Clinical phenotypes of sarcoidosis using cluster analysis: a Spanish population-based cohort study

R. Fernández-Ramón[1], J.J. Gaitán-Valdizán[1], J.L. Martín-Varillas[2], R. Demetrio-Pablo[1], I. Ferraz-Amaro[3], S. Castañeda[4], R. Blanco[5,6]

[1]Department of Ophthalmology, Hospital Universitario Marqués de Valdecilla, Santander; [2]Department of Rheumatology, Hospital de Laredo, Cantabria; [3]Department of Rheumatology, Hospital Universitario de Canarias, Tenerife; [4]Department of Rheumatology, Hospital Universitario de La Princesa, IIS-Princesa, Madrid; Cátedra UAM-Roche, EPID-Future, Universidad Autónoma de Madrid (UAM), Madrid; [5]Department of Rheumatology, Hospital Universitario Marqués de Valdecilla, Santander; [6]Immunopathology Group, Marqués de Valdecilla University Hospital-IDIVAL, Santander, Spain.

Abstract

Objective

Sarcoidosis is a clinically heterogenous disease. The objective of this study is the identification of clinical phenotypes using cluster analysis.

Methods

A model-based clustering relaying on 19 clinical variables was performed in a retrospective cohort of 342 sarcoidosis patients, diagnosed and followed-up from 1999 to 2019 in a tertiary hospital at Northern Spain. Chi-square test and ANOVA were used to compare categorical and continuous variables among groups. Two-sample t-tests and the partition of Pearson's chi-square statistic were used in pairwise comparisons. The Wasfi severity score was calculated and compared among clusters.

Results

Cluster analysis identified five groups: C1 (16.1%), C2 (14.3%), C3 (24.3%), C4 (5.0%), and C5 (40.4%). Lung involvement was predominant, ranging from 55.1% (C2) to 100% (C1 and C4). Extrapulmonary involvement was significantly higher in C2 (96.4%) and C3 (98.0%). A significant lower FEV1 percent predicted was detected in C5 (90.5±21.8) versus C1 (102.0±22.9), C3 (102.3±17.6) and C4 (105.8±20.8). The cluster 5 had a lower FVC percent predicted (96.6±18.9) than others, ranging from 108.1±18.0 (C3) to 111.5±21.7 (C4). The prescription of systemic glucocorticoids and non-corticosteroid immunosuppressants was higher in the clusters 1, 3 and 5. Chronicity rates were higher in C3 (31.3%) and C5 (32.6%) compared to C1 (9.1%) and C4 (0%), as well as the Wasfi severity score values.

Conclusion

Five phenotypes with different clinical and prognostic characteristics are proposed in our study. Cluster analysis can be a useful tool for identifying clinical patterns in a disease as heterogeneous as sarcoidosis and optimising its management.

Key words

clinical phenotypes, sarcoidosis, cluster analysis, treatment, prognosis

Raúl Fernández-Ramón, MD
Jorge J. Gaitán-Valdizán, MD
José L Martín-Varillas, MD
Rosalía Demetrio-Pablo, MD, PhD
Iván Ferraz-Amaro, MD, PhD
*Santos Castañeda, MD, PhD **
*Ricardo Blanco, MD, PhD **

**Share senior authorship.*

Please address correspondence to:
Ricardo Blanco
Hospital Universitario
Marqués de Valdecilla,
Avda. Valdecilla s/n,
39008 Santander, Spain.
E-mail: rblancovela@gmail.com

Received on January 10, 2024; accepted in revised form on April 8, 2024.

Competing interests: R. Blanco has received grants/research support from AbbVie, MSD and Roche, and had consultation fees/participation in a company-sponsored speaker's bureau from AbbVie, Bristol-Myers, Galapagos, GSK, Janssen, Lilly, MSD, Novartis, Pfizer and Roche.
The other authors have declared no competing interests.

Introduction

Sarcoidosis is a granulomatous disease of unknown aetiology, affecting any organ of the body, with a very heterogeneous course (1-3). There is a well-defined and widely accepted acute pattern characterised by a self-limited clinical course and good prognosis (Löfgren's syndrome). This is clinically defined by the triad of erythema nodosum, pulmonary hilar lymphadenopathy and arthritis (4-6). However, in the subacute and chronic forms of the disease, often there is no such specific picture (3, 7). Therefore, phenotyping of sarcoidosis can be a useful tool in its management, by grouping patients with similar clinical characteristics and identifying those with a worse prognosis and need for long-term therapy (8). Interestingly, several clinical characteristics have been related to disease prognosis, including race, organ involvement, body mass index (BMI), therapeutic status, Scadding lung stage, and respiratory function (3, 9-11), and can help us in the classification of these patients.

In recent years, the use of cluster analysis to identify phenotypes in diseases with clinical heterogeneity such as sarcoidosis has become popular. The clustering methodology employs multivariate techniques to establish groups with homogeneous characteristics. Overall, it is considered a relatively objective tool due to the statistical methods it uses. However, the selection of input variables remains a subjective procedure (8). There are studies that have used this methodology in the phenotyping of sarcoidosis according to organic involvement. However, there are few that help predict severity and prognosis of this disease (12-14).

Therefore, the aim of this study is to present a phenotyping model using a cluster analysis in a cohort of patients with sarcoidosis, collected and followed over two decades, in a tertiary hospital in Northern Spain. With this method of analysis, we seek to transform clinical variables into "disease phenotypes" that allow us to classify better the different kind of patients and improve their treatment accordingly.

Patients and methods

Study design and data collection

A retrospective study on sarcoidosis cases diagnosed and followed-up from January 1st 1999 to December 31th 2019 at a sarcoidosis referral hospital in Northern Spain was performed. Our centre is the only referral hospital in a population of about 300,000 inhabitants, corresponding to Healthcare Area 1 of Cantabria, Spain. Therefore, all patients with sarcoidosis are diagnosed and followed exclusively by different Units (Rheumatology, Pulmonology, Ophthalmology, Dermatology, Internal Medicine, Cardiology, Neurology, Radiology, Nuclear Medicine and Pathology) in our hospital. Patients with no follow-up due to belonging to other areas were excluded from this study. Sarcoidosis was diagnosed according to American Thoracic Society/European Respiratory Society/World Association of Sarcoidosis and Other Granulomatous Disorders (ATS/ERS/WASOG) criteria (15).

The collected variables included were sex, age at diagnosis, ethnicity, specific organ involvement, radiological stage, laboratory tests results, pulmonary function tests (PFTs), and treatment program. In the absence of histological confirmation, organ involvement was determined using the "highly probable" and "probable" criteria according the WASOG organ assessment instrument (16). Radiological stage was established by an expert radiologist using the chest x-ray closest to enrolment date in accordance with Scadding's classification (17). PFTs used in the analysis included forced vital capacity (FVC), forced expiratory volume at first second (FEV1), and FEV1/FVC ratio. FEV1 and FVC<80%, and FEV1/FVC<70% of the predicted values were considered abnormal.

Treatment schedule was classified in corticosteroid and non-corticosteroid immunosuppressive (IS) therapy. Among IS agents, we included conventional synthetic immunosuppressants (methotrexate, azathioprine, mycophenolate mofetil, and leflunomide) as well as biologic drugs (infliximab, adalimumab, golimumab, rituximab and tocilizumab).

Chronicity definition

Chronic sarcoidosis was defined by the presence of symptoms after 5 years of disease and/or irreversible organ dam-

age including pulmonary fibrosis and FVC<60% and/or the diffusing capacity of the lungs for carbon monoxide (DLCO)<50%, pulmonary hypertension, neurosarcoidosis with sequelae, cardiac sarcoidosis, chronic skin involvement, chronic uveitis, chronic renal failure, portal hypertension or chronic symptomatic osteoarticular involvement. Resolution of the disease was considered when symptoms, radiological findings and laboratory tests normalised.

Severity score

To evaluate the severity of the disease, we used the severity score proposed by Wasfi *et al*. (18). It was calculated using the following equation:

Severity score = 11.46 + 3.9(C) + 2.51(N) + 1.56(IS) – 0.051 (FVC% predicted) + 1.75(AA) – 0.054(FEV1/FVC),

where C=1 if there is cardiac involvement, 0 if not; N=1 if there is neurologic involvement, 0 if not; IS=1 if the patient is receiving non-corticosteroid immunosuppression therapy, 0 if not; AA=1 if the subject is Afro-American.

Those patients without respiratory function tests were excluded of the analysis.

Statistical analysis

A model-based clustering was performed to identify sarcoidosis phenotypes based on demographics and clinical characteristics. They were summarised as absolute number and percentage for categorical variables or mean and standard deviation (SD) for continuous variables. A hierarchical ascendant clustering analysis was performed using Ward's minimum variance method. Results were graphically represented in a dendrogram and the number of clusters was estimated by a visual distance criterion.

Chi-square test and ANOVA were used to compare, respectively, categorical and continuous variables among groups. Two-sample t-tests for quantitative features and the partition of Pearson's chi-square statistic for categorical features were used in pairwise comparisons among clusters. False Discovery Rate was calculated using the Benjamini-Hochberg method to correct for multiple comparisons; its results are displayed as q-values. Results with q-values <0.05 were considered as statistically significant. All statistical analyses were performed with IBM SPSS Statistics for Windows, v. 20.0 (IBM Corp. Armonk, N.Y., USA).

Ethical approval

The study was approved by the Cantabria Clinical Research Ethics Committee (protocol no. 2020.257) and was conducted in according to the latest modifications of the Helsinki Declaration (Fortaleza, Brazil, October 2013).

Results

Demographical and clinical features of study population

A total of 342 patients were included in the study. Demographic and clinical characteristics of our cohort are summarised in Table I. A slight female predominance was observed (51.8%) and most of subjects were Caucasian (94.2%). The mean age at diagnosis was 47.7±15.1 years. Lung was the most commonly affected organ (88.3%). The most frequently involved extra-pulmonary organs included skin (34.2%), joints (27.8%), eye (17.8%) and liver (9.6%). The mean number of organs involved was 2.0±1.0. Most individuals were treated with systemic corticosteroids (60.2%), while 28.1% required non-corticosteroid immunosuppressive (IS) therapy within 21 years of enrolment.

Cluster analysis

The hierarchical clustering identified five homogeneous groups: C1 (n=55; 16.1%), C2 (n=49; 14.3%), C3 (n=83; 24.3%), C4 (n=17; 5.0%), and C5 (n=138; 40.4%). The variables used and their distribution in the cluster analysis are represented in the dendrogram of Figure 1. Differences across clusters in demographic, clinical and therapeutic variables as shown in Table I.

The proportion of males was significantly higher in cluster 5 (65.2%) than in the remaining groups, ranging from 28.6% (C2) to 47.1% (C4). The mean age at diagnosis was significantly lower in C1 (44.6±13.5 years) and C5 (46.6±14.3 years), compared to C2 (52.0±14.5 years) (Table I).

Lung involvement was clearly predominant in all clusters, ranging from 89.9% (C5) to 100% (C1 and C4), except for C2 (55.1%). The Scadding lung stages were significantly variable, being predominant stage I in C1 (100%), C3 (95.2%) and C4 (100%); while in cluster 5, parenchymal lung involvement prevails, mostly stage II (67.4%), and all cases with pulmonary fibrosis were included in this group. The high percentage of patients with stage 0 in cluster 2 (44.9%) is remarkable. Extra-pulmonary involvement was significantly higher in C1 (96.4%) and C2 (98.0%), with skin lesions, predominating erythema nodosum in C1, and granulomatous lesions in C2, as well as joint involvement in C1. Other predominantly affected organs were the liver in C2 (34.7%), as well as the eyes (36.1%) and the central nervous system (CNS) (24.1%) in C3. It is worth noting the exclusive hilar adenopathy in cluster 4, with no parenchymal lung and/or extrathoracic involvement. Löfgren's syndrome, an acute form of sarcoidosis, was significantly more frequent in C1 (43.6%) compared to the other clusters (Table I).

Regarding the symptoms at the onset of the disease, respiratory symptoms such as cough and dyspnea were common in C5 (23.9% and 44.2%, respectively), asthenia predominated in C2 (36.7%) and C3 (50.6%); and fever was notable in C1 (29.1%).

Respiratory function tests were available for 243 patients and their results according to clusters are represented in Figure 2. A significant lower mean FEV1 were detected in C5 (90.5±21.8%), *versus* C1 (102.0±22.9%), C3 (102.3±17.6%) and C4 (105.8±20.8). The cluster 5 had a significantly lower mean FVC (96.6±18.9%) than the other clusters, ranging from 105.6±20.3% (C2) to 111.5±21.7% (C4). No differences were noted in FEV1/FVC ratio and DLCO.

Systemic steroids were significantly more used in cluster 5 (76.1%) compared to C1: 54.5%; C2: 34.7%; C3: 61.4%, and C4: 17.6%. In C1, mean maximum dose of oral prednisone was significantly lower than in the rest of the clusters. Nevertheless, no differences were observed regarding the duration of treatment. The prescription of non-corticosteroid IS drugs was signifi-

Table I. Demographic and clinical characteristics of the study population.

Characteristics n (%)	Whole cohort	C1	C2	C3	C4	C5	*p*-value	q-value
Total	342 (100)	55 (16.1)	49 (14.3)	83 (24.3)	17 (5.0)	138 (40.4)		
Gender (female)	177 (51.8)	36 (65.5) [A]	35 (71.4) [B]	49 (59.0) [C]	9 (52.9)	48 (34.8) [A,B,C]	<0.001	<0.001
Follow-up period (years), mean±SD	9.1±5.8	9.0±6.2	7.5±5.8 [A]	8.5±6.0 [B]	9.1±6.2	10.1±5.3 [A,B]	0.06	0.07
Age at diagnosis (years), mean±SD	47.7±15.1	44.6±13.5 [A]	52.0±14.5 [A,B]	48.4±16.8	50.3±17.0	46.6±14.3 [B]	0.1	0.1
Ethnicity							0.2	0.2
Caucasian	322 (94.2)	53 (94.5)	45 (91.8)	76 (92.7)	16 (94.1)	133 (96.4)		
Hispanic	15 (4.4)	2 (3.6)	2 (4.1)	5 (6.1)	0	4 (2.9)		
Black	4 (1.2)	1 (1.8)	2 (4.1)	1 (1.2)	1 (5.9)	1 (0.7)		
Smoking history	118 (34.5)	15 (27.3)	17 (34.7)	22 (26.5) [A]	6 (35.3)	58 (42.0) [A]	0.1	0.1
Lung involvement	302 (88.3)	55 (100) [A,B]	27 (55.1) [A,C,D,E]	79 (95.2) [C]	17 (100) [D]	124 (89.9) [B,E]	<0.001	<0.001
Scadding stage							<0.001	<0.001
0	40 (11.7)	0 [A,B]	22 (44.9) [A,C,D,E]	4 (4.8) [C]	0 [D]	14 (10.1) [B,E]	<0.001	<0.001
I	173 (50.6)	55 (100) [A,E]	22 (44.9) [A,B,C,D,F]	79 (95.2) [B,G]	17 (100) [C,,H]	2 (1.4) [E,F,G,H]	<0.001	<0.001
II	100 (29.2)	0 [A,D]	5 (10.2) [A,B,C,E]	0 [F]	0 [C,G]	93 (67.4) [D,E,F,G]	<0.001	<0.001
III	20 (5.8)	0 [A]	0 [B]	0 [C]	0	20 (14.5) [A,B,C]	<0.001	<0.001
IV	9 (2.6)	0	0	0 [A]	0	9 (6.5) [A]	0.008	0.009
Extra-pulmonary involvement	234 (68.6)	53 (96.4) [A,B,C]	48 (98.0) [D,E]	55 (66.3) [A,D,E]	0 [B,D,E]	78 (56.5) [C,E]	<0.001	<0.001
Skin	117 (34.2)	38 (69.1) [A,B,C]	32 (65.3) [D,E]	17 (20.5) [A,D]	0 [B, D]	30 (21.7) [C,E,F]	<0.001	<0.001
Nodosum erythema	70 (20.5)	35 (63.6) [A,B,C,D]	0 [A,E,F]	16 (19.3) [B,E]	0 [C]	19 (13.8) [D,F]	<0.001	<0.001
Granulomatous lesions	47 (13.8)	3 (5.5) [A]	32 (65.3) [A,B,C,D]	1 (1.2) [B,E]	0 [C]	11 (8.0) [D,E]	<0.001	<0.001
Eye	61 (17.8)	0 [A,B]	1 (2.0) [C,D]	30 (36.1) [A,C,E]	0 [E]	30 (21.7) [B,D,E]	0.02	0.02
Anterior uveitis	31 (9.1)	0 [A,B]	1 (2.0) [C]	19 (22.9) [A,C,D,E]	0 [D]	11 (8.0) [B,E]	<0.001	<0.001
Intermediate uveitis	1 (0.3)	0	0	1 (1.2)	0	0	0.5	0.4
Posterior uveitis	7 (2.0)	0	0	1 (1.2)	0	6 (4.3)	0.2	0.2
Panuveitis	12 (3.5)	0	0	5 (6.0)	0	7 (5.1)	0.1	0.1
Liver	33 (9.6)	1 (1.8) [A]	17 (34.7) [A,B,C,D]	3 (3.6) [B]	0 [C]	12 (8.7) [D]	<0.001	<0.001
Spleen	7 (2.0)	0	1 (2.0)	1 (1.2)	0	5 (3.6)	0.5	0.4
Bone	4 (1.2)	1 (1.8)	1 (2.0)	1 (1.2)	0	1 (0.7)	0.9	0.7
Joint	95 (27.8)	37 (67.3) [A,B,C,D]	8 (16.3) [A]	19 (22.9) [B,E]	0 [C,E,F]	33 (23.9) [D,F]	<0.001	<0.001
Parotid	8 (2.3)	0	0	4 (4.8)	0	4 (2.9)	0.3	0.3
Kidney	13 (3.8)	0 [A]	4 (8.2) [A,B,D]	0 [B]	0	2 (1.4) [D]	0.006	0.007
Hypercalcaemia	17 (5.0)	1 (1.8) [A]	8 (16.3) [A,B,C]	2 (2.4) [B]	0	6 (4.3) [C]	0.002	0.002
CNS	29 (8.5)	1 (1.8) [A]	0 [B]	20 (24.1) [A,B,C,D]	0 [C]	8 (5.8) [D]	<0.001	<0.001
Heart	5 (1.5)	0	1 (2.0)	0	0	4 (2.9)	0.4	0.4
Löfgren's syndrome	43 (12.6)	24 (43.6) [A,B,C,D]	2 (4.1) [A]	9 (11.0) [B]	0 [C]	8 (5.8) [D]	<0.001	<0.001
Number of involved organs	2.0±1.0	2.4±0.7 [A,B,C,D]	1.9±0.8 [A,E]	2.1±1.1 [B,F]	1.0±0.0 [C,E,F,G]	1.8±1.1 [D,G]	<0.001	<0.001
Cough	51 (14.9)	3 (5.5) [A]	1 (2.0) [B,C]	14 (16.9) [B]	0 [D]	33 (23.9) [A,C,D]	<0.001	<0.001
Dyspnoea	94 (27.5)	11 (20.0) [A]	5 (10.2) [B]	17 (20.5) [C,D]	0 [D,E]	61 (44.2) [A,B,C,E]	<0.001	<0.001
Asthenia	96 (28.1)	1 (1.8) [A,B,C]	18 (36.7) [A,D]	42 (50.6) [B,E,F]	0 [D,E,G]	35 (25.4) [C,F,G]	<0.001	<0.001
Fever	38 (11.1)	16 (29.1) [A,B,C,D]	0 [A,E,F]	7 (8.4) [B,E]	0 [C]	15 (10.9) [D,F]	<0.001	<0.001
Treatment								
Corticosteroids	206 (60.2)	30 (54.5) [A,B,C]	17 (34.7) [A,D,E]	51 (61.4) [D,F]	3 (17.6) [B,F]	105 (76.1) [C,E,F]	<0.001	<0.001
Duration (weeks)	176.7±177.5	127.8.5±144.8	115.5±101.1	194.6±195.8	8.7±1.5	195.3±184.1	0.08	0.08
Maximum dose (mg/kg)	41.6±18.5	21.8±11.7 [A,B,C,D]	45.9±15.4 [A]	41.3±17.7 [B]	50.00±17.3 [C]	46.5±17.2 [D]	<0.001	<0.001
Immunosuppressants	96 (28.1)	13 (23.6) [A,B]	7 (14.3) [C,D]	35 (42.2) [A,C,E]	0 [B,E,F]	41 (29.7) [D,F]	<0.001	<0.001
CIS	87 (25.4)	13 (23.6) [A]	6 (12.2) [B,C]	32 (38.6) [B,D]	0 [A,D,E]	36 (26.1) [C,E]	0.001	0.002
Methotrexate	75 (21.9)	10 (18.2) [A]	4 (8.2) [B,C]	31 (37.3) [A,B,D]	0 [D,E]	30 (21.7) [C,E]	<0.001	<0.001
Duration (weeks)	189.0±174.0	133.5±178.3	135.1±156.3	168.7±133.7	0	236.4±205.4	0.3	0.3
Azathioprine	23 (6.7)	2 (3.6)	2 (4.1)	6 (7.2)	0	13 (9.4)	0.4	0.4
Duration (weeks)	192.4±174.9	275.9±72.9	172.6±43.0	319.5±153.6	0	128.9±182.5	0.2	0.2
Mycophenolate	2 (0.6)	0	0	1 (1.2)	0	1 (0.7)	0.9	0.9
Duration (weeks)	181.5±220.5	0	0	25.6	337.4	0		
Leflunomide	6 (1.8)	1 (1.8)	1 (2.0)	2 (2.4)	0	2 (1.4)	0.9	0.9
Duration (weeks)	244.3±326.0	14.9	178.1	175.1±21.1	0	450.9±606.6	0.8	0.8
Biological therapy	44 (12.9)	2 (3.6) [A,B]	1 (2.0) [C,D]	18 (21.7) [A,C,E]	0 [E,F]	23 (16.7) [B,D,F]	0.001	0.002
Infliximab	28 (8.2)	1 (1.8) [A]	1 (2.0) [B]	12 (14.5) [A,B]	0	14 (10.1)	0.02	0.03
Duration (weeks)	115.8±111.2	81.4	178.3	111.3±95.0	0	117.8±134.4	0.9	0.9
Adalimumab	31 (9.1)	1 (1.8) [A,B]	2 (4.1)	10 (12.0) [A]	0	18 (13.0) [B]	0.04	0.04
Duration (weeks)	121.8±126.9	195.9	63.6±10.7	101.1±80.9	0	136.4±155.7	0.8	0.8
Golimumab	5 (1.5)	0	0	4 (4.8)	0	1 (0.7)	0.07	0.07
Duration (weeks)	208.2±219.5	0	0	265.0±230.0	0	37.9	0.5	0.5
Tocilizumab	2 (0.6)	0	0	1 (1.2)	0	1 (0.7)	0.9	0.9
Duration (weeks)	189.3±239.4	0	0	358.6	0	19.9		
Rituximab	3 (0.9)	1 (1.8)	0	0	0	2 (1.4)	0.7	0.7
Duration (weeks)	37.9±34.1	1	0	0	0	56.3±17.0	0.3	0.3
Chronicity	87 (26.5)	5 (9.1) [A,B]	10 (20.4)	26 (31.3) [A,C]	0 [C,D]	45 (32.6) [B,D]	0.001	0.002
Sequelae / irreversible organ damage	47 (13.7)	2 (3.6) [A]	6 (12.2)	11 (13.3)	0 [B]	27 (19.6) [A,B]	0.02	0.03

CNS: central nervous system; CIS: conventional synthetic immunosuppressant agents; SD: standard deviation.
Data presented: n (%) or mean ± SD.
C1: erythema nodosum and articular involvement; C2: miscellaneous extrapulmonary sarcoidosis; C3: ocular and/or neurological involvement; C4: isolated hilar adenopathy; C5: parenchymal lung involvement with dyspnea.
Clusters sharing the same letter (*A, B, C, D, E, F, G*) means that they have statistically significant differences (q<0.05) in the pairwise comparison.

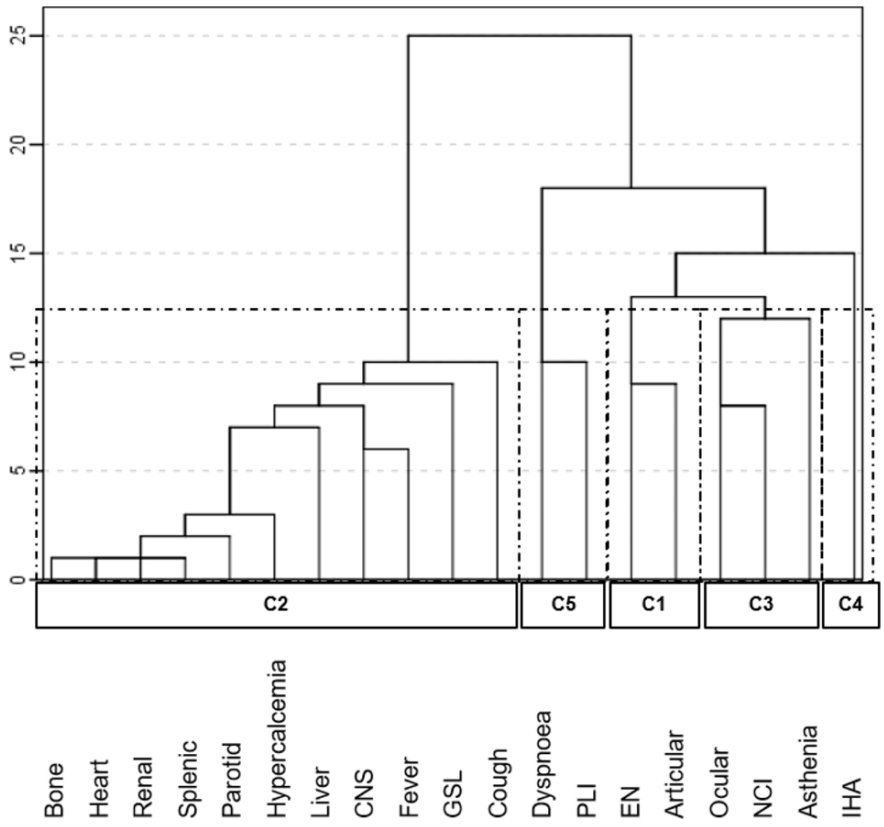

Fig. 1. Dendrogram of the hierarchical clustering leading to five clusters. **C1**: erythema nodosum and articular involvement; **C2**: miscellaneous extrapulmonary sarcoidosis; **C3**: ocular and/or neurological involvement; **C4**: isolated hilar adenopathy; **C5**: parenchymal lung involvement with dyspnea.
CNS: central nervous system; GSL: granulomatous skin lesions; PLI: parenchymal lung involvement; EN: erythema nodosum; NCI: non-corticosteroid immunosuppressant drugs; IHA: isolated hilar adenopathy.

cantly more frequent in C3 (42.2%) and C5 (29.7%), highlighting a higher prescription of conventional synthetic IS agents in C3 (38.6%), C5 (26.1%) and C1 (23.6%). Biological therapy was significantly more used in C3 (21.7%) and C5 (16.7%) (Table I).

Chronicity

A chronic course of the disease was significantly more common in C3 (31.3%) and C5 (32.6%) compared to C1 (9.1%) and C4 (0%). Regarding the development of irreversible organic damage, C5 showed a significantly higher percentage (19.6%) compared to C1 (3.6%) and C4 (0%) (Table I).

Severity score

The Wasfi severity score was calculated out of a total of 243 patients (Fig. 3). A higher mean value was detected in C3 (2.9±2.1) and C5 (3.0±1.2) compared to the other groups (C1: 1.9±1.6; C2: 2.2±1.7; C4: 1.3±1.2).

Discussion

The cluster analysis has allowed us to identify five different groups, with some clinical characteristics that are differentials between them.

The cluster 1 includes patients with pulmonary involvement exclusively restricted to the hilar nodes, where joint involvement (67.6%) and erythema nodosum (63.6%) also are predominant features. These three manifestations are part of the classic definition of Löfgren's syndrome (LS), so it is not surprising to observe that most patients with this acute presentation of sarcoidosis in our cohort are concentrated in this group (55.8% of the cases of LS of the entire cohort). As expected, the chronicity rate in this cluster was low, however, the same does not occur with the prescription of treatment. Bearing in mind that Löfgren's syndrome usually presents a self-limited course and that there is usually no indication for systemic treatment in mild radiological stages (19-22), it is possible to think that our cohort may have been over treated. In fact, corticosteroid treatment was not included as an important variable in the cluster analysis given our initial suspicion that there were patients treated with glucocorticoids without a correct clinical indication.

The cluster 2 represents a miscellany of extra-pulmonary sarcoidosis, emphasizing granulomatous skin (65.3%) and liver involvement (34.7%); where lung involvement is not clearly predominant (55.1%), unlike the rest of the clusters.

The cluster 3 stands out for encompassing a high percentage of patients who developed ocular and neurological involvement in our cohort. Considering that these organic manifestations are often related to a chronic course and usually require immunosuppressive treatment (23-26), it is not surprising to find higher rates of chronicity (31.3%) and higher use of IS (38.6%) in this group of patients.

The cluster 4 is characterised by exclusively hilar adenopathy involvement, without parenchymal lung or extrathoracic involvement. The low severity of the disease in this cluster justifies the low rate of intensive treatment, without the need for IS drugs in any case, with no progression of the disease to a chronic course in any patient.

Finally, the cluster 5 represents a form of sarcoidosis where lung parenchymal involvement predominates. As expected, the higher prevalence of advanced radiological stages in this group (88.4% presented a Scadding stage ≥ II) is reflected in worse PFT results, specifically FEV1pp and FVCpp values. In addition, a relationship has been observed between the alteration of the respiratory function tests and the Scadding stage, so that the more advanced stages, more frequent are the spirometric abnormalities (1) (27). It is also remarkable the predominance of the male sex in this cluster, unlike the others. In a previous study, we detected a greater tendency of more advanced radiological stages in men, which has also been observed in other series (14, 28, 29). Respiratory symptoms (cough and dyspnea) were more predominant in this cluster, which we could associate with the worse re-

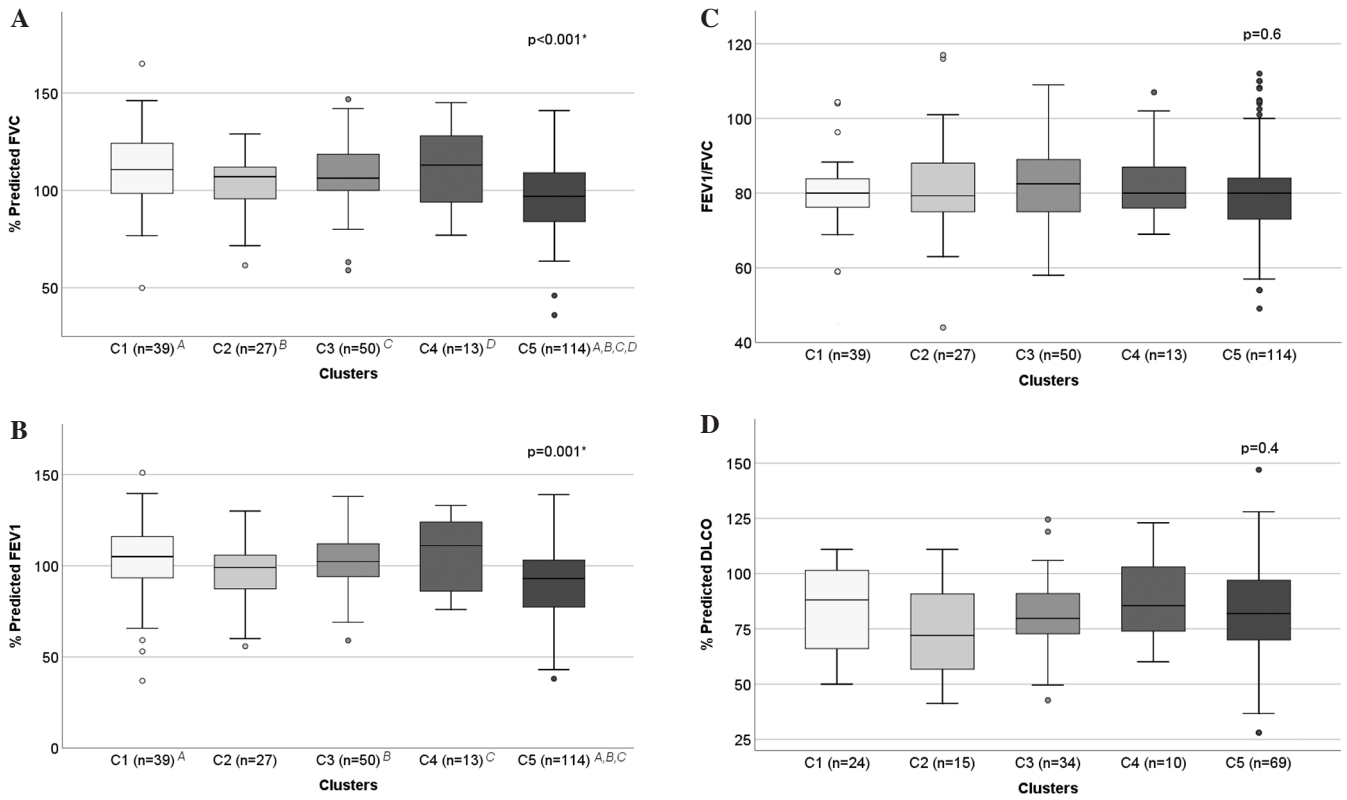

Fig. 2. Comparison of lung function parameters among clusters. Median and IQR are shown by boxplots.
A: Percentage (%) of predicted values of forced vital capacity (FVC).
B: Percentage (%) of predicted values of forced expiratory volume at first second (FEV1).
C: FEV1/FVC ratio.
D: Percentage (%) of predicted values of diffusing lung capacity for carbon monoxide (DLCO).
Statistical test performed: Welch's ANOVA test (*) and 2-sample t-test for pairwise comparisons. Significant differences were detected between clusters sharing the same letter (*A, B, C, D*).

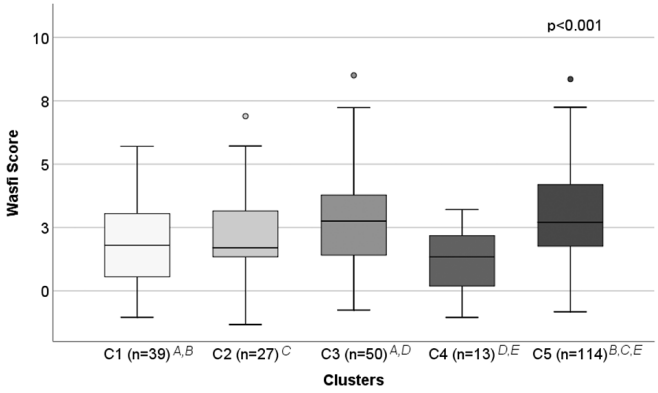

Fig. 3. Comparison of Wasfi Scores among clusters.
Statistical test performed: Welch's ANOVA test (*) and 2-sample t-test for pairwise comparisons. Significant differences were detected between clusters sharing the same letter (*A, B, C, D, E*). Median and IQR are shown by boxplots.

sults in the PFTs and radiological stages observed in the patients of this group. However, a dissociation between the level of dyspnea, PFT results and radiological findings has been reported (30, 31). In addition, lung involvement is not the only condition for dyspnea in sarcoidosis. Pulmonary hypertension, cardiac involvement, anaemia, musculoskeletal involvement, neurological disease or parasarcoidosis syndrome have been related to the development of dyspnea (32).

In recent years, several publications focused on the phenotyping of sarcoidosis have appeared (12, 14, 18, 33-37). Therefore, we have compared the results of some of them with those obtained in our study.

In 2006, Wasfi *et al*. developed a severity score using a multivariate regression model (18). Cardiac and neurological involvement, treatment with non-corticosteroids immunosuppressants, DLCO, FEV1/FVC ratio, and Afro-American race were identified as poor prognostic factors and were included in an equation to calculate a numerical index of severity. Higher values of this score were related to greater severity of the disease. Despite being based on a limited cohort of 104 patients and not having been validated in other populations, the results obtained in our cohort were consistent with those proposed by these authors. Thus, clusters 3 and 5 had significantly higher values of this score, in agreement with the higher rates of chronicity and severe organic involvement observed during follow-up in these groups. The neurological involvement in C3, and the worst results in the PFTs in C5, were the main determinants of this result in this severity score.

In GenPhenReSa, a multicentre European study published in 2018, five phenotypic groups were identified in a

cohort of 2163 patients: 1) abdominal, 2) ocular-cardiac-cutaneous-CNS, 3) musculoskeletal-cutaneous, 4) pulmonary and lymphonodal, and 5) extrapulmonary involvement (33). Although the large number of collected patients represents one of its main strengths, we can find limitations in a possible restricted multidisciplinary nature in the clinical characterisation of the patients, since Respiratory Departments were mainly involved; the heterogeneity of the diagnosis and management of the disease inherent to the international multicenter nature of the study; and the ethnicity homogeneity of the cohort (all patients were Caucasian), which makes it difficult to extrapolate the results to other populations. We can identify similarities between the clusters of the GenPhenReSa study and those of our cohort. The subgroup called "abdominal" and "extrapulmonary" could be included in our C2 phenotype, where hepatic involvement and other extrapulmonary manifestations are predominant. The ocular-cardiac-cutaneous-CNS, could correspond to our C3 phenotype, considering the ocular and CNS involvement. The presentation "musculoskeletal-cutaneous" would correspond to our C1 cluster, covering most patients with LS. Finally, the cluster called "pulmonary-lymphonodal" would be more difficult to identify with our clustering, due to the lack of data on radiological staging. Still, we could relate this group to C4 and C5, two groups that showed clear differences in the degree of pulmonary involvement and in the course of the disease in our cohort. Unlike our study, genetics was used to define different phenotypes of sarcoidosis, but not to predict the chronicity of these groups.

In 2020, Rubio-Rivas *et al*. developed a study to identify phenotypes and predict the chronicity of sarcoidosis in a Spanish cohort of 694 patients collected over four decades (12). After performing a cluster analysis with 26 variables, they describe six different phenotypes. The high percentage of patients with Löfgren's syndrome (44.8%) compared to our sample (12.6%) stands out noticeably. In fact, these authors divide their patients as those with LS into different phenotypes (C1, C2, and C3) and those with pulmonary and/or extrapulmonary involvement (C4, C5 and C6). We found similarities between these phenotypes and ours. In the first place, although we did not identify a pure form of LS, our C1 encompasses a large part of the cases of this variant in our cohort, being comparable with the "Non-febrile LS with periarticular ankle inflammation" subgroup. Regarding lung involvement, our C5 could be partially equated to the "Pure pulmonary sarcoidosis" subgroup. Although in our series there is associated extrapulmonary involvement, lung parenchyma (Scadding stage ≥II) is always involved in both cases. The rest of the clusters are more difficult to match. Our C2 clustering could be included in the so-called "stage I miscellaneous sarcoidosis", given the variability of extrapulmonary manifestations shown in both subgroups. Our C3 phenotype presents similarities with the group "abdominal sarcoidosis with pulmonary sarcoidosis", being in both cases the clusters with the highest percentage of ocular and neurological involvement. Finally, C4 (exclusive hilar adenopathy) could be part of the three types of LS identified by the Rubio-Rivas group. These similarities are also shared in terms of chronicity, such that those clusters related to LS in both cohorts presented lower chronicity rates than the rest of groups.

In a more recent study conducted at National Jewish Health (NJH) in 2022, six phenotypic subgroups were also proposed. This American cohort includes 554 patients, mainly Caucasians (81.7%), although with a higher percentage of blacks (16.6%) than the previous studies mentioned(14). As mentioned, six pulmonary phenotypes were identified by means of a cluster analysis, with no contribution of extrapulmonary involvement, unlike our study. These phenotypes were classified according to the pulmonary involvement stage and respiratory function tests. The most severe phenotypes present worse results in the PFTs, higher values in the Wasfi severity score and a higher rate of treatment prescription. Although we cannot make a clear cluster correlation between both studies, we observe that the above coincides with our C5 cluster (parenchymal lung involvement) if we compare it with the rest of the phenotypes.

Considering the gender distribution of the clusters, the female predominance stands out in C1 and C2, where the high proportion of patients who developed Löfgren syndrome (C1) and skin manifestations (C2) stands out. This coincides with several articles that have shown a higher frequency of these clinical forms of sarcoidosis in women (38-40). In contrast, cluster 5 includes a higher percentage of men than the other groups. Taking into account that patients with more advanced radiological stages predominate in this group, it is possible to think that men have a greater tendency to develop more severe forms of pulmonary sarcoidosis. In fact, in several cohorts it has been observed that women more often presented with stage 0–1 compared to men (41-43). The inverse relationship reported between the radiological stage and spontaneous resolution, together with the acute course that usually characterises Löfgren's syndrome, may determine the differences in the chronicity rates observed between clusters 1 and 4 with respect to cluster 5 (6, 44). At the same time, it allows us to associate male sex as a risk factor for chronicity, as has already been described (45).

Our study has several strengths. It has been carried out with a large cohort of more than 300 patients diagnosed and followed for more than 20 years by a multidisciplinary team including multiple specialties (rheumatology, internal medicine, radiology, pathology, dermatology, ophthalmology, nephrology, cardiology, and neurology). Only patients with certain diagnosis, according the ATS/ERS/WASOG criteria, have been included. In addition, our hospital is the only reference hospital in the entire region, so it includes all diagnosed cases of sarcoidosis, avoiding biases such as the overrepresentation of complicated sarcoidosis.

The major limitations of this study are related to its retrospective nature. The absence of a standardised protocol for the diagnosis and management of these

patients leads to subjective variability in the data collected, such as symptoms or indication for treatment, as well as the lack of some variables that should have been included in the cluster analysis. In contrast, the Caucasian homogeneity of our cohort prevents the extrapolation of our results to populations with different ethnic composition.

Even though we identified five groups in the cluster analysis, we observed that these are not completely homogeneous. There is an uneven distribution of certain characteristics that could make it difficult to categorise certain patients into a specific group. In this way, a patient with Löfgren's syndrome could belong to both cluster 1 and cluster 5. While the first group was associated with lower rates of chronicity and severity and, therefore, it correlates with the benign course that this form of sarcoidosis usually develops; cluster 5 presented a worse prognosis of the disease. Therefore, we are aware that the classification we propose should be used with caution and for future directions, a larger cohort followed longitudinally would allow a better cluster analysis for the identification of more homogeneous groups.

Truly, in a disease as heterogeneous and complex as sarcoidosis, any attempt to stratify and differentiate groups of patients with similar characteristics is very stimulating to deepen the pathophysiological knowledge of this entity, try to establish genetic differences between the different subgroups, assess appropriate treatments to the profile of every subgroup and even plan specific clinical trials aimed at the groups or profiles at highest risk and with the worst prognosis. Therefore, we believe that cluster analysis could improve the identification and categorisation of patients with sarcoidosis to optimise the treatment of this disease and improve its prognosis.

Conclusions

In conclusion, the present study allowed us to identify five clinical phenotypes of sarcoidosis in a cohort of 342 patients with variable organic involvement, using an objective method such as cluster analysis. In a disease with such enormous clinical variability, we were able to distinguish subgroups of patients who share clinical features at diagnosis as well as throughout the course of the disease. In this way, two subgroups presented a more self-limited course (C1 and C4), while the remaining three (C2, C3 and C5) developed a chronic disease more frequently, being also more prone to the prescription of IS treatment. The early identification of these phenotypes could make it possible to prevent the clinical course of the disease, facilitating its correct management from diagnosis. However, the retrospective nature of this study prevents obtaining better results in clustering, by limiting the number of variables and their control. For this reason, this study is a starting point for developing future projects for a better clinical categorisation of this so heterogeneous disease and not for immediate use in clinical practice.

References

1. 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 ER. *Am J Respir Crit Care Med* 1999; 160(2): 736-55. https://doi.org/10.1164/ajrccm.160.2.ats4-99
2. BAUGHMAN RP, TEIRSTEIN AS, JUDSON MA *et al*.: Clinical characteristics of patients in a case control study of sarcoidosis. *Am J Respir Crit Care Med* 2001; 164(10 Pt 1): 1885-89. https://doi.org/10.1164/ajrccm.164.10.2104046
3. 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 Vasc Diffus Lung Dis* 2012; 29(2): 119-27.
4. LÖFGREN S: Primary pulmonary sarcoidosis. 1: II. Clinical course and prognosis. *Acta Med Scand* 1953; 145(6): 465-74.
5. MANA J, GOMEZ-VAQUERO C, MONTERO A *et al*.: Lofgren's syndrome revisited: a study of 186 patients. *Am J Med* 1999; 107(3): 240-45. https://doi.org/10.1016/s0002-9343(99)00223-5
6. BROWN F, TANNER LS: Lofgren syndrome. In Treasure Island (FL); 2020.
7. SCHUPP JC, FREITAG-WOLF S, BARGAGLI E *et al*.: Phenotypes of organ involvement in sarcoidosis. *Eur Respir J* 2018; 51(1): 1-11. https://doi.org/10.1183/13993003.00991-2017
8. PEREIRA CAC, DORNFELD MC, BAUGHMAN R, JUDSON MA: Clinical phenotypes in sarcoidosis. *Curr Opin Pulm Med* 2014; 20(5): 496-502. https://doi.org/10.1097/mcp.0000000000000077
9. ISRAEL HL, KARLIN P, MENDUKE H, DELISSER OG: Factors affecting outcome of sarcoidosis. Influence of race, extrathoracic involvement, and initial radiologic lung lesions. *Ann N Y Acad Sci* 1986; 465(1): 609-18. https://doi.org/10.1111/j.1749-6632.1986.tb18537.x
10. COZIER YC, COOGAN PF, GOVENDER P, BERMAN JS, PALMER JR, ROSENBERG L: Obesity and weight gain in relation to incidence of sarcoidosis in US black women: data from the Black Women's Health Study. *Chest* 2015; 147(4): 1086-93. https://doi.org/10.1378/chest.14-1099
11. LOPES MC, AMADEU TP, RIBEIRO-ALVES M *et al*.: Defining prognosis in sarcoidosis. *Medicine* (Baltimore) 2020; 99(48): e23100. https://doi.org/10.1097/md.0000000000023100
12. RUBIO-RIVAS M, CORBELLA X: Clinical phenotypes and prediction of chronicity in sarcoidosis using cluster analysis in a prospective cohort of 694 patients. *Eur J Intern Med* 2020; 77: 59-65. https://doi.org/10.1016/j.ejim.2020.04.024
13. LHOTE R, ANNESI-MAESANO I, NUNES H *et al*.: Clinical phenotypes of extrapulmonary sarcoidosis: an analysis of a French, multiethnic, multicentre cohort. *Eur Respir J* 2021; 57(4): 2001160. https://doi.org/10.1183/13993003.01160-2020
14. LIN NW, ARBET J, MROZ MM *et al*.: Clinical phenotyping in sarcoidosis using cluster analysis. *Respir Res* 2022; 23(1): 88. https://doi.org/10.1186/s12931-022-01993-z
15. COSTABEL U, HUNNINGHAKE GW; on behalf of THE SARCOIDOSIS STATEMENT COMMITTEE: ATS/ERS/WASOG statement on sarcoidosis. *Eur Respir J* 1999; 14(4): 735. https://doi.org/10.1034/j.1399-3003.1999.14d02.x
16. JUDSON MA, COSTABEL U, DRENT M *et al*.: The WASOG Sarcoidosis Organ Assessment Instrument: an update of a previous clinical tool. *Sarcoidosis Vasc Diffus Lung Dis* 2014; 31(1): 19-27.
17. NUNES H, BRILLET P-Y, VALEYRE D, BRAUNER MW, WELLS AU: Imaging in sarcoidosis. *Semin Respir Crit Care Med* 2007; 28(1): 102-20. https://doi.org/10.1055/s-2007-970336
18. WASFI YS, ROSE CS, MURPHY JR *et al*.: A new tool to assess sarcoidosis severity. *Chest* 2006; 129(5): 1234-45. https://doi.org/10.1378/chest.129.5.1234
19. KARAKAYA B, KAISER Y, VAN MOORSEL C, GRUNEWALD J: Löfgren's syndrome: diagnosis, management, and disease pathogenesis. *Semin Respir Crit Care Med* 2017; 38(4): 463-76. https://doi.org/10.1055/s-0037-1602380
20. CASTRO MDC, PEREIRA CAC: Nonlife-threatening sarcoidosis. *Semin Respir Crit Care Med* 2020; 41(5): 733-40. https://doi.org/10.1055/s-0040-1710371
21. NUNES H, JENY F, BOUVRY D, UZUNHAN Y, VALEYRE D: Indications for treatment of sarcoidosis. *Curr Opin Pulm Med* 2019; 25(5): 505-18. https://doi.org/10.1097/mcp.0000000000000604
22. BAUGHMAN RP, VALEYRE D, KORSTEN P *et al*.: ERS clinical practice guidelines on treatment of sarcoidosis. *Eur Respir J* 2021; 58(6): 2004079. https://doi.org/10.1183/13993003.04079-2020
23. BAUGHMAN R, LOWER E, KAUFMAN A: Oc-

ular sarcoidosis. *Semin Respir Crit Care Med* 2010; 31(04): 452-62. https://doi.org/10.1055/s-0030-1262213
24. YANG SJ, SALEK S, ROSENBAUM JT: Ocular sarcoidosis. *Curr Opin Pulm Med* 2017; 23(5): 458-67. https://doi.org/10.1097/mcp.0000000000000409
25. BRADSHAW MJ, PAWATE S, KOTH LL, CHO TA, GELFAND JM: Neurosarcoidosis. *Neurol Neuroimmunol Neuroinflamm* 2021; 8(6): e1084. https://doi.org/10.1212/nxi.0000000000001084
26. VOORTMAN M, DRENT M, BAUGHMAN RP: Management of neurosarcoidosis: a clinical challenge. *Curr Opin Neurol* 2019; 32(3): 475-83. https://doi.org/10.1097/wco.0000000000000684
27. HUNNINGHAKE GW, GILBERT S, PUERINGER R *et al*.: Outcome of the treatment for sarcoidosis. *Am J Respir Crit Care Med* 1994; 149(4): 893-98. https://doi.org/10.1164/ajrccm.149.4.8143052
28. FERNÁNDEZ-RAMÓN R, GAITÁN-VALDIZÁN JJ, SÁNCHEZ-BILBAO L *et al*.: Epidemiology of sarcoidosis in northern Spain, 1999-2019: a population-based study. *Eur J Intern Med* 2021; 91: 63-69. https://doi.org/10.1016/j.ejim.2021.05.011
29. LUNDKVIST A, KULLBERG S, ARKEMA EV *et al*.: Differences in disease presentation between men and women with sarcoidosis: a cohort study. *Respir Med* 2022; 191: 106688. https://doi.org/10.1016/j.rmed.2021.106688
30. JUDSON MA, PRESTON S, HU K *et al*.: Quantifying the relationship between symptoms at presentation and the prognosis of sarcoidosis. *Respir Med* 2019; 152: 14-19. https://doi.org/10.1016/j.rmed.2019.03.012
31. BAYDUR A, ALSALEK M, LOUIE SG, SHARMA OP: Respiratory muscle strength, lung function, and dyspnea in patients with sarcoidosis. *Chest* 2001; 120(1): 102-8. https://doi.org/10.1378/chest.120.1.102
32. DE BOER S, KOLBE J, WILSHER ML: The relationships among dyspnoea, health-related quality of life and psychological factors in sarcoidosis. *Respirology* 2014;19(7): 1019-24. https://doi.org/10.1111/resp.12359
33. SCHUPP JC, FREITAG-WOLF S, BARGAGLI E *et al*.: Phenotypes of organ involvement in sarcoidosis. *Eur Respir J* 2018; 51(1): 1700991. https://doi.org/10.1183/13993003.00991-2017
34. PRASSE A, KATIC C, GERMANN M, BUCHWALD A, ZISSEL G, MÜLLER-QUERNHEIM J: Phenotyping sarcoidosis from a pulmonary perspective. *Am J Respir Crit Care Med* 2008; 177(3): 330-36. https://doi.org/10.1164/rccm.200705-742oc
35. BAUGHMAN RP, NAGAI S, BALTER M *et al*.: Defining the clinical outcome status (COS) in sarcoidosis: results of WASOG Task Force. *Sarcoidosis Vasc Diffus Lung Dis* 2011; 28(1): 56-64.
36. RODRIGUES SCS, ROCHA NAS, LIMA MS *et al*.: Factor analysis of sarcoidosis phenotypes at two referral centers in Brazil. *Sarcoidosis Vasc Diffus Lung Dis* 2011; 28(1): 34-43.
37. WALSH SL, WELLS AU, SVERZELLATI N *et al*.: An integrated clinicoradiological staging system for pulmonary sarcoidosis: a case-cohort study. *Lancet Respir Med* 2014; 2(2): 123-30. https://doi.org/10.1016/S2213-2600(13)70276-5
38. MORIMOTO T, AZUMA A, ABE S *et al*.: Epidemiology of sarcoidosis in Japan. *Eur Respir J* 2008; 31(2): 372-79. https://doi.org/10.1183/09031936.00075307
39. YANARDAĞ H, PAMUK ON, KARAYEL T: Cutaneous involvement in sarcoidosis: analysis of the features in 170 patients. *Respir Med* 2003; 97(8): 978-82. https://doi.org/10.1016/s0954-6111(03)00127-6
40. NEWMAN LS, ROSE CS, MAIER LA: Sarcoidosis. *N Engl J Med* 1997; 336(17): 1224-34. https://doi.org/10.1056/nejm199704243361706
41. TE HS, PERLMAN DM, SHENOY C *et al*.: Clinical characteristics and organ system involvement in sarcoidosis: comparison of the University of Minnesota Cohort with other cohorts. *BMC Pulm Med* 2020; 20(1). https://doi.org/10.1186/s12890-020-01191-x
42. BRITO-ZERON P, SELLARES J, BOSCH X *et al*.: Epidemiologic patterns of disease expression in sarcoidosis: age, gender and ethnicity-related differences. *Clin Exp Rheumatol* 2016; 34: 380-88.
43. LUNDKVIST A, KULLBERG S, ARKEMA EV *et al*.: Differences in disease presentation between men and women with sarcoidosis: a cohort study. *Respir Med* 2022; 191: 106688. https://doi.org/10.1016/j.rmed.2021.106688
44. SOTO-GOMEZ N, PETERS JI, NAMBIAR AM: Diagnosis and management of sarcoidosis. *Am Fam Physician* 2016; 93(10): 840-48.
45. MAÑÁ J, RUBIO-RIVAS M, VILLALBA N *et al*.: Multidisciplinary approach and long-term follow-up in a series of 640 consecutive patients with sarcoidosis: Cohort study of a 40-year clinical experience at a tertiary referral center in Barcelona, Spain. *Medicine* (Baltimore) 2017; 96(29): e7595. https://doi.org/10.1097/md.0000000000007595