EUROPEAN RESPIRATORY *journal*

FLAGSHIP SCIENTIFIC JOURNAL OF ERS

Early View

Original article

Clinical phenotypes of extrapulmonary sarcoidosis: an analysis of a French, multiethnic, multicenter cohort

Raphael Lhote, Isabella Annesi-Maesano, Hilario Nunes, David Launay, Raphael Borie, Karim Sacré, Nicolas Schleinitz, Mohamed Hamidou, Matthieu Mahevas, Hervé Devilliers, Philippe Bonniaud, François Lhote, Julien Haroche, Pierre Rufat, Zahir Amoura, Dominique Valeyre, Fleur Cohen Aubart

Please cite this article as: Lhote R, Annesi-Maesano I, Nunes H, *et al*. Clinical phenotypes of extrapulmonary sarcoidosis: an analysis of a French, multiethnic, multicenter cohort. *Eur Respir J* 2020; in press (https://doi.org/10.1183/13993003.01160-2020).

Clinical phenotypes of extrapulmonary sarcoidosis: an analysis of a French, multiethnic, multicenter cohort

Running Title: Sarcoidosis phenotypes in a multicenter study

Raphael Lhote[1,2], MD, Isabella Annesi-Maesano[1], MD, PhD, DSc, Hilario Nunes[3], MD, PhD, David Launay[4], MD, PhD, Raphael Borie[5], MD, PhD, Karim Sacré[6], MD, PhD, Nicolas Schleinitz[7], MD, PhD, Mohamed Hamidou[8], MD, PhD, Matthieu Mahevas[9], MD, PhD, Hervé Devilliers[10], MD, PhD, Philippe Bonniaud[11], MD, François Lhote[12], MD, Julien Haroche[1], MD, PhD, Pierre Rufat[13], MD, Zahir Amoura[1], MD, MSc, Dominique Valeyre[3], MD, and Fleur Cohen Aubart[1], MD, PhD

Affiliations

[1] Sorbonne Université, Institut Pierre Louis d'Epidémiologie et de Santé Publique (IPLESP UMRS 1136), Epidémiologie des maladies Allergiques et Respiratoires (EPAR), Faculté de Médecine Saint-Antoine, 75012-Paris, France

[2] Sorbonne Université, Assistance Publique Hôpitaux de Paris, Hôpital de la Pitié-Salpêtrière, Service de Médecine Interne 2, Centre National de Référence Maladies Systémiques Rares, Lupus et Syndrome des Anticorps Antiphospholipides, Centre National de Référence Histiocytoses, 75013-Paris, France

[3] Assistance Publique Hôpitaux de Paris, Hôpital Avicenne, Service de Pneumologie, 93000-Bobigny, France

[4] Université de Lille, INSERM U995 – LIRIC – Lille Inflammation Research International Center, CHRU de Lille, Département de Médecine Interne et Immunologie Clinique, Centre de Référence des Maladies Auto-immunes Systémique Rares du Nord et du Nord-Ouest de France (CeRAINO), 59000-Lille, France

[5] Université Paris Diderot, Assistance Publique Hôpitaux de Paris, Hôpital Bichat, Service de Pneumologie A, et INSERM U1152, Paris-75018, France

[6] Université Paris Diderot, Assistance Publique Hôpitaux de Paris, Hôpital Bichat, Service de Médecine Interne, Paris-75018, France

[7] Université Aix Marseille, Assistance Publique Hôpitaux de Marseille, Hôpital de la Timone, Service de Médecine Interne, Marseille-13000, France

[8] Université de Nantes, Centre Hospitalier Universitaire, Service de Médecine Interne, Nantes-44093, France

[11] Assistance Publique Hôpitaux de Paris, Hôpital Mondor, Service de Médecine Interne, Créteil-94010, France

[10] Centre Hospitalier Dijon-Bourgogne, Service de Médecine Interne 2, Dijon-21000, France

[11] Centre Hospitalier Dijon-Bourgogne, Service de Pneumologie et Soins Intensifs Respiratoires, Centre de référence constitutif des maladies pulmonaires rares de l'adulte, Dijon-21000, France

[12] Hôpital Delafontaine, Service de Médecine Interne, St-Denis-93200, France

[13] Sorbonne Université, Assistance Publique Hôpitaux de Paris, Hôpital de la Pitié-Salpêtrière, Département d'Information Médicale et Biostatistiques, Paris-75013, France

Correspondence to

Dr. Fleur Cohen Aubart, Service de Médecine Interne 2, Groupe Hospitalier Pitié-Salpêtrière, 47-83 Boulevard de l'hôpital, 75 651 PARIS CEDEX 13, France, email fleur.cohen@aphp.fr

Phone + 33 (0)1 42 17 82 42

Fax + 33 (0)1 42 16 58 04

Word count in the text: 2970

Word count in abstract: 242

Number of Figures: 3

Number of Tables: 3

Number of references: 21

Keywords: sarcoidosis, phenotypes, clustering analysis, granuloma

Contributions

I.A.-M., H.N., D.V., Z.A., and F.C.A. designed the study.

R.L., R.B., K.S., N.S., D.L., H.D., P.B., M.H., M.M., F.L., J.H., and F.C.-A. collected the data.

R.L., I.A.-M., and F.C.-A. conducted the statistical analysis.

R.L., I.A.-M., H.N., D.V., and F.C.-A. analyzed and interpreted the data.

R.L., I.A.-M., and F.C.-A. wrote the manuscript.

All authors critically reviewed and approved the final version of the manuscript.

R.L. was supported by a Master and PhD grants of the Open Health Institute, the Fondation pour la Recherche Médicale (FRM), and Gérard Trouillet S.A.

The authors declare that they have no conflicts of interest to report.

Abstract

Background Sarcoidosis is a rare disease of unknown cause with wide heterogeneity in clinical features and outcomes. We aimed to explore sarcoidosis phenotypes and their clinical relevance with particular attention to extrapulmonary subgroups.

Patients and methods The EpiSarc (Epidemiology of Sarcoidosis) study is a French retrospective multicenter study. Sarcoidosis patients were identified through national hospitalization records using appropriate codes from 11 hospital centers between 2013 and 2016 according to a standardized protocol. Medical charts were reviewed. The phenotypes of sarcoidosis were defined using a hierarchical cluster analysis.

Results A total of 1,237 patients were included (562 men and 675 women). The mean age at sarcoidosis diagnosis was 43.5±13 years. Hierarchical cluster analysis identified five distinct phenotypes according to organ involvement and disease type and symptoms: 1) (n=180) *erythema nodosum*, joint involvement and hilar lymph nodes; 2) (n= 137) eye, neurological, digestive and kidney involvement; 3) (n=630) pulmonary involvement with fibrosis and heart involvement; 4) (n=41) *lupus pernio* and a high percentage of severe involvement; and 5) (n=249) hepatosplenic, peripheral lymph node and bone involvement. Phenotype 1 was associated with being European and female and with nonmanual work; phenotype 2 with being European; and phenotypes 3 and 5 with being non-European. The labor worker proportion was significantly lower in phenotype 5 than in the other phenotypes.

Answer to the question This multicenter study confirms the existence of distinct phenotypes of sarcoidosis, with a nonrandom distribution of organ involvement. These phenotypes differ according to gender, geographical origin and socioprofessional categories.

Sarcoidosis is a rare heterogeneous multisystemic granulomatous disease of unknown cause. It preferentially affects the lungs and intrathoracic lymph nodes. Extrapulmonary involvement is observed in 30% to 50% of patients and occurs more frequently in skin, eye, lymphatic organ, liver and spleen involvements [1]. Sarcoidosis has a variable clinical presentation and heterogeneous outcomes, hitherto unexplained.

The first clinical phenotype of sarcoidosis was described by the Swedish pulmonologist Sven Löfgren in 1946 [2]. The so-called "Löfgren syndrome" is mostly seen in young European patients with acute-onset *erythema nodosum,* bilateral hilar lymphadenopathy, fever, or migratory polyarthritis and typically has a favorable outcome. Later, several studies described the heterogeneity of sarcoidosis according to ethnicity, sex, and age [3-7] and identified phenotypes, but others have still to be identified. The determination of sarcoidosis phenotypes was recently supported by the GenPhenReSa study through a clustering approach of Caucasian patients in an international panel of centers [8]. This approach allowed the identification of 5 subgroups of patients characterized by pulmonary and extrapulmonary organ involvement. Subgroups of extrapulmonary involvement have also been reported in a significant American cohort, in which skin lesions were the most important feature [9]. More recently, ^{18}F-fluorodeoxyglucose positrons emission tomography, coupled with cluster analysis, succeeded to identify in a small sample of patients an ordered stratification into 4 phenotypes [10]. However, other studies are necessary to confirm these findings and allow a better understanding of sarcoidosis, phenotypes expression, and treatment guidance [11].

The aim of the EpiSarc (Epidemiology of Sarcoidosis) study was to identify and explore sarcoidosis phenotypes using an unsupervised analytical approach in a well-characterized cohort of French sarcoidosis patients with extrathoracic involvements.

Patients and methods

The EpiSarc study was a multicenter (11 French centers) retrospective transversal study. Patients were identified using the French hospitalizations database (*Programme de Médicalisation des Systèmes d'Information*), which records the medical codes of hospitalizations

(according to the International Classification of Diseases, **ICD**-10). Selected patients were hospitalized between January 2013 and December 2016 with a medical code corresponding to sarcoidosis (D86, G532, and/or M633), as newly diagnosed or chronic disease cases. Medical charts were manually reviewed for the verification of inclusion criteria. The inclusion criteria were as follows: 1) a sarcoidosis diagnosis according to the American Thoracic Society and World Association for Sarcoidosis and Other Granulomatous Diseases (**WASOG**) recommendations, including compatible clinico-radiological findings, histological evidence of noncaseating granuloma, and the exclusion of other granulomatous diseases [12]; and 2) the involvement of at least one extrapulmonary organ. Patients with Löfgren syndrome, defined by the association of fever, *erythema nodosum*, arthralgia, and bilateral hilar lymphadenopathy with or without histological evidence, were also included. Demographic, clinical, biological, and imaging data were extracted from medical records. Organ involvements were recorded according to the WASOG organ assessment instrument and were classified as present if they were observed at any time in the sarcoidosis history [13]. We studied organ involvement listed in the extrapulmonary Physician Organ Severity Tool (ePOST) score and pulmonary involvement [14]. In addition, the analysis also considered the sociodemographic characteristics of the patients (age, self-reported geographic origin, and occupation at the time of diagnosis, classified according to the *Institut National de la Statistique et des Etudes Economiques (INSEE)* tool (https://www.insee.fr/fr/information/2497958) and to the International Standard Classification of Occupation (ISCO) 2008), comorbidities (thromboembolism disease, neoplasm, autoimmune disease, chronic infectious disease), treatments, and outcomes. Occupations were categorized into 6 groups according to INSEE classification and ISCO 2008: craft and trade-related workers (ISCO 2008 group 7); labor workers (ISCO 2008 groups 8 and 9); farmers and skilled agricultural workers (ISCO 2008 group 6); clerk support workers (ISCO 2008 group 4); intermediary occupations (ISCO 2008 groups 0, 3, and 5); and upper class, framework and higher intellectual professions (ISCO 2008 groups 1 and 2).

Statistical analysis

Variables were presented as percentages (categorical variables) or mean ± standard deviation (**SD**) (continuous variables) as appropriate. The phenotypes of sarcoidosis were defined using a hierarchical cluster analysis. For hierarchic ascendant clustering, preferential variable associations were obtained after the inclusion of the variables of organ involvement included in the ePOST score, in addition to the following, referring to intra-thoracic involvement: mediastinal lymph nodes, parenchymal involvement, lung fibrosis, and represented in a dendrogram plot ("pvclust packages" in R 3.3) generated by Ward's method. The principal component analysis **(PCA)** was applied to select the final clinical variables of the model that were included in the model used to identify phenotypes according to statistical significance. The population was then separated into clusters ("factoMineR packages" in R 3.3) [15]. Unsupervised hierarchical cluster analysis was performed using the "Euclidian" method to create the distance matrix and Ward's minimum variance method. Differences in sarcoidosis phenotypes were compared using the chi-squared test or Fisher's exact test when necessary (categorical variables) and Student's t-test, ANOVA, the Wilcoxon test, or the Kruskal-Wallis test (continuous variables) when appropriate. In addition, multinomial logistic regression was used to confirm the relationships of the different clusters to the set of independent variables. The multinomial logistic regression models were adjusted for gender, geographical origin, and socioprofessional activities. All tests were bilateral, with type I errors of 5% and 95% confidence intervals. The analyses were performed in R software, version 3.3. This study was approved according to the French legislation by the *Commission Nationale Informatique et Libertés* (CNIL) organization.

Results

Demographic and clinical characteristics

During the study period, 2,090 patients were screened. Among them, 853 patients were excluded (Figure 1) because of the absence of extrapulmonary organ involvement (n=275, 13%), the absence of histological documentation (n=321, 15%), a lack of data in medical records (n=95, 5%), redundant follow-up in 2 participating centers (n=7, 0.05%) and the

absence of sarcoidosis diagnosis (n=155, 7%). The final study included 1,237 patients, 562 (44.5%) males and 675 (54.5%) females. The mean age at sarcoidosis diagnosis was 43.5 (±13) years. Most of the patients were European (n=541, 44%), followed by Afro-Caribbean (n=314, 25.5%). The detailed clinical characteristics of the population are described in **Table 1**.
The clinical characteristics of the patients are described in the first column of the **Table 2**. Organ involvement mainly consisted of hilar lymphadenopathy (n=1115, n=90%), parenchymal lung involvement (n=792, 64%), pulmonary fibrosis (n=176, 14%), peripheral lymphadenopathy (n=518, 42%), skin involvement (n=396, 32%) and joint involvement (n=350, 28%). The mean number of organs involved was 3.4 (1.5), and 520 patients (42%) had >3 organs involved.

Sarcoidosis phenotypes

Five groups of patients were identified, corresponding to 5 sarcoidosis phenotypes, using hierarchical cluster analysis. The clinical characteristics of the population according to these 5 phenotypes are described in **Table 2**. The 5 phenotypes were as follows: 1) (n=180): a higher frequency of hilar lymph nodes (99%) and joint involvement (90%), a lower frequency of parenchymal lung involvement (39%), and the presence of erythema nodosum (67%); 2) (n=137) a higher frequency of neurological (31%), digestive (24%), and kidney (12%) involvement and a lower frequency of mediastino-pulmonary involvement (33%); 3) (n=630) a higher frequency of parenchymal lung involvement (72%), including pulmonary fibrosis (19%), cardiac involvement (19%), and cutaneous sarcoids (37%); 4) (n=41): an aggregation of all patients with *lupus pernio* (95%), and a high frequency of ear, nose and throat (**ENT)** (68%) and parenchymal lung involvement (88), including pulmonary fibrosis (29%); and 5 (n=249): a high frequency of parenchymal lung involvement (82%), peripheral nodes (68%), and hepatic (68%), splenic (68%), and bone (29%) involvement. Finally, parenchymal lung involvement, including pulmonary fibrosis, was present in clusters 3, 4, and 5. Cluster 3 associated neurological and abdominal involvement, cluster 4 aggregated cases of *lupus pernio* and ENT involvement, and cluster 5 comprised peripheral lymph nodes and hepatosplenic and bone

involvement, in addition to parenchymal localization. There was no difference for age at diagnosis regarding phenotype status. To confirm reproducibility and the stability of the results, the cluster analysis was made twice, once after inclusion of the patients from 7 centers (corresponding to 1,081 patients), then after the inclusion of the 11 centers with 1237 patients, leading to a concordance of 85% of the results, with the same clustering.

In addition to cluster separation, nonsupervised preferential associations of organ involvement are shown in **Figure 2**. *Erythema nodosum* was associated with joint involvement. Peripheral nodes were associated with bone and hepatosplenic involvement. Cardiac localizations were associated with parenchymal lung involvement and fibrosis. *Lupus pernio* was associated with ENT manifestations, uveitis, and neurological involvement. These associations strengthened the results of the cluster distributions.

Treatments and outcomes

Treatments and outcomes according to the 5 phenotypes are described in **Table 2.** The mean duration of follow-up was 8.1 (±8) years, and was longer in cluster 4. The most common treatment regimens were corticosteroids (n=907, 73%), methotrexate (n=388, 31%), hydroxychloroquine (n=283, 23%), azathioprine (n=161, 13%), TNF alpha antagonists (n=113, 9%) and cyclophosphamide (n=61, 5%). During follow-up, 23 patients died (2%).

Treatment regimens *(i.e.,* corticosteroids, methotrexate, azathioprine, TNF antagonists and/or cyclophosphamide) were significantly different among the 5 phenotypes.

Phenotype 1 was treated less than other phenotypes with corticosteroids (n=95, 53% versus 77%, p<0.001), with methotrexate (n=34, 19% versus 34%, p<0.001), with azathioprine (n=13, 7% versus 14%, p=0.02), with TNF antagonists (n= 7, 4% versus 10%, p=0.01) and with cyclophosphamide (n=3, 2% versus 5%, p=0.046). Phenotype 2 was treated more than other phenotypes with cyclophosphamide (n=15, 11% versus 9%, p=0.02) and treated less with azathioprine (n=10, 7% versus 14%, p=0.046). Phenotype 3 was treated more than other phenotypes with corticosteroids (n=493, 78% versus 68%, p<0.001). Phenotype 4 was treated more than other phenotypes with corticosteroids (n=36, 88% versus 73%, p=0.05),

methotrexate (n= 31, 76% versus 30%, p<0.001), azathioprine (n=14, 34% versus 12%, p<0.001) and TNF antagonists (n= 17, 41% versus 8%, p<0.001). Phenotype 5 was treated less than the other clusters with TNF antagonists (n=12, 5% versus 10%, p=0.01). The number of treatments was also different regarding the phenotypes (**Figure 3**). Patients received a lower number of treatments in cluster 1 (mean number of treatments of 1.1±1.3) than in cluster 2 (mean number of treatments of 1.6±1.3), cluster 3 (mean number of treatments of 1.7±1.2), and cluster 5 (mean number of treatments of 1.6±1.2). The number of treatments was higher among patients in cluster 4 (3.6±1.7) (p<0.001 for the comparison of the number of treatments among the 5 clusters). The number of deaths was also significantly different in the clusters, with no death in cluster 1; 0.5% to 3% in clusters 2, 3 and 5; and 5% in cluster 4 (p=0.01).

Socio-economic factors related to sarcoidosis phenotypes

The phenotypes varied according to the socioeconomic characteristics (**Table 3**). Females were represented differently according to the phenotypes (p<0.002), and the female to male ratio was higher in phenotypes 1 (66% of women) and 4 (68% of women) than in phenotypes 2 (56% of women), 3 (51% of women) and 5 (53% of women). There was no difference in the age of diagnosis of sarcoidosis regarding phenotype status.

The occupational repartition differed significantly between the 5 phenotypes (p=0.04). Labor workers (40.5%) were the most frequent occupational population among the whole cohort. The proportion of labor workers was higher in cluster 3 (46%), cluster 2 (44%), and cluster 5 (35%).

The geographical origin repartition was significantly different regarding the phenotype status (p=0.01). Patients of European origin (44%) were the most represented geographic category in the study population. The proportion of Europeans was higher in phenotype 1 (51%) and phenotype 2 (62.5%) than in phenotypes 3 (42%), 5 (37.5%) or 4 (20%).

The adjusted-multinomial analysis confirmed the results of the univariate analysis. The female proportion was significantly higher in cluster 1 than in the other clusters (OR=0.6 95% CI [0.4-0.9], p=0.008); the labor worker proportion was significantly lower in cluster 1 (OR=0.6 95% CI

[0.4-0.7], p=0.03) and in cluster 5 (OR=0.6 95% CI [0.4-0.9], p=0.01) than in the other clusters. European origin was significantly higher in phenotype 1 (OR=1.7 95% CI [1.1-2.5], p=0.01) and phenotype 2 (OR=2.2 95% CI [1.4-3.4], p<0.001) than in the phenotypes.

Discussion

The EpiSarc study identified 5 phenotypes of sarcoidosis in a large French multicenter and multiethnic cohort of 1,237 well-characterized patients. These phenotypes may be useful for a better management of the disease [11]. The 5 phenotypes significantly differed by gender, geographic origin and occupational categories. The 5 phenotypes were 1) *erythema nodosum*, joint involvement and hilar lymph nodes, mainly European female; 2) neurological, digestive, and/or kidney involvement; 3) parenchymal lung involvement and fibrosis, cardiac and skin involvement, mainly in non-European patients; 4) *lupus pernio* and severe involvement; and 5) parenchymal pulmonary involvement, peripheral nodes, and hepatic, splenic, and bone involvement, mainly in non-European patients. In addition to clustering phenotypes, we identified a preferential association of organ involvements, characterized by the association of *erythema nodosum* with joint involvement; hepatic, splenic, bone and peripheral node involvement; pulmonary with cardiac involvements; kidney with digestive involvement; *lupus pernio* with ENT involvements; and uveitis with neurological involvements.

In the GenPhenReSa (Genotype-Phenotype Relationship in Sarcoidosis) study, which included 2,163 patients, the phenotypes were slightly distinct from those of the EpiSarc study. They included 1) abdominal organ involvement, 2) ocular-cardiac-cutaneous-central nervous system disease involvement, 3) musculoskeletal-cutaneous involvement, 4) pulmonary and intrathoracic lymph node involvement, and 5) extrapulmonary involvement. However, in GenPhenReSa study, only Caucasian patients were included. Organ involvement was different in the GenPhenReSa study: we observed more skin (32% versus 16%), eye (24% versus 8%), CNS (22% versus 3%), cardiac (10% versus 3%), and gastrointestinal involvement (3% versus 0.6%). Several explanations may be proposed: our cohort is a multiethnic cohort, and Afro-American patients usually have more organ involvement [7]. Moreover, our patients were

mostly enrolled from internal medicine centers, whereas the GenPhenReSa study participants were mostly enrolled from pulmonology centers. Our French centers are highly specialized centers, and we can suspect that we recruited more severe patients, particularly for cardiac, eye, or CNS involvement. We recruited only hospitalized patients. Finally, we excluded from our cohort patients with isolated intra-thoracic sarcoidosis, in order to highlight the extra-pulmonary localizations in the clustering analysis. Thus, we think that our results are complementary to those of the GenPhenReSa study. In another study of 1230 patients including 91.4% of white patients, those with pulmonary involvements had a lower frequency of skin and salivary glands involvements, and a higher frequency of liver involvement [16]. This study used multinomial logistic regression analyses to test the associations, whereas we used an unsupervised hierarchical clustering. Thus we better detected phenotypes than in this Spanish study, which studied associations. In another study including 195 patients from pulmonology centers, the authors used ^{18}F-fluorodeoxyglucose positrons emission tomography to identify 4 phenotypes based not only on an organ involvement basis, but on disease activity [10]. However, the phenotypes were not compared for severity and socio-economic factors. The impact of this study is also limited because not all patients with sarcoidosis need ^{18}F-fluorodeoxyglucose positrons emission tomography, which is an expensive exam, with radiations exposure.

In the EpiSarc study, the unsupervised hierarchical clustering allowed the aggregation of organ involvement corresponding to the "Löfgren" syndrome [17]. We also demonstrated the female and European predominance of these patients. Löfgren syndrome was first described in 1946 and was recognized as a distinct phenotype of sarcoidosis. Patients typically have an acute disease onset, together with bilateral hilar lymphadenopathy, *erythema nodosum*, and/or bilateral ankle arthritis or arthralgia. Patients with Löfgren syndrome typically have favorable outcomes, especially in HLA-DRB1*03+ individuals [18]. However, we found that Löfgren/phenotype 1 was less treated, with a mean number of treatments of 1.1±1.3 and lower rates for each molecule, and was associated with better outcomes with no deaths.

Extrapulmonary sarcoidosis specificities were recently reported in an American cohort [9]. In

this multicenter study, the skin was the most frequent extrathoracic organ involved in patients with or without lung involvement. Our study focused on hospitalized patients, and this can explain the difference in organ involvement frequencies with other cohorts.

As previously reported, in the EpiSarc study, *lupus pernio* was strongly associated with a severe phenotype [19, 20]. In this phenotype, TNF antagonists were more frequently used (41%), and patients received more treatments than in other phenotypes. *Lupus pernio* was associated with pulmonary fibrosis and ENT, CNS, bone, and joint involvement and was more frequently seen in Afro-Caribbean women.

This study has several strengths. The patients were recruited from 11 centers highly specialized in sarcoidosis. The cohort was well characterized: in addition to the verification of inclusion criteria, organ involvement was assessed from individual medical records using the WASOG instrument, leading to a strong validity of organ involvement. The main limitations of the study were the retrospective design and a possible selection bias of particularly severe patients with complicated, multiorgan involvements since we selected only hospitalized patients. We had only one cohort, so we could not obtain external validation. Additionally, the time of appearance of organ involvement since the onset of disease was not taken into account in the analysis. Both newly diagnosed cases and cases that had chronic disease were included in the cohort: for patients included as incident cases, additional organ involvements could occur during the disease course and accumulate, even if most of the time, the organ involvements of sarcoidosis are present at diagnosis. We provided a clustering based only on an organ involvement basis, but not on disease activity. With this study, we did not obtain a complete understanding of the reasons why sarcoidosis phenotypes exist: some genetic, environmental, epigenetic, and immunologic factors are increasingly evaluated as possible mechanisms for sarcoidosis etiology, but they are probably also major determinants of disease heterogeneity [21]. This should be evaluated in further studies.

In conclusion, our multicentric study highlights and confirms the existence of distinct phenotypes of sarcoidosis, with a nonrandom distribution of organ involvement, and are

associated with gender, geographical origin and socioprofessional category, in a multiethnic cohort. These phenotypes should be used to understand sarcoidosis not as a single disease but as various syndromes. The genetic and environmental determinants of such phenotypes have to be elucidated in future studies.

References

1. Valeyre D, Bernaudin JF, Uzunhan Y, Kambouchner M, Brillet PY, Soussan M, Nunes H. Clinical presentation of sarcoidosis and diagnostic work-up. *Seminars in respiratory and critical care medicine* 2014: 35(3): 336-351.
2. Lofgren S, Lundback H. The bilateral hilar lymphoma syndrome; a study of the relation to tuberculosis and sarcoidosis in 212 cases. *Acta Med Scand* 1952: 142(4): 265-273.
3. Newman LS, Rose CS, Bresnitz EA, Rossman MD, Barnard J, Frederick M, Terrin ML, Weinberger SE, Moller DR, McLennan G, Hunninghake G, DePalo L, Baughman RP, Iannuzzi MC, Judson MA, Knatterud GL, Thompson BW, Teirstein AS, Yeager H, Jr., Johns CJ, Rabin DL, Rybicki BA, Cherniack R, Group AR. A case control etiologic study of sarcoidosis: environmental and occupational risk factors. *American journal of respiratory and critical care medicine* 2004: 170(12): 1324-1330.
4. Baughman RP, Teirstein AS, Judson MA, Rossman MD, Yeager H, Jr., Bresnitz EA, DePalo L, Hunninghake G, Iannuzzi MC, Johns CJ, McLennan G, Moller DR, Newman LS, Rabin DL, Rose C, Rybicki B, Weinberger SE, Terrin ML, Knatterud GL, Cherniak R, Case Control Etiologic Study of Sarcoidosis research g. Clinical characteristics of patients in a case control study of sarcoidosis. *American journal of respiratory and critical care medicine* 2001: 164(10 Pt 1): 1885-1889.
5. Loddenkemper R, Kloppenborg A, Schoenfeld N, Grosser H, Costabel U. Clinical findings in 715 patients with newly detected pulmonary sarcoidosis--results of a cooperative study in former West Germany and Switzerland. WATL Study Group. Wissenschaftliche Arbeitsgemeinschaft fur die Therapie von Lungenkrankheitan. *Sarcoidosis, vasculitis, and diffuse lung diseases : official journal of WASOG / World Association of Sarcoidosis and Other Granulomatous Disorders* 1998: 15(2): 178-182.
6. Pereira CA, Dornfeld MC, Baughman R, Judson MA. Clinical phenotypes in sarcoidosis. *Current opinion in pulmonary medicine* 2014: 20(5): 496-502.
7. Judson MA, Boan AD, Lackland DT. The clinical course of sarcoidosis: presentation, diagnosis, and treatment in a large white and black cohort in the United States. *Sarcoidosis, vasculitis, and diffuse lung diseases : official journal of WASOG / World Association of Sarcoidosis and Other Granulomatous Disorders* 2012: 29(2): 119-127.
8. Schupp JC, Freitag-Wolf S, Bargagli E, Mihailovic-Vucinic V, Rottoli P, Grubanovic A, Muller A, Jochens A, Tittmann L, Schnerch J, Olivieri C, Fischer A, Jovanovic D, Filipovic S, Videnovic-Ivanovic J, Bresser P, Jonkers R, O'Reilly K, Ho LP, Gaede KI, Zabel P, Dubaniewicz A, Marshall B, Kieszko R, Milanowski J, Gunther A, Weihrich A, Petrek M, Kolek V, Keane MP, O'Beirne S, Donnelly S, Haraldsdottir SO, Jorundsdottir KB, Costabel U, Bonella F, Wallaert B, Grah C, Peros-Golubicic T, Luisetti M, Kadija Z, Pabst S, Grohe C, Strausz J, Vasakova M, Sterclova M, Millar A, Homolka J, Slovakova A, Kendrick Y, Crawshaw A, Wuyts W, Spencer L, Pfeifer M, Valeyre D, Poletti V, Wirtz H, Prasse A, Schreiber S, Krawczak M, Muller-Quernheim J. Phenotypes of organ involvement in sarcoidosis. *The European respiratory journal* 2018: 51(1).
9. James WE, Koutroumpakis E, Saha B, Nathani A, Saavedra L, Yucel RM, Judson MA. Clinical Features of Extrapulmonary Sarcoidosis Without Lung Involvement. *Chest* 2018: 154(2): 349-356.
10. Papiris SA, Georgakopoulos A, Papaioannou AI, Pianou N, Kallergi M, Kelekis NL, Gialafos H, Manali ED, Chatziioannou S. Emerging phenotypes of sarcoidosis based on 18F-FDG PET/CT: a hierarchical cluster analysis. *Expert Rev Respir Med* 2020: 14(2): 229-238.

11. Baughman RP, Scholand MB, Rahaghi FF. Clinical phenotyping: role in treatment decisions in sarcoidosis. *Eur Respir Rev* 2020: 29(155).
12. Statement on sarcoidosis. Joint Statement of the American Thoracic Society (ATS), the European Respiratory Society (ERS) and the World Association of Sarcoidosis and Other Granulomatous Disorders (WASOG) adopted by the ATS Board of Directors and by the ERS Executive Committee, February 1999. *American journal of respiratory and critical care medicine* 1999: 160(2): 736-755.
13. Judson MA, Costabel U, Drent M, Wells A, Maier L, Koth L, Shigemitsu H, Culver DA, Gelfand J, Valeyre D, Sweiss N, Crouser E, Morgenthau AS, Lower EE, Azuma A, Ishihara M, Morimoto S, Tetsuo Yamaguchi T, Shijubo N, Grutters JC, Rosenbach M, Li HP, Rottoli P, Inoue Y, Prasse A, Baughman RP, Organ Assessment Instrument Investigators TW. The WASOG Sarcoidosis Organ Assessment Instrument: An update of a previous clinical tool. *Sarcoidosis, vasculitis, and diffuse lung diseases : official journal of WASOG / World Association of Sarcoidosis and Other Granulomatous Disorders* 2014: 31(1): 19-27.
14. Judson MA, Baughman RP, Costabel U, Flavin S, Lo KH, Kavuru MS, Drent M, Centocor TSI. Efficacy of infliximab in extrapulmonary sarcoidosis: results from a randomised trial. *The European respiratory journal* 2008: 31(6): 1189-1196.
15. Lê S JJ, Husson F. FactoMineR: an R package for multivariate analysis. *J Stat Softw* 2008: 25: 1-28.
16. Perez-Alvarez R, Brito-Zeron P, Kostov B, Feijoo-Masso C, Fraile G, Gomez-de-la-Torre R, De-Escalante B, Lopez-Dupla M, Alguacil A, Chara-Cervantes J, Perez-Conesa M, Rascon J, Garcia-Morillo JS, Perez-Guerrero P, Fonseca-Aizpuru E, Akasbi M, Bonet M, Callejas JL, Pallares L, Ramos-Casals M, Sarco G-SR. Systemic phenotype of sarcoidosis associated with radiological stages. Analysis of 1230 patients. *European journal of internal medicine* 2019: 69: 77-85.
17. Karakaya B, Kaiser Y, van Moorsel CHM, Grunewald J. Lofgren's Syndrome: Diagnosis, Management, and Disease Pathogenesis. *Seminars in respiratory and critical care medicine* 2017: 38(4): 463-476.
18. Grunewald J, Eklund A, Olerup O. Human leukocyte antigen class I alleles and the disease course in sarcoidosis patients. *American journal of respiratory and critical care medicine* 2004: 169(6): 696-702.
19. Aubart FC, Ouayoun M, Brauner M, Attali P, Kambouchner M, Valeyre D, Nunes H. Sinonasal involvement in sarcoidosis: a case-control study of 20 patients. *Medicine* 2006: 85(6): 365-371.
20. Heidelberger V, Ingen-Housz-Oro S, Marquet A, Mahevas M, Bessis D, Bouillet L, Caux F, Chapelon-Abric C, Debarbieux S, Delaporte E, Duval-Modeste AB, Fain O, Joly P, Marchand-Adam S, Monfort JB, Noel N, Passeron T, Ruivard M, Sarrot-Reynauld F, Verrot D, Bouvry D, Fardet L, Chosidow O, Seve P, Valeyre D. Efficacy and Tolerance of Anti-Tumor Necrosis Factor alpha Agents in Cutaneous Sarcoidosis: A French Study of 46 Cases. *JAMA Dermatol* 2017: 153(7): 681-685.
21. Moller DR, Rybicki BA, Hamzeh NY, Montgomery CG, Chen ES, Drake W, Fontenot AP. Genetic, Immunologic, and Environmental Basis of Sarcoidosis. *Ann Am Thorac Soc* 2017: 14(Supplement_6): S429-S436.

Figure Legends

Figure 1. Flow chart. Description of the EpiSarc study sample enrolled from the French hospitalization database (*Programme de médicalisation des Systèmes d'Information - PMSI*) in the 11 participating centers, from January 2013 to December 2016.

Figure 2. Preferential association of organ involvements in sarcoidosis in the EpiSarc study. Hierarchical cluster analysis (A), factor map (B), and dendrogram (C).

The dendrogram includes organ involvement listed in the extrapulmonary Physician Organ Severity Tool (ePOST) score and pulmonary involvement.

Figure 3. Treatments administered in the 5 phenotypes of sarcoidosis.

Percentages of patients receiving corticosteroids, methotrexate, azathioprine, hydroxychloroquine, cyclophosphamide, and tumor necrosis factor (TNF)-alpha antagonists in the 5 phenotypes of sarcoidosis.

	All (n=1237)	Cluster 1 (n=180)	Cluster 2 (n=137)	Cluster 3 (n=630)	Cluster 4 (n=41)	Cluster 5 (n=249)	p
Male, n (%)	562 (45)	61 (34)	60 (44)	310 (49)	13 (32)	118 (47)	<0.002
Age at sarcoidosis diagnosis, mean (SD)	43.5 (13)	42 (10.2)	43.2 (16.7)	44.5 (13)	42.2 (10.5)	42.2 (13.1)	0.8
Number of organs affected, mean (SD)	3.4±1.5	2.8±1.1	2.7±1.5	3.2±1.2	4.4±1.7	4.8±1.4	<0.001
Intrathoracic involvement, n (%)	1143 (92)	178 (99)	45 (33)	630 (100)	41 (100)	249 (100)	<0.001
Intrathoracic nodes, n (%)	1115 (90)	178 (99)	23 (17)	627 (99.5)	41 (100)	246 (99)	<0.001
Parenchymal involvement, n (%)	792 (64)	71 (39)	27 (20)	453 (72)	36 (88)	205 (82)	<0.001
Fibrosis, n (%)	176 (14)	2 (1)	3 (2)	120 (19)	12 (29)	39 (16)	<0.001
Peripheral nodes, n (%)	518 (42)	43 (24)	49 (36)	204 (32)	11 (27)	211 (85)	<0.001
Skin involvement, n (%)	396 (32)	27 (15)	26 (19)	235 (37)	39 (95)	69 (28)	<0.001
Cutaneous sarcoids, n (%)	357 (29)	26 (14)	26 (19)	234 (37)	2 (5)	69 (28)	<0.001
Lupus pernio, n (%)	38 (3)	0 (0)	0 (0)	0 (0)	38 (93)	0 (0)	<0.001
Erythema nodosum, n (%)	136 (11)	120 (67)	3 (2)	3 (0.5)	2 (5)	8 (3)	<0.001
Joint involvement, n (%)	350 (28)	162 (90)	20 (15)	96 (15)	18 (44)	54 (22)	<0.001
Liver involvement, n (%)	301 (24)	11 (6)	32 (23)	85 (13)	4 (10)	169 (68)	<0.001
Eye involvement, n (%)	297 (24)	24 (13)	35 (26)	184 (29)	6 (15)	48 (19)	<0.001
Conjunctival nodules, n (%)	24 (2)	1 (0.5)	2 (1.5)	14 (2)	0 (0)	7 (3)	0.5
Uveitis, n (%)	265 (21)	22 (12)	33 (24)	166 (26)	4 (10)	40 (16)	<0.001
Anterior uveitis, n (%)*	125 (11)	13 (7)	15 (12)	77 (13)	1 (2)	19 (8)	0.04
Posterior uveitis, n (%)*	36 (3)	6 (3)	6 (5)	22 (4)	0 (0)	2 (1)	0.1
Panuveitis, n (%)*	57 (5)	4 (2)	9 (7)	34 (6)	1 (2)	9 (4)	0.2
Ear, nose and throat involvement, n (%)	280 (23)	23 (13)	25 (18)	132 (21)	28 (68)	72 (29)	<0.001
Nasosinusal involvement, n (%)	131 (11)	6 (3)	8 (6)	72 (11)	20 (50)	25 (10)	<0.001

Laryngopharyngeal involvement, n (%)	28 (2)	0 (0)	5 (4)	14 (2)	4 (10)	5 (2)	0.003
Parotid and lacrimal gland involvement, n (%)	137 (11)	16 (9)	13 (10)	55 (9)	3 (7)	50 (20)	<0.001
Neurological involvement, n (%)	271 (22)	13 (7)	43 (31)	181 (28)	10 (24)	24 (10)	<0.001
Cranial nerve involvement, n (%)	130 (11)	12 (7)	19 (14)	87 (14)	3 (7)	9 (4)	<0.001
Optical nerve involvement, n (%)	55 (4)	4 (2)	9 (7)	39 (6)	1 (2.5)	2 (1)	0.003
Medullary involvement, n (%)	37 (3)	2 (1)	7 (5)	23 (4)	2 (5)	3 (1)	0.07
Parenchymal involvement, n (%)	97 (8)	2 (1)	20 (15)	67 (11)	4 (10)	4 (2)	<0.001
Pituitary infiltration, n (%)	33 (3)	0 (0)	6 (5)	21 (4)	2 (5)	4 (2)	0.051
Cerebrospinal fluid inflammation, n (%)	103 (8)	4 (2)	16 (12)	73 (12)	3 (7)	7 (3)	<0.001
Neuromuscular involvement, n (%)	42 (3)	2 (1)	9 (7)	21 (3)	2 (5)	8 (3)	0.1
Spleen involvement, n (%)	195 (16)	3 (2)	10 (7)	11 (2)	3 (7)	168 (67)	<0.001
Cardiac involvement, n (%)	180 (15)	11 (6)	8 (6)	120 (19)	7 (17)	34 (14)	<0.001
Bone involvement, n (%)	118 (10)	7 (4)	8 (6)	20 (3)	10 (24)	73 (29)	<0.001
Kidney involvement, n (%)	75 (6)	3 (2)	16 (12)	37 (6)	1 (2.5)	18 (7)	0.004
Digestive involvement, n (%)	41 (3)	1 (0.5)	33 (24)	2 (0.5)	0 (0)	5 (2)	<0.001
Hypercalcemia, n (%)	99 (8)	7 (4)	6 (4)	51 (8)	3 (7)	31(12)	0.01

Table 1. Clinical presentation of patients in the EpiSarc study Characteristics of the whole cohort, and according to the cluster hierarchization

* Out of 1,188 patients (among the 265 patients with uveitis, the characteristics of uveitis were not available in 49)

The results are expressed as n (%) except if otherwise specified

SD: standard deviation

	All (n=1237)	Cluster 1 (n=180)	Cluster 2 (n=137)	Cluster 3 (n=630)	Cluster 4 (n=41)	Cluster 5 (n=249)	p
Occupational status	**964 (78)**	**137 (76)**	**102 (74)**	**493 (78)**	**27 (66)**	**205 (82)**	
Employed	**887 (92)**	**125 (92)**	**88 (86)**	**461 (94)**	**25 (93)**	**188 (92)**	**<0.001**
Upper class: managers and professionals	66 (8)	9 (7)	9 (10)	31 (7)	3 (12)	14 (8)	
Clerk support workers	259 (29)	41 (33)	22 (25)	116 (25)	11 (44)	69 (37)	
Intermediary occupations	168 (19)	37 (30)	12 (14)	85 (18)	4 (16)	30 (16)	
Craftsmen and trade-related workers	26 (3)	1 (0.5)	5 (6)	12 (3)	0 (0)	8 (4)	
Labor workers	360 (41)	37 (30)	39 (44)	211 (46)	7 (28)	66 (35)	
Skilled agricultural workers	8 (1)	0 (0)	1 (1)	6 (1)	0 (0)	1 (0.5)	
Student	**14 (1.5)**	**1 (0.5)**	**7 (7)**	**2 (0.5)**	**0 (0)**	**4 (2)**	
Unemployed	**63 (6.5)**	**11 (8)**	**7 (7)**	**30 (6)**	**2 (7)**	**13 (6)**	
Geographical origins, n (%)	123 (99.5)	179 (99)	136 (99)	627 (99)	40 (98)	248 (99.5)	
European	541 (44)	91 (51)	85 (62.5)	264 (42)	8 (20)	93 (37.5)	0.01
Northern African	305 (25)	47 (26)	24 (18)	159 (25)	15 (37.5)	60 (24)	
Sub-Saharan Africa	208 (17)	26 (15)	17 (12.5)	106 (17)	6 (15)	53 (21.5)	
Overseas departments	106 (8.5)	9 (5)	5 (3.5)	57 (9)	8 (20)	27 (11)	
Others	70 (5.5)	6 (3)	5 (3.5)	41 (7)	3 (7.5)	15 (6)	
Familial sarcoidosis*	62 (8)	4 (4)	6 (7)	35 (10)	3 (11)	14 (8)	0.4
Follow-up duration, (years), mean±SD	8.1±8	6.3±7.1	7±7.1	8.8±8.6	14.6±9.7	7.5±8	0.03
Death, n (%)	23 (2)	0 (0)	2 (1)	18 (3)	2 (5)	1 (0.5)	0.01
Treatments							
Corticosteroids	906 (73)	95 (53)	98 (72)	492 (78)	36 (88)	185 (74)	<0.001
Methotrexate	388 (31)	34 (19)	40 (22)	209 (33)	31 (76)	74 (30)	<0.001
Azathioprine	161 (13)	13 (7)	10 (7)	85 (14)	14 (34)	39 (16)	<0.001
Hydroxychloroquine	283 (23)	43 (24)	20 (15)	138 (22)	20 (49)	62 (25)	<0.001
TNF-alpha blockers	113 (9)	7 (4)	15 (11)	62 (10)	17 (41)	12 (5)	<0.001
Cyclophosphamide	61 (5)	3 (2)	13 (9)	36 (6)	3 (7)	6 (2)	0.003

Table 2. Occupational activities, geographical origins, and outcomes of patients in the EpiSarc Study

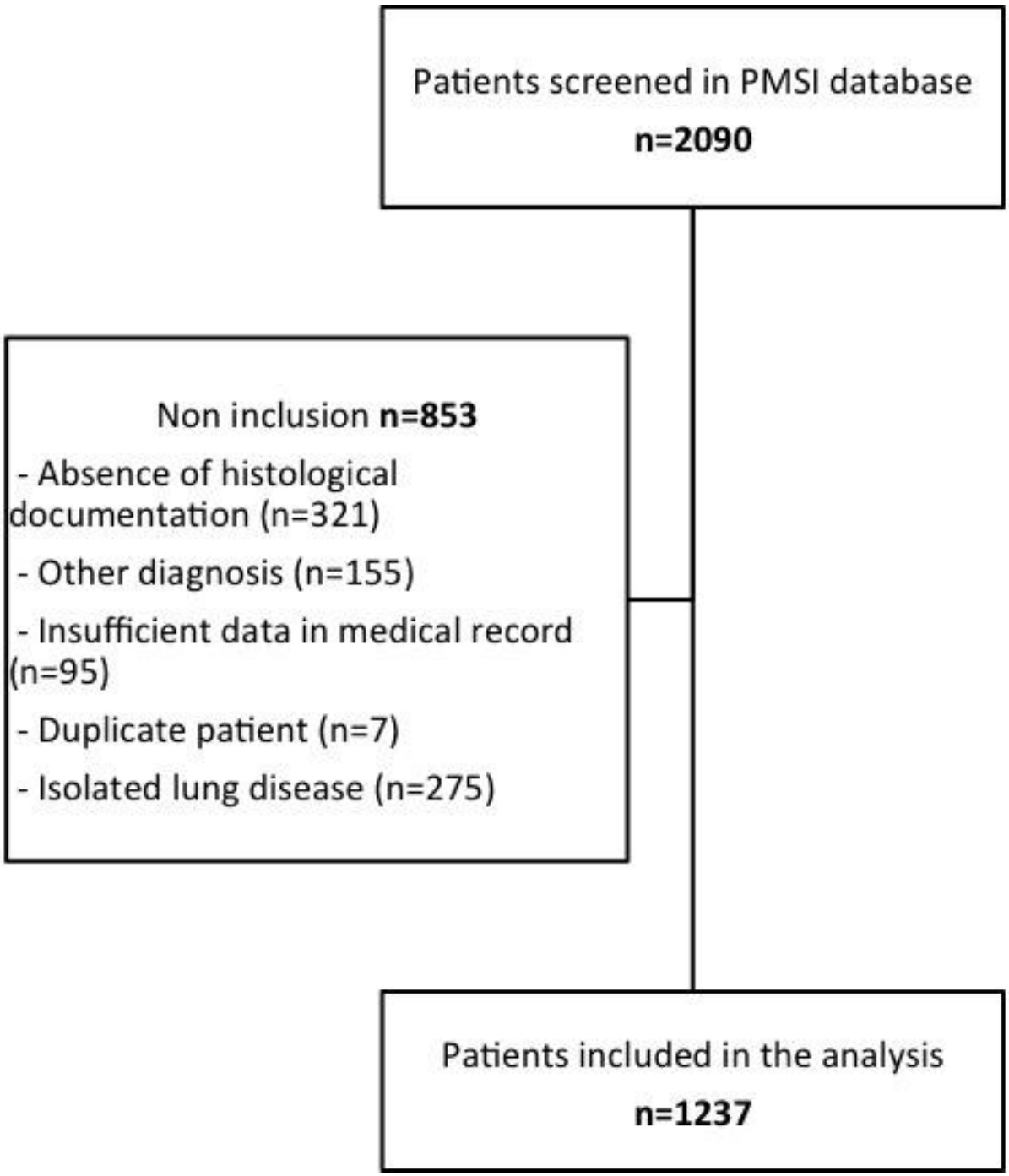
* Data presented are from 775 patients (for the others, this information was not available)

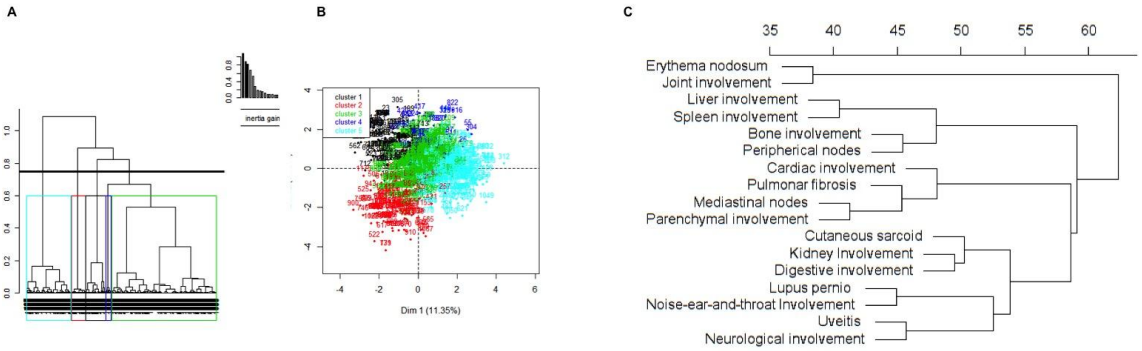
The results are expressed as n (%) except if otherwise specified

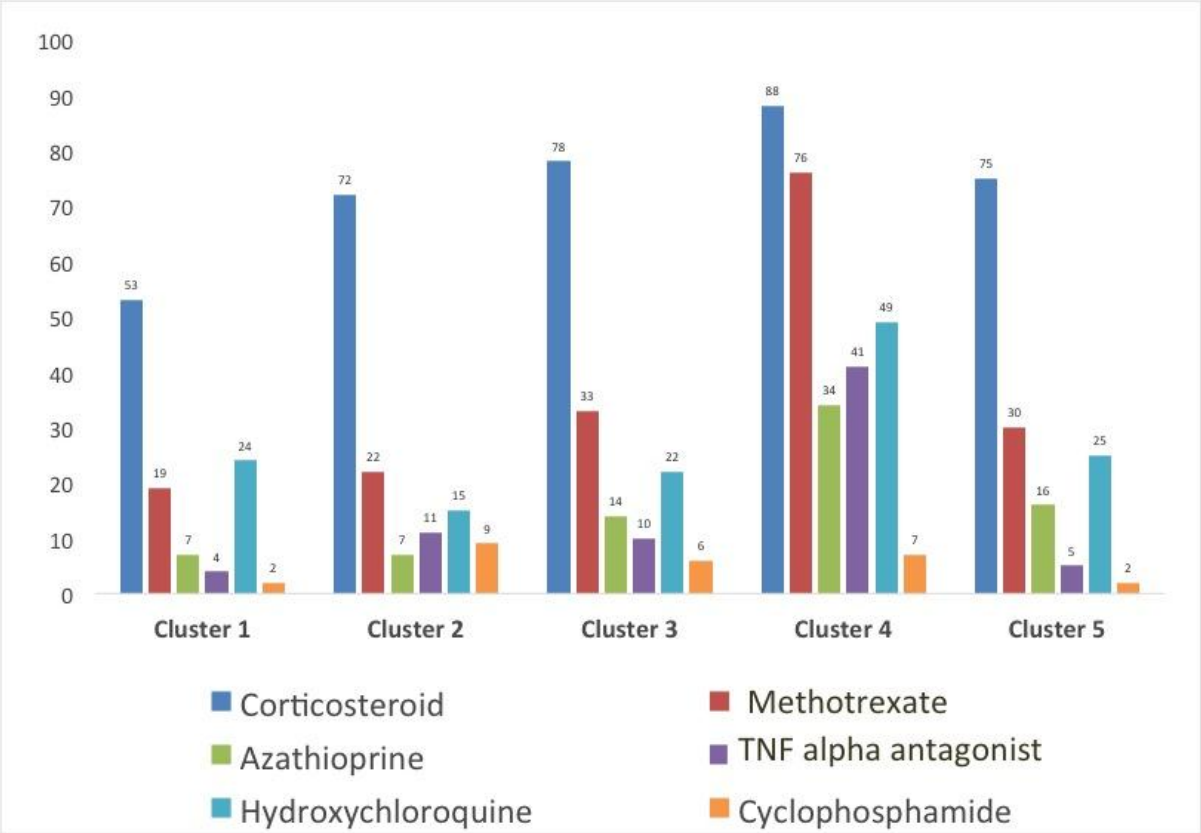
SD: standard deviation

	Cluster 1 (n=180)	Cluster 2 (n=137)	Cluster 3 (n=630)	Cluster 4 (n=41)	Cluster 5 (n=249)
Sex (male)					
OR (95% CI)	**0.6 [0.4-0.9]**	0.8 [0.5-1.3]	1	0.6 [0.3-1.5]	0.9 [0.6-1.3]
P	**0.008**	0.4	NA	0.3	0.5
Labor worker					
OR (95% CI)	**0.6 [0.4-0.7]**	0.9 [0.6-1.5]	1	0.5 [0.2-1.3]	**0.6 [0.4-0.9]**
P	**0.03**	0.7	NA	0.1	**0.01**
European					
OR (95% CI)	1.7 [1.1-2.5]	**2.2 [1.4-3.4]**	1	0.4 [0.2-1.1]	0.9 [0.6-1.3]
P	0.01	**<0.001**	NA	0.08	0.6
Sub-Saharan African					
OR (95% CI)	0.6 [0.4-1.2]	0.7 [0.4-1.3]	1	1.1 [0.4-2.9]	1.2 [0.8-1.8]
p	0.1	0.3	NA	0.9	0.5
Caribbean					
OR (95% CI)	0.5 [0.2-1.2]	0.4 [0.2-1.3]	1	2.5 [0.9-7.1]	1.3 [0.7-2.2]
p	0.1	0.1	NA	0.08	0.4
North African					
OR (95% CI)	1 [0.6-1.6]	0.7 [0.4-1.2]	1	1.3 [0.6-3.2]	0.9 [0.6-1.4]
p	0.98	0.2	NA	0.5	0.7
Others					
OR (95% CI)	0.5 [0.2-1.4]	0.6 [0.2-1.7]	1	1.2 [0.3-5.3]	0.9 [0.5-1.8]
p	0.2	0.4	NA	0.8	0.8

Table 3. Factors related to clusters of sarcoidosis of patients in the EpiSarc study
OR odds ratios with 95% confidence intervals (CI) in each cluster adjusted for gender, geographical origin and socioprofessional activities. Reference is cluster 3.


Patients screened in PMSI database
n=2090
Non inclusion n=853
- Absence of histological documentation (n=321)
- Other diagnosis (n=155)
- Insufficient data in medical record (n=95)
- Duplicate patient (n=7)
- Isolated lung disease (n=275)
Patients included in the analysis
n=1237


A
B
C
inertia gain
cluster 1
cluster 2
cluster 3
cluster 4
cluster 5
Dim 1 (11.35%)
35
40
45
50
55
60
Erythema nodosum
Joint involvement
Liver involvement
Spleen involvement
Bone involvement
Peripherical nodes
Cardiac involvement
Pulmonar fibrosis
Mediastinal nodes
Parenchymal involvement
Cutaneous sarcoid
Kidney Involvement
Digestive involvement
Lupus pernio
Noise-ear-and-throat Involvement
Uveitis
Neurological involvement


100
90
80
70
60
50
40
30
20
10
0
53
19
7
4
24
2
72
22
7
11
15
9
78
33
14
10
22
6
88
76
34
41
49
7
75
30
16
5
25
2
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Corticosteroid
Methotrexate
Azathioprine
TNF alpha antagonist
Hydroxychloroquine
Cyclophosphamide