS)*§	51 (20-77)	37 (10-60)	52 (20-78)	67 (48-80)	<0.0001†	5 (2-7)	4 (2-6)	6 (3-8)	6 (4-7)	<0.0001†
Fatigue (VAS or VNS)*§	62 (40-80)	49 (20-70)	61 (38-81)	76 (60-87)	<0.0001†	6 (4-8)	5 (3-8)	7 (5-8)	7 (6-8)	<0.0001†
Overall dryness (VAS or VNS)*§	61 (40-77)	51 (31-72)	66 (46-80)	65 (47-80)	<0.0001†	6 (4-7)	5 (3-7)	6 (5-8)	6 (5-7)	<0.0001†
ESSPRI score§	6 (4-7)	5 (3-6)	6 (4-7)	7 (5-8)	<0.0001†	6 (4-7)	5 (3-6)	6 (5-7)	6 (5-7)	<0.0001†
Systemic manifestations according to ESSDAI domains										
Constitutional*	6 (1%)	0	6 (4%)	0	0.0007¶	16 (4%)	0	16 (10%)	0	<0.0001‡
Lymphadenopathy*	29 (5%)	2 (1%)	27 (17%)	0	<0.0001‡	12 (3%)	0	12 (8%)	0	<0.0001‡
History of lymphoma before inclusion	24 (4%)	3 (1%)	19 (12%)	2 (1%)	<0.0001‡	18 (5%)	2 (1%)	16 (10%)	0	<0.0001‡
Glandular*	152 (28%)	78 (38%)	47 (29%)	27 (16%)	<0.0001‡	47 (12%)	24 (13%)	18 (11%)	5 (10%)	0.80¶
Articular*	176 (33%)	58 (28%)	58 (36%)	60 (36%)	0.20‡	73 (18%)	33 (18%)	32 (20%)	8 (16%)	0.70¶
Cutaneous*	14 (3%)	0	14 (9%)	0	<0.0001¶	16 (4%)	1 (1%)	15 (9%)	0	<0.0001‡
Pulmonary*	21 (4%)	0	21 (13%)	0	<0.0001‡	57 (14%)	5 (3%)	52 (33%)	0	<0.0001¶
Renal*	3 (1%)	0	3 (2%)	0	0.027¶	11 (3%)	0	11 (7%)	0	0.0001‡
Muscular	2 (<1%)	0	1 (1%)	1 (1%)	0.52¶	13 (3%)	4 (2%)	6 (4%)	3 (6%)	0.30‡
Peripheral nervous system*	20 (4%)	0	20 (13%)	0	<0.0001‡	38 (10%)	4 (2%)	34 (22%)	0	<0.0001‡
Central nervous system	1 (<1%)	0	1 (1%)	0	0.30¶	8 (2%)	4 (2%)	4 (3%)	0	0.80‡
Haematological*	78 (15%)	29 (14%)	34 (21%)	15 (9%)	0.0063‡	62 (16%)	32 (17%)	29 (18%)	1 (2%)	0.02¶
Biological*	266 (50%)	124 (60%)	95 (59%)	47 (28%)	<0.0001‡	146 (37%)	71 (38%)	72 (46%)	3 (6%)	<0.0001¶
ESSDAI score	2 (1-5)	2 (1-4)	6 (3-10)	2 (0-3)	<0.0001†	3 (2-8)	2 (1-4)	8 (3-13)	0 (0-2)	<0.0001†
Paraclinical parameter										
Lymphocyte count (G/L)§	2 (1-2)	2 (1-2)	2 (1-2)	2 (1-2)	0.023†	1 (1-2)	1 (1-2)	1 (1-2)	2 (1-2)	0.051†
IgG concentration > 15 g/L*	203 (38%)	108 (53%)	64 (40%)	31 (18%)	<0.0001‡	147 (37%)	91 (49%)	55 (35%)	1 (2%)	<0.0001¶
Monoclonal gammopathy*	44 (8%)	20 (10%)	22 (14%)	2 (1%)	0.0001‡	44 (11%)	8 (4%)	35 (22%)	1 (2%)	<0.0001¶
Rheumatoid factor*	234 (44%)	113 (55%)	71 (44%)	50 (30%)	<0.0001‡	122 (31%)	60 (32%)	50 (32%)	12 (24%)	0.47¶
Anti-SSA antibody*	368 (69%)	171 (83%)	101 (63%)	96 (57%)	<0.0001‡	232 (59%)	126 (68%)	85 (54%)	21 (41%)	0.0008¶
Anti-SSB antibody*	182/530 (34%)	95/202 (47%)	47/160 (26%)	45/168 (27%)	<0.0001‡	132 (33%)	76 (41%)	48 (30%)	8 (16%)	0.0019¶
Anti-RNP antibody*	15 (3%)	0	15 (9%)	0	<0.0001¶	5 (1%)	0	5 (3%)	0	0.028‡
Anti-centromere antibody*	20 (4%)	0	20 (13%)	0	<0.0001‡	9 (2%)	0	9 (6%)	0	0.0007‡
Anti-DNA antibody*	29 (5%)	23 (11%)	5 (3%)	1 (1%)	<0.0001‡	38 (10%)	25 (13%)	13 (8%)	0	0.0020‡
Cryoglobulinaemia*	10 (2%)	0	10 (6%)	0	<0.0001¶	57 (14%)	24 (13%)	27 (17%)	6 (12%)	0.46¶
Low C4 concentration*	92 (17%)	53 (26%)	28 (18%)	11 (7%)	<0.0001‡	72 (18%)	38 (20%)	34 (22%)	0	0.0014¶

Data are median (IQR), n (%), or n/N (%). Percentages might not total 100 where expected owing to rounding. BALS—B-cell active disease and a low symptom burden. ESSDAI—EULAR Sjögren's Syndrome Disease Activity Index. ESSPRI—EULAR Sjögren's Syndrome Patient Reported Index. EULAR—European Alliance of Associations for Rheumatology. HSA—high systemic disease activity. LSAHS—low systemic disease activity and a high symptom burden. RNP—ribonucleoprotein. SSA—Sjögren's syndrome A. SSB—Sjögren's syndrome B. VAS—visual analogue scale. VNS—visual numeric scale. *Variables included in the multiple correspondence analysis and the hierarchical clustering. VAS scores are given for the Paris-Saclay cohort and VNS scores for the ASSESS cohort. †Kruskal-Wallis rank sum test. ‡Fisher's exact test. §Data on lymphocyte count were available for 529 (99%) of 534 patients in the Paris-Saclay cohort and 383 (97%) of 395 patients in the ASSESS cohort. Of the patients in the ASSESS cohort, pain VNS was available for 366 (93%), fatigue VNS for 366 (93%), overall dryness VNS for 357 (90%), and ESSPRI score for 361 (91%). ¶Pearson's χ^2 test.

Table 1: Main characteristics of patients with Sjögren's disease in each of the derived clusters

Cluster two (n=160) had a higher proportion of men (32 [20%]) than other clusters, with high symptom burdens of dryness and fatigue but lower levels of pain (median VAS for pain 52 [IQR 20–78]). Patients in this cluster had high systemic disease activity (median ESSDAI score 6 [3–10]), as evidenced by lymphadenopathy (27 [17%]); cutaneous (14 [9%]), pulmonary (21 [13%]), or peripheral nervous system (20 [13%]) involvement; the frequent presence of autoantibodies other than anti-SSA or anti-SSB (15 [9%] with anti-RNP and 20 [13%] with anti-centromere antibodies); and cryoglobulinemia (ten [6%]). This cluster was named high systemic disease activity (HSA).

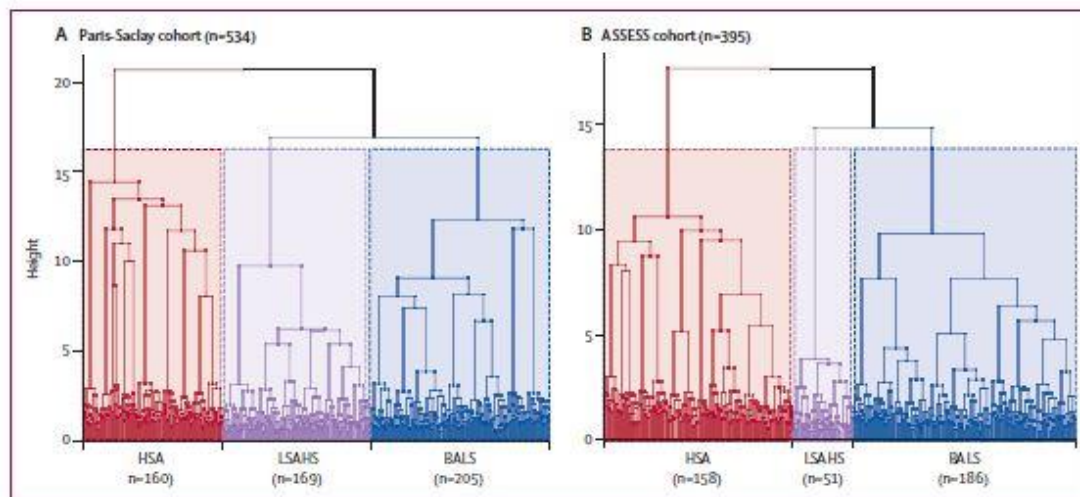


Figure 1: Dendrogram of identified clusters
 Clusters identified in the Paris-Saclay cohort (A) and the ASSESS cohort (B). Horizontal branches represent the combination of two clusters, and vertical branches the degree of dissimilarity between combined clusters. The areas enclosed within the dotted lines represent the three groups after truncation. ASSESS=Assessment of Systemic Signs and Evolution of Sjögren's Syndrome. BALS=B-cell active disease and low symptom burden. HSA=high systemic disease activity. LSAHS=low systemic disease activity and high symptom burden.

Cluster three (n=169) included patients with a high reported burden of subjective symptoms (median ESSPRI score 7 [IQR 6–8]) but low systemic disease activity (median ESSDAI score 2 [0–3]). Most disease activity involved the articular (60 [36%]), glandular (27 [16%]), and biological (47 [28%]) ESSDAI domains. This cluster was named low systemic disease activity with high symptom burden (LSAHS). Notably, all 33 (100%) Black patients within the Paris-Saclay cohort were in the HSA cluster. 30 (88%) of 34 Asian patients were in the BALS cluster and four (12%) were in the HSA cluster.

24 (4%) of the 534 patients in the Paris-Saclay cohort had a history of lymphoma; of these, 19 (79%) were in the HSA cluster, three (13%) were in the BALS cluster, and two (8%) were in the LSAHS cluster.

We next analysed the ASSESS cohort, using the same variables as for the Paris-Saclay cohort with the exception of race. Similarly to the Paris-Saclay cohort, this analysis also resulted in three distinct subgroups of patients (table 1), within which homogeneity was apparent both visually on the dendrogram (figure 1B) and in terms of the gain in inertia (appendix p 8).

The three clusters obtained were similar to those of the Paris-Saclay cohort in terms of characteristics, and were also termed BALS, HSA, and LSAHS. The LSAHS cluster of the ASSESS cohort contained a smaller proportion of patients (51 [13%] of 395) than the analogous cluster in the Paris-Saclay cohort (169 [32%] of 534). Men in the ASSESS cohort were grouped into the BALS (17 [9%] of 186) and HSA (eight [5%] of 158) clusters. The

difference in ESSDAI scores across clusters was greatest in the ASSESS cohort (median ESSDAI scores BALS 2 [IQR 1–4], HSA 8 [3–13], and LSAHS 0 [0–2] in the ASSESS cohort; BALS 2 [1–4], HSA 6 [3–10], and LSAHS 2 [0–3] in the Paris-Saclay cohort).

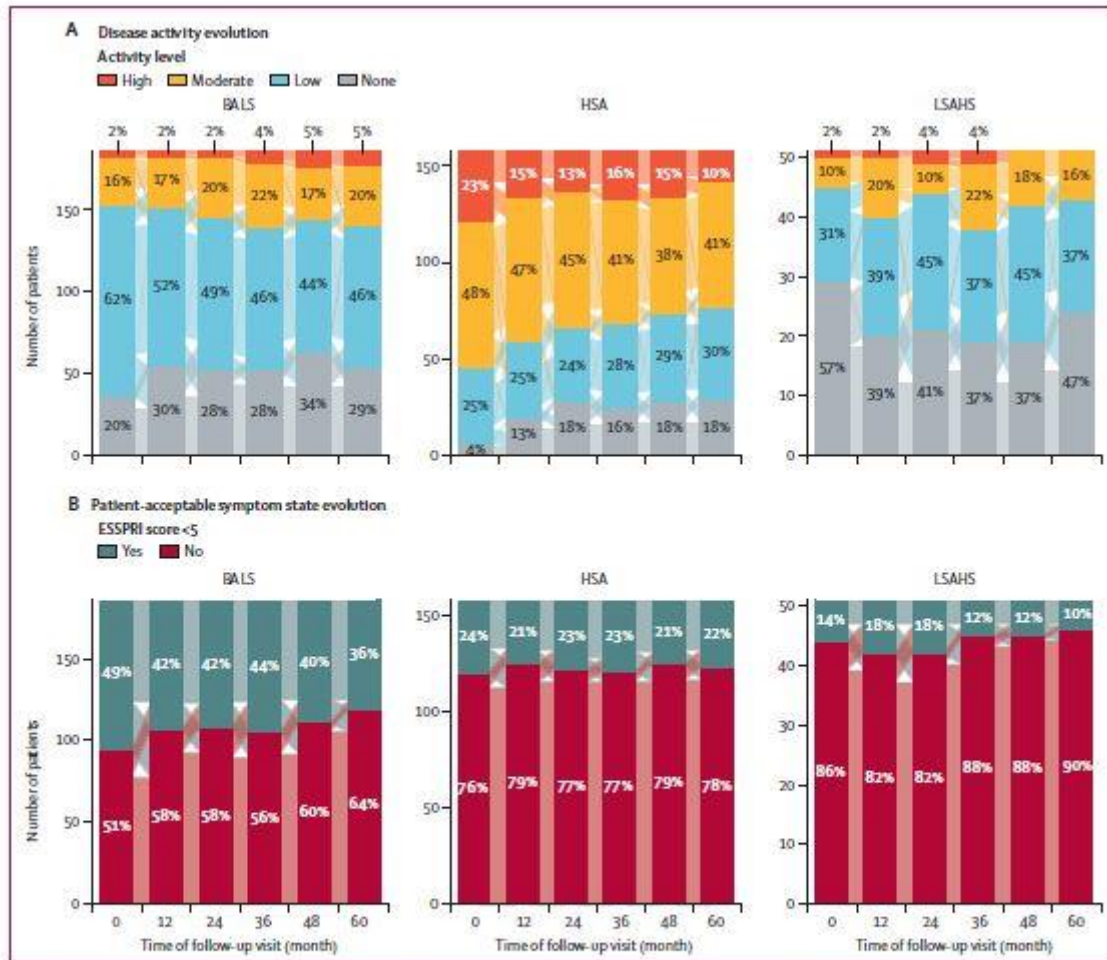


Figure 2: Evolution of disease activity and patient-acceptable symptom state over 5 years in each cluster of the ASSESS cohort (A) Evolution of disease activity in each cluster of the ASSESS cohort (n=395). Activity levels are defined as: none (ESSDAI score=0), low (ESSDAI score<5), moderate (ESSDAI score 5–13), and high (ESSDAI score ≥14). Percentages might not total 100 owing to rounding. (B) Changes in patient-acceptable symptom state, defined as an ESSPRI score <5, in each cluster for the ASSESS cohort (n=395). ASSESS=Assessment of Systemic Signs and Evolution of Sjögren's Syndrome. BALS=B-cell active disease and low symptom burden. ESSDAI=EULAR Sjögren's Syndrome Disease Activity Index. ESSPRI=EULAR Sjögren's Syndrome Patient Reported Index. EULAR= European Alliance of Associations for Rheumatology. HSA=high systemic disease activity. LSAHS=low systemic disease activity and high symptom burden.

222 (56%) of 395 patients in the ASSESS cohort received hydroxychloroquine, with a similar proportion receiving the drug in each cluster. However, only a few patients (six [12%] of 51) from the LSAHS cluster received immunosuppressive drugs, whereas patients in the HSA cluster were more frequently treated with glucocorticoids (56 [35%] of 158 patients) and more frequently had a history of treatment with rituximab (12 [8%]) than those in other clusters (appendix p 4).

Follow-up ESSDAI scores were available for 361 (91%) of 395 patients in the ASSESS cohort at the month 12 visit, 323 (82%) at month 24, 313 (79%) at month 36, 299 (76%) at month 48, and 291 (74%) at month 60. Changes in disease activity levels in each cluster over the 5-year follow-up period are shown in figure 2A. No substantial change was observed in the LSAHS cluster (p=0.17, comparing inclusion and month 60), whereas systemic disease activity levels improved in the HSA cluster (p<0.0001) and worsened in the BALS cluster (p=0.041). Details of the systemic manifestations between inclusion and 5 years in the BALS cluster are shown in the appendix (p 5). At the 5-year follow-up visit, 46 (25%) of 186

patients in the BALS cluster, 81 (51%) of 158 in the HSA cluster, and eight (16%) of 51 in the LSAHS cluster had moderate to high systemic disease activity levels ($p < 0.0001$).

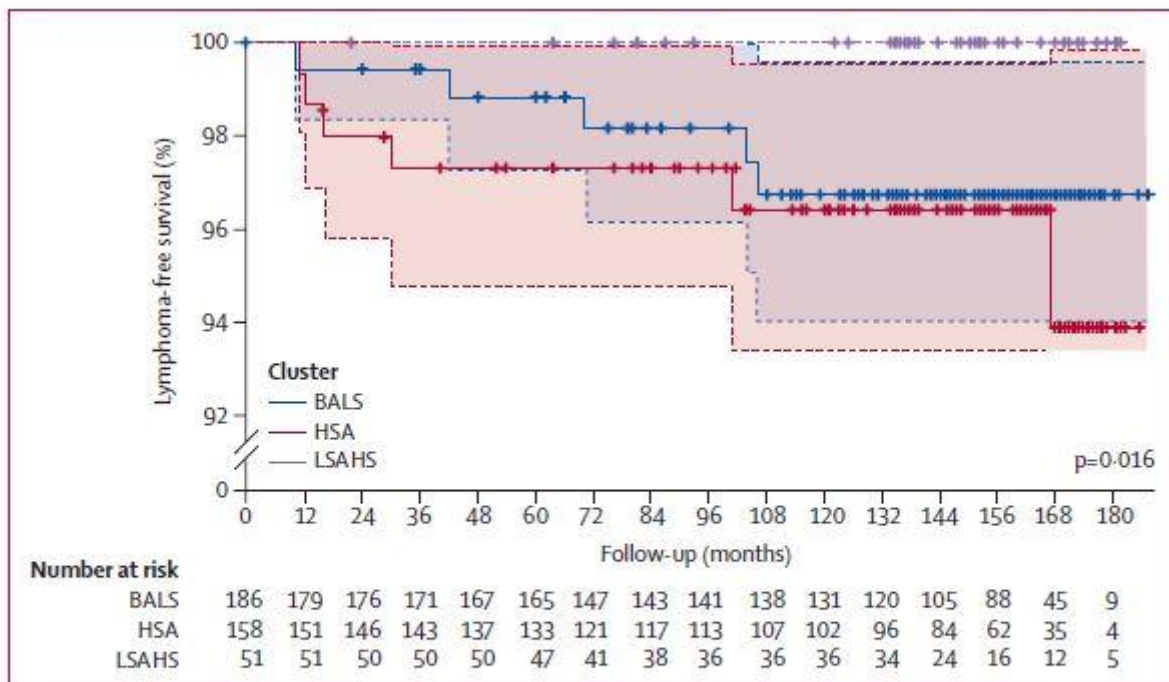


Figure 3: Kaplan-Meier plot estimating the risk of incident lymphoma in each cluster of the ASSESS cohort p-value was calculated by log-rank test. ASSESS=Assessment of Systemic Signs and Evolution of Sjögren's Syndrome. BALS=B-cell active disease and low symptom burden. HSA=high systemic disease activity. LSAHS=low systemic disease activity and high symptom burden. Shading represents the 95% CIs.

Follow-up ESSPRI scores were available for 326 (83%) of 395 patients in the ASSESS cohort at month 12, 308 (78%) at month 24, 295 (75%) at month 36, 283 (72%) at month 48, and 261 (66%) at month 60. The proportion of patients with patient-acceptable symptom states (ESSPRI score < 5) in each subgroup over the follow-up period is shown in figure 2B. In the BALS cluster, the proportion of patients with an ESSPRI of less than 5 was significantly lower at month 60 (67 [36%] of 186 patients) than at inclusion (92 [49%] of 186; $p < 0.0001$). Conversely, no substantial changes were observed in the HSA cluster (35 [22%] of 158 patients at month 60 vs 38 [24%] of 158 at inclusion; $p = 0.86$) or the LSAHS cluster (five [10%] of 51 patients at month 60 vs seven [14%] of 51 at inclusion; $p = 0.58$).

In the 180 months after inclusion in the ASSESS cohort, 11 (3%) of 395 patients were diagnosed with lymphoma, after a median of 42 months (IQR 14–103); 46 (12%) patients died before a diagnosis of lymphoma, and 338 (86%) were censored at their last-follow-up visit. Cases of lymphoma occurred in both the BALS cluster (five [3%] of 186 patients; diagnosed after a median of 70 months [IQR 42–104]) and the HSA cluster (six [4%] of 158 patients; diagnosed after 23 [13–83] months; log-rank $p = 0.016$; figure 3). No cases of lymphoma occurred in the LSAHS cluster.

We compared our clusters with the symptom-based subgroups described by Tarn and colleagues,⁷ assessed by the NSST, on which data were available for 316 (80%) of the 395 patients from the ASSESS cohort (figure 4). Patients in the low symptom burden subgroup (N=63) were split between our BALS (47 [75%] of the 63) and HSA (16 [25%]) clusters, with

none in the LSAHS cluster. Patients in the high symptom burden, pain dominant with fatigue, and dryness dominant with fatigue subgroups were present in all three of our clusters.

Discussion

We used an unsupervised clustering method, based on symptoms and clinical and biological manifestations, to distinguish three homogeneous subgroups of Sjögren's disease—BALS, HSA, and LSAHS—in two independent cohorts of patients. These clusters had different prognoses in terms of the risk of systemic and symptomatic evolution and of incident lymphoma. Notably, lymphoma occurred later in the BALS cluster than in the HSA cluster. We found a poor correlation between this new classification and a previous symptom-based classification.⁷

Our results have similarities to the subgroups described in previous studies. Beydon and colleagues¹⁸ reported that patients of African ancestry had higher disease activity with more systemic complications, as measured by the cumulative ESSDAI score (7.5 vs 4.0, $p=0.002$), than White patients. Such patients are highly represented in our HSA cluster. Two studies found that patients with anti-centromere²⁰ or anti-RNP21 antibodies had higher systemic disease activity than those without—consistent with the grouping of patients with these antibodies exclusively in our HSA cluster. However, these studies did not use unsupervised clustering methods, instead using supervised analyses that focused on the comparison of manifestations according to a single characteristic or factor.

Over the past 5 years, attempts have been made to stratify patients with Sjögren's disease using unsupervised clustering; however, the stratification methods used were based only on subjective symptoms (appendix pp 6–7).^{7,8,22–24} In an international, cross-sectional study, Gairy and colleagues²² used latent class analysis to identify five distinct clusters: two were characterised by a low burden, with or without articular involvement; one by a moderate burden consisting of fatigue, pain, and joint discomfort; and two clusters showed multi-organ involvement, with differing levels of burden (one moderate and one high). A symptom-based cluster analysis using five common symptoms (pain, fatigue, dryness, anxiety, and depression), conducted on data from the UKPSSR,⁷ identified four subgroups and was confirmed in two independent validation cohorts: one from Norway and the ASSESS cohort in France.⁷ Notably, reanalysis of data from the JOQUER trial after stratification of patients into these four subgroups suggested a treatment effect with hydroxychloroquine in the high symptom burden subgroup and with rituximab in the dryness dominant with fatigue subgroup, compared with placebo.^{7,25} In addition, clear differences in transcriptomic and serum proteomic profiles,⁷ as well as differences in health-related quality-of-life outcome measures, were observed between the four subgroups.²⁶ However, these stratifications were also based only on symptoms and did not include all aspects of Sjögren's disease. We found that the correlation between the subgroups from the UKPSSR and our obtained clusters was poor, as patients from the high symptom burden, dryness dominant with fatigue, and pain dominant with fatigue subgroups were present in each of our three clusters. This finding is consistent with the poor correlation between patient-reported outcomes and systemic disease activity.⁹ These two stratification approaches could be complementary, and both should be considered in clinical trials and daily practice depending on which question is to be addressed. Symptom-based classifications might be a useful tool to identify which patients are more likely to consult health services or have an impaired quality of life, whereas the stratification we

present here could identify patients with high systemic disease activity (HSA cluster) and those with the highest risks of incident lymphoma (BALS and HSA clusters).

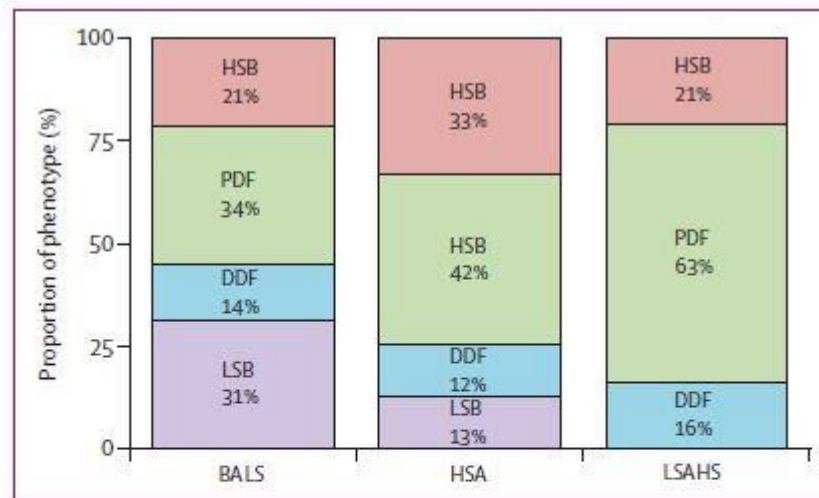


Figure 4: Proportion of patients in each symptom-based subgroup, assessed by the NSST, in each cluster in the ASSESS cohort

Data were available for 316 (80%) of the 395 patients in the ASSESS cohort. Symptom-based subgroups were described by Tarn and colleagues.⁷ BALS=B-cell active disease and low symptom burden. DDF=dryness dominant with fatigue. HSA=high systemic disease activity. HSB=high symptom burden. LSAHS=low systemic disease activity and high symptom burden. LSB=low symptom burden. NSST=Newcastle Sjögren's Stratification Tool. PDF=pain dominant with fatigue.

To our knowledge, our study is the first to identify clusters based on patient-reported outcomes, objective measures, and biological parameters. Our findings also highlight the fact that, even in patients who present with predominantly systemic manifestations, the symptom burden is high and should not be neglected. Overall, these results reinforce the need to adequately evaluate patient-reported outcomes in all subgroups of Sjögren's disease, regardless of systemic activity, as proposed in the Sjögren's Tool for Assessing Response (STAR) 27 or the Composite of Relevant Endpoints for Sjögren's Syndrome (CRESS).²⁸

In addition to describing different subgroups of patients with Sjögren's disease, we also assessed their long-term outcomes in terms of systemic activity, symptom evolution, and risk of lymphoma. Compared with those in other clusters, patients in the BALS cluster were more likely to reach a moderate to high disease activity state and to experience worsening of their subjective symptoms over time. Of note, follow-up ESSDAI and ESSPRI scores were available for only 5 years, and long-term studies would be needed to establish whether more patients develop widespread systemic disease. This follow-up is ongoing within the ASSESS cohort. However, we had 15-year follow-up data on the incidence of lymphoma, which revealed that incident lymphoma occurred exclusively in the BALS and HSA clusters. Notably, in the BALS cluster, lymphoma occurred later than in the HSA cluster and, after 5 years, systemic manifestations in this cluster tended to be similar to those in the HSA cluster at inclusion. The BALS cluster could therefore represent an earlier stage of the disease and carry the risk of progressing towards a more systemic phenotype (HSA).

We acknowledge the limitations of our study. Some data in the ASSESS cohort were missing, requiring the use of multiple imputations. However, the rate of missing was low (<10%). In addition, even though we found similar clusters in the two cohorts, the respective proportions of patients in each cluster were different. These differences can be explained by the recruitment strategies of each cohort: as the primary goal of the ASSESS cohort was to establish the risk of systemic changes and the risk factors for lymphomas, patients with low systemic disease activity were less likely than those with higher activity to be included in this cohort. Additionally, we did not account for current and previous treatments for Sjögren's disease in the analysis, which could have influenced particular manifestations. Nevertheless, only a few patients were treated with disease-modifying antirheumatic drugs or immunosuppressants. Furthermore, ESSPRI and ESSDAI scores for individual patients might vary over time, and an annual assessment of these indexes might not give the full picture of symptoms and disease progression. Finally, Cox models are subject to built-in selection bias when calculating hazard ratios.

In conclusion, using an unsupervised clustering method encompassing all features of Sjögren's disease, we identified three subgroups with distinct disease evolution and prognosis. Our study suggests the role of heterogeneous pathophysiological mechanisms within Sjögren's disease, and could aid further stratifications of patients with this disease.

Contributors

YN, XM, and RS conceived and designed the study. YN, GN, FD, JH, RB, EB, J-EG, XM, and RS collected and analysed the data. All authors interpreted the data. YN wrote the first version of the manuscript, and all authors critically revised and approved the final version. YN and RS directly accessed and verified the underlying data. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Declaration of interests

GN received honoraria from Boehringer and Novartis and travel fees from Amgen and AbbVie. W-FN received consulting fees from Resolve Therapeutics, Argenx, and Novartis and participated on data safety monitoring boards or advisory boards for Sanofi and Janssen Pharmaceuticals. JM received grants from Bristol Myers Squibb (BMS), Fresenius Kabi, Lilly, Novartis, Pfizer, and Roche-Chugai; received honoraria from AbbVie, Boehringer Ingelheim, Biogen, Lilly, Mylan, Pfizer, Sanofi, BMS, Fresenius Kabi, Galapagos, Medac, Novartis, and Roche-Chugai; received travel fees from BMS, Lilly, and Fresenius Kabi; and participated on advisory boards for AbbVie, Pfizer, and Galapagos. ED received consulting fees from BMS, Celgene, Lilly, Merck Sharp & Dohme (MSD), Novartis, and UCB; honoraria for lectures from AbbVie, BMS, Janssen Pharmaceuticals, Lilly, Medac, MSD, Novartis, Roche-Chugai, Sanofi, UCB, Celgene, Amgen, and Galapagos; and travel fees from AbbVie, BMS, Janssen Pharmaceuticals, Lilly, Medac, MSD, Novartis, Roche-Chugai, Sanofi, UCB, Celgene, Amgen, and Galapagos. PD received grants from Novartis and consulting fees from Pfizer, Roche-Chugai, BMS, AbbVie, and MSD. MC received travel fees from Janssen Pharmaceuticals, UCB, and Pfizer. CS received honoraria from Novartis and Roche-Chugai and travel fees from Novartis, Biogen, and Lilly. VLG received travel fees from AstraZeneca and Novartis. J-EG received grants from Pfizer, AbbVie, and Lilly and

consulting fees from AbbVie, AstraZeneca, Sanofi, Lilly, Galapagos, Gilead Sciences, Roche-Chugai, Pfizer, BMS, and MSD. XM received consulting fees from AstraZeneca, BMS, Galapagos, GSK, Novartis, and Pfizer. RS received consulting fees from GSK, BMS, Boehringer Ingelheim, and Janssen Pharmaceuticals; honoraria from GSK, BMS, Boehringer Ingelheim, Amgen, Pfizer, and Roche-Chugai; and travel fees from Amgen and GSK. All other authors declare no competing interests.

Data sharing

Anonymised data from both cohorts, along with the data dictionary, will be available upon publication. All requests for access must be directed to XM (xavier.mariette@aphp.fr). Data transfer will be subject to an access protocol and will require authorisation from the scientific board of the Paris-Saclay or the ASSESS cohort. Data sharing will be contingent upon a data sharing agreement prior to any transfer, to ensure all users of the data adhere to the legal requirements of using personal data.

Acknowledgments

The authors thank all patients for their participation and all physicians who included patients in the Paris-Saclay and ASSESS cohorts. YN is supported by a grant from the Fondation pour la Recherche Médicale (SPF202209015782). The Assessment of Systemic Signs and Evolution in Sjögren's Syndrome (ASSESS) national multicentre prospective cohort was formed in 2006 with a French Ministry of Health grant (Programme Hospitalier de Recherche Clinique 2005 P060228). The ASSESS cohort is promoted by the French Society of Rheumatology and receives research grants from the French Society of Rheumatology. The NECESSITY project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (IMI 2 JU; NECESSITY grant 806975). RS, XM, GN, W-FN, VD-P, and J-EG are members of the NECESSITY consortium. The Joint Undertaking received support from the European Union's Horizon 2020 Research and Innovation Program and from the European Federation of Pharmaceutical Industries and Associations. This Article reflects only the authors' views and the Joint Undertaking is not responsible for any use that might be made of the information it contains. The NSST development was supported by funding from the Medical Research Council UK (G0800629) and the Foundation for Research in Rheumatology (FOREUM).

Conflict of interest statement

Conflicts of interest GN received honoraria from Boehringer and Novartis and travel fees from Amgen and AbbVie. W-FN received consulting fees from Resolve Therapeutics, Argenx, and Novartis and participated on data safety monitoring boards or advisory boards for Sanofi and Janssen Pharmaceuticals. JM received grants from Bristol Myers Squibb (BMS), Fresenius Kabi, Lilly, Novartis, Pfizer, and Roche-Chugai; received honoraria from AbbVie, Boehringer Ingelheim, Biogen, Lilly, Mylan, Pfizer, Sanofi, BMS, Fresenius Kabi, Galapagos, Medac, Novartis, and Roche-Chugai; received travel fees from BMS, Lilly, and Fresenius Kabi; and participated on advisory boards for AbbVie, Pfizer, and Galapagos. ED received consulting fees from BMS, Celgene, Lilly, Merck Sharp & Dohme (MSD), Novartis, and UCB; honoraria for lectures from AbbVie, BMS, Janssen Pharmaceuticals, Lilly, Medac, MSD, Novartis, Roche-Chugai, Sanofi, UCB, Celgene, Amgen, and Galapagos; and travel fees from AbbVie, BMS, Janssen Pharmaceuticals, Lilly, Medac, MSD, Novartis, Roche-Chugai, Sanofi, UCB, Celgene, Amgen, and Galapagos. PD received grants from Novartis and consulting fees from Pfizer, Roche-Chugai, BMS, AbbVie, and MSD. MC received travel fees from Janssen Pharmaceuticals, UCB, and Pfizer. CS received honoraria from Novartis

and Roche-Chugai and travel fees from Novartis, Biogen, and Lilly. VLG received travel fees from AstraZeneca and Novartis. J-EG received grants from Pfizer, AbbVie, and Lilly and consulting fees from AbbVie, AstraZeneca, Sanofi, Lilly, Galapagos, Gilead Sciences, Roche-Chugai, Pfizer, BMS, and MSD. XM received consulting fees from AstraZeneca, BMS, Galapagos, GSK, Novartis, and Pfizer. RS received consulting fees from GSK, BMS, Boehringer Ingelheim, and Janssen Pharmaceuticals; honoraria from GSK, BMS, Boehringer Ingelheim, Amgen, Pfizer, and Roche-Chugai; and travel fees from Amgen and GSK. All other authors declare no competing interests.

References

- 1 Mariette X, Criswell LA. Primary Sjögren's syndrome. *N Engl J Med* 2018; 378: 931–39.
- 2 Beydon M, McCoy S, Nguyen Y, Sumida T, Mariette X, Seror R. Epidemiology of Sjögren syndrome. *Nat Rev Rheumatol* 2023; published online Dec 18. <https://doi.org/10.1038/s41584-023-01057-6>.
- 3 Mahr A, Katsahian S, Varet H, et al. Revisiting the classification of clinical phenotypes of anti-neutrophil cytoplasmic antibody-associated vasculitis: a cluster analysis. *Ann Rheum Dis* 2013; 72: 1003–10.
- 4 Dion J, Costedoat-Chalumeau N, Sène D, et al. Relapsing polychondritis can be characterized by three different clinical phenotypes: analysis of a recent series of 142 patients. *Arthritis Rheumatol* 2016; 68: 2992–3001.
- 5 Font J, Cervera R, Ramos-Casals M, et al. Clusters of clinical and immunologic features in systemic lupus erythematosus: analysis of 600 patients from a single center. *Semin Arthritis Rheum* 2004; 33: 217–30.
- 6 Nguyen Y, Yelnik CM, Morel N, et al. Determination of four homogeneous subgroups of patients with antiphospholipid syndrome: a cluster analysis based on 509 cases. *Rheumatology (Oxford)* 2023; 62: 2813–19.
- 7 Tarn JR, Howard-Tripp N, Lendrem DW, et al. Symptom-based stratification of patients with primary Sjögren's syndrome: multi-dimensional characterisation of international observational cohorts and reanalyses of randomised clinical trials. *Lancet Rheumatol* 2019; 1: e85–94.
- 8 McCoy SS, Woodham M, Bartels CM, et al. Symptom-based cluster analysis categorizes Sjögren's disease subtypes: an international cohort study highlighting disease severity and treatment discordance. *Arthritis Rheumatol* 2022; 74: 1569–79.
- 9 Seror R, Gottenberg JE, Devauchelle-Pensec V, et al. European League Against Rheumatism Sjögren's Syndrome Disease Activity Index and European League Against Rheumatism Sjögren's Syndrome Patient- Reported Index: a complete picture of primary Sjögren's syndrome patients. *Arthritis Care Res (Hoboken)* 2013; 65: 1358–64.

- 10 Hay EM, Thomas E, Pal B, Hajeer A, Chambers H, Silman AJ. Weak association between subjective symptoms or and objective testing for dry eyes and dry mouth: results from a population based study. *Ann Rheum Dis* 1998; **57**: 20–24.
- 11 Duret P-M, Schleiss C, Kawka L, et al. Association between Bruton's tyrosine kinase gene overexpression and risk of lymphoma in primary Sjögren's syndrome. *Arthritis Rheumatol* 2023; **75**: 1798–811.
- 12 Nocturne G, Seror R, Fogel O, et al. CXCL13 and CCL11 serum levels and lymphoma and disease activity in primary Sjögren's syndrome. *Arthritis Rheumatol* 2015; **67**: 3226–33.
- 13 Tobón GJ, Saraux A, Gottenberg J-E, et al. Role of Fms-like tyrosine kinase 3 ligand as a potential biologic marker of lymphoma in primary Sjögren's syndrome. *Arthritis Rheum* 2013; **65**: 3218–27.
- 14 Vitali C, Bombardieri S, Jonsson R, et al. Classification criteria for Sjögren's syndrome: a revised version of the European criteria proposed by the American–European Consensus Group. *Ann Rheum Dis* 2002; **61**: 554–58.
- 15 Gottenberg J-E, Seror R, Miceli-Richard C, et al. Serum levels of beta2-microglobulin and free light chains of immunoglobulins are associated with systemic disease activity in primary Sjögren's syndrome. Data at enrollment in the prospective ASSESS cohort. *PLoS One* 2013; **8**: e59868.
- 16 Seror R, Ravaud P, Bowman SJ, et al. EULAR Sjogren's syndrome disease activity index: development of a consensus systemic disease activity index for primary Sjogren's syndrome. *Ann Rheum Dis* 2010; **69**: 1103–09.
- 17 Seror R, Ravaud P, Mariette X, et al. EULAR Sjogren's Syndrome Patient Reported Index (ESSPRI): development of a consensus patient index for primary Sjogren's syndrome. *Ann Rheum Dis* 2011; **70**: 968–72.
- 18 Beydon M, Seror R, Le Guern V, Chretien P, Mariette X, Nocturne G. Impact of patient ancestry on heterogeneity of Sjögren's disease. *RMD Open* 2023; **9**: e002955.
- 19 Seror R, Bootsma H, Saraux A, et al. Defining disease activity states and clinically meaningful improvement in primary Sjögren's syndrome with EULAR primary Sjögren's syndrome disease activity (ESSDAI) and patient-reported indexes (ESSPRI). *Ann Rheum Dis* 2016; **75**: 382–89.
- 20 Liu Y, Guo L, Lin W, et al. Anticentromere antibody positive patients with primary Sjögren's syndrome have distinctive clinical and immunological characteristics. *Clin Exp Rheumatol* 2023; **41**: 2371–78.
- 21 Abbara S, Seror R, Henry J, et al. Anti-RNP positivity in primary Sjögren's syndrome is associated with a more active disease and a more frequent muscular and pulmonary involvement. *RMD Open* 2019; **5**: e001033.
- 22 Gairy K, Knight C, Anthony P, Hoskin B. Burden of illness among subgroups of patients with primary Sjögren's syndrome and systemic involvement. *Rheumatology (Oxford)* 2021; **60**: 1871–81.
- 23 Lee JJ, Park YJ, Park M, Yim HW, Park SH, Kwok S-K. Longitudinal analysis of symptom-based clustering in patients with primary Sjogren's syndrome: a prospective cohort study with a 5-year follow-up period. *J Transl Med* 2021; **19**: 394.
- 24 Kramer E, Seeliger T, Skripuletz T, et al. Multimodal assessment and characterization of Sicca syndrome. *Front Med (Lausanne)* 2021; **8**: 777599.
- 25 Collins A, Lendrem D, Wason J, et al. Revisiting the JOQUER trial: stratification of primary Sjögren's syndrome and the clinical and interferon response to hydroxychloroquine. *Rheumatol Int* 2021; **41**: 1593–600.

- 26 Tarn J, Lendrem D, McMeekin P, Lendrem C, Hargreaves B, Ng W-F. Primary Sjögren's syndrome: longitudinal real-world, observational data on health-related quality of life. *J Intern Med* 2022; 291: 849–55.
- 27 Seror R, Baron G, Camus M, et al. Development and preliminary validation of the Sjögren's Tool for Assessing Response (STAR): a consensual composite score for assessing treatment effect in primary Sjögren's syndrome. *Ann Rheum Dis* 2022; 81: 979–89.
- 28 Arends S, de Wolff L, van Nimwegen JF, et al. Composite of Relevant Endpoints for Sjögren's Syndrome (CRESS): development and validation of a novel outcome measure. *Lancet Rheumatol* 2021; 3: e553–62.