Review

Geoepidemiological big data approach to sarcoidosis: geographical and ethnic determinants

P. Brito-Zerón[1,2], B. Kostov[2,3], D. Superville[4,5], R.P. Baughman[6], M. Ramos-Casals[2], on behalf of the Autoimmune Big Data Study Group.

[1]*Autoimmune Diseases Unit, Department of Medicine, Hospital CIMA-Sanitas, Barcelona, Spain;*
[2]*Laboratory of Autoimmune Diseases Josep Font, IDIBAPS-CELLEX, Department of Autoimmune Diseases, ICMiD, Hospital Clínic, University of Barcelona, Spain;*
[3]*Primary Healthcare Transversal Research Group, IDIBAPS, Barcelona, Spain;*
[4]*Massachusetts Institute of Technology (MIT), Cambridge, MA, USA;*
[5]*University of California, San Francisco, CA, USA.*
[6]*University of Cincinnati Medical Center, Department of Medicine, Cincinnati, USA.*

Pilar Brito-Zerón, MD, PhD
Belchin Kostov, PhD
Daphne Superville, BS
Robert P. Baughman, MD
Manuel Ramos-Casals, MD, PhD

Please address correspondence to:
Dr Manuel Ramos-Casals,
Servei de Malalties Autoimmunes, Hospital Clínic, C/Villarroel 170, 08036-Barcelona, Spain.
E-mail: mramos@clinic.cat

Received on November 26, 2018; accepted in revised form on February 11, 2019.

Clin Exp Rheumatol 2019; 37; 1052-1064.

Key words: sarcoidosis, ethnicity, epidemiology, geographical, phenotype

Competing interests: R.P. Baughman has received grant funding for sarcoidosis studies from Genetech, Gilead, Bayer, Mallinckrodt, aTyr, and Actelion; he has been a consultant for Actelion and Mallinckrodt, and is on the speaker bureau for Mallinckrodt.
The other co-authors have declared no competing interests.

ABSTRACT

In sarcoidosis, a rare multiorgan disease of unknown aetiology characterised by non-caseating epitheloid cell granulomas, three geoepidemiological factors are major aetiopathogenic factors: geolocation, ethnicity, and personal environment. Geographically, sarcoidosis is mainly reported in the Northern Hemisphere, with the highest incidence rates uniformly reported in countries located at the highest latitudes. The main geoepidemiological-driven differences across the world are of greater female involvement in Southern Europe, the Southern US and Japan, a differentiated radiological pattern (predominance of stage I in Southern Europe and Middle East/Asia and of stage II in Northern Europe, China and India, with the US and Japan having the highest frequencies of stages III/IV) and the extrathoracic phenotype: the most frequent extrathoracic organs involved are the skin in Southern Europe and Middle East/Asia, the eyes in Northern Europe, Northeast US and Japan, the liver in India and the lymph nodes in China. In addition, there are large ethnicity-driven variations in the frequency, epidemiology, clinical expression and outcome of sarcoidosis. The highest incidence rates are uniformly reported in Black/African-American people, independently of the geographical location, with rates between 2- and 10-fold higher than those reported in White people living in the same geographical area. Furthermore, ethnicity heavily influences the clinical phenotype by modifying the age at diagnosis and the rates of thoracic and extrathoracic involvements. Geoepidemiological studies enhanced by big data may yield important clues to understanding the role of these factors in the frequency and clinical phenotypes of sarcoidosis.

Introduction

Sarcoidosis, a systemic disease of unknown aetiology characterised by non-caseating epitheloid cell granulomas occurs predominantly in persons aged 20–50 years (1). Sarcoidosis has a heterogeneous clinical presentation, although thoracic involvement (hilar adenopathies and/or interstitial lung disease) occurs in >90% of cases (2). Epidemiologically, sarcoidosis is rare, with an estimated incidence in Europe ranging between 1 and 19 cases per 100,000 persons/year (3). The potential influence of epidemiological determinants on the frequency and phenotypic expression of sarcoidosis has been extensively studied, and three major epidemiological factors have been identified: geolocation, ethnicity, and the personal environment. Ethnicity significantly influences sarcoidosis, as highlighted by US studies showing it is more common and severe in Black/African Americans (BAA), with incidence rates as high as 40 cases per 100,000 persons (3). Geolocation also plays a key role in understanding the epidemiology of sarcoidosis (4). Its rarity means that the larger the population analysed, the better the characterisation of the influence of geoepidemiological and ethnic factors in the phenotypical disease expression, and the more likely the findings will be close to the real population. Big data-driven research may help provide a high-definition picture of infrequent, heterogeneous diseases such as sarcoidosis (5). Using a literature search combining the Google search engine with PubMed (see the Methods section), we selected studies with ≥100 patients with sarcoidosis (Supplementary Table 1, Suppl. Fig. 1). This provided 128,955 patients whose features were merged to provide a

worldwide picture that might be used as a standard reference for geoepidemiological comparisons (Suppl. Table II). This review provides an update on the presentation of sarcoidosis across geographic regions and ethnic groups and how clinical phenotypes are modulated by geoepidemiological and environmental factors.

Methods

Methodological approach

This review used the potential of combining the Google search engine with PubMed to collect large series (>100 patients) of sarcoidosis reported in the PubMed library in the last 20 years. We made a text-word search in Google (www.google.com) between 1st and 31st October, 2017 using the methodology previously described by our group (5).

Studies were selected according to the following inclusion criteria: description of ≥100 patients, available data for ≥2 of the main geoepidemiological variables (age, gender, ethnicity, city/country) and diagnosis of sarcoidosis according to one of the following classifications:

- American Thoracic Society/European Respiratory Society/World Association of Sarcoidosis and Other Granulomatous Disorders (WASOG) statement on sarcoidosis (6)
- Tissue biopsy with histologic evidence of non-caseating granulomas in the absence of other causes of granulomatous disease;
- Clinical and radiological findings consistent with sarcoidosis (overwhelmingly clinical presentations such as Löfgren syndrome) in the absence of biopsy;
- Administrative codification for sarcoidosis according to the International Classification of Diseases (ICD) of the World Health Organisation codes.

Reviews, meta-analyses and case series were excluded. Two reviewers (PBC, DS) independently examined the entries retrieved by the Google search for potential eligibility. Studies selected for possible inclusion by either reviewer underwent dual, independent, review in PubMed. If reviewers disagreed, conflicts were resolved by consensus. A data extraction form was developed by PBZ prior to manuscript review to gather relevant data from each article. All data extractions were reviewed for completeness and accuracy by the remaining authors. No restrictions were placed on language or type of publication. Reviewers manually searched reference lists of selected articles for relevant citations missed by the search. Identification of duplicated cohorts was managed by contrasting the following variables between studies: name of authors, participant centres, period of patient recruitment, name of database/multicentre group number of patients included, epidemiological features, and patterns of organ involvement. When a selected study referenced other clinical studies carried out by the same research group, we evaluated the information offered by these referenced manuscripts and selected only one study among the various studies reported by the group (those containing the most detailed information).

Definition of variables collected

Variables collected from each study included:

- number of patients (included, excluded and evaluated);
- data sources (administrative databases used primarily for purposes other than medical care, or medical databases collected specifically for clinical purposes);
- type of database (population-based, including studies involving a defined general population from a specific geographical area, or patient-based, including inpatient/hospital-based and/or outpatient/primary care databases);
- city and country of the study;
- inclusion and exclusion criteria;
- gender distribution;
- mean age at diagnosis (time of fulfilment of classification criteria/clinical/histopathological diagnosis);
- mean age at study (current age);
- ethnic definitions based on the FDA classification;

a) American Indian (AI) or Alaska Native. A person having origins in any of the original peoples of North and South America (including Central America).

b (Asian (A). A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent, including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.

c) Black or African American (AA). A person having origins in any of the black racial groups of Africa.

d) Hispanic (H) or Latino. A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race.

e) Native Hawaiian or Other Pacific Islander (PI). A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.

f) White (W). A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.

- radiographic stages classified as stage 0 (normal), stage I (bilateral hilar lymphadenopathy [BHL] without pulmonary infiltrates), stage II (BHL plus pulmonary infiltrates), stage III (parenchymal infiltrates without BHL) and stage IV (extensive fibrosis with distortion or bullae);
- extrathoracic organ involvement classified according to the 2014 WASOG organ assessment instrument (7). Clinical scenarios are classified into four categories (highly probable, at least probable, possible and indeterminate). For all the 16 organs evaluated, the experts reached consensus that organ involvements classified as "highly/at least probable" represent sarcoidosis;
- sarcoidosis-specific therapies: use of corticosteroids, immunosuppressive agents and/or biologics;
- follow-up (in months) and death.

Statistical analysis

Descriptive data were presented as means and range for continuous variables and numbers and percentages for categorical variables. The Spearman correlation coefficients were computed to study the influence of the ethnicity on the phenotypic expression of sarcoidosis. The number of cohorts with

available data that were considered to compute correlations was detailed as well as the number of patients included. The correlation coefficients were computed considering the percentages of different ethnicity groups and the percentages of phenotypic expression covariates, except for age at diagnosis and gender ratio where mean values were considered. Data visualisation techniques were used to summarise information from the aggregated data (8). Choropleth Map was used to visualise variation of Sarcoidosis across the countries. Bar charts were used to compare thoracic and extrathoracic phenotypic expression of sarcoidosis between studies carried out in Northern

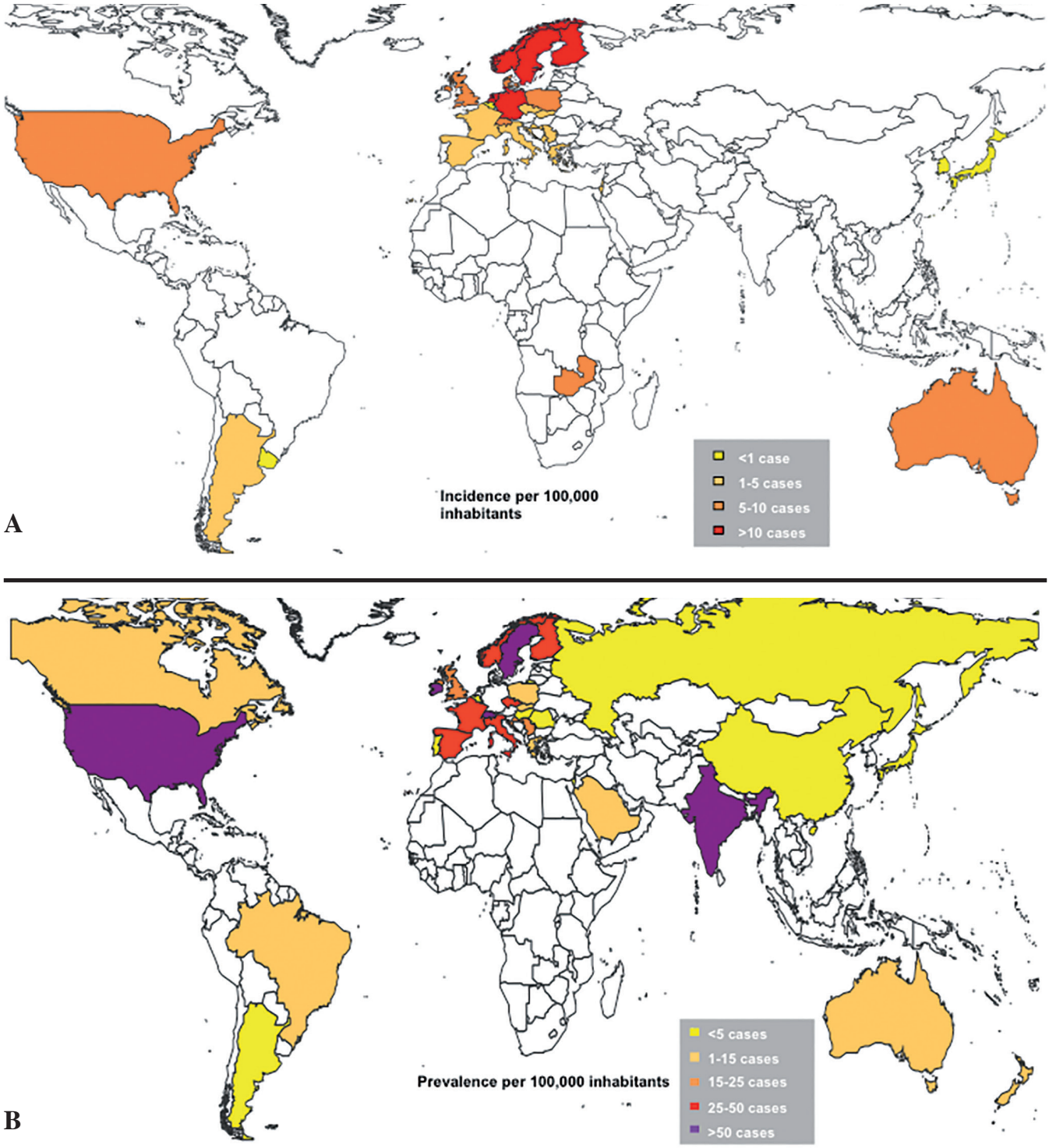

Fig. 1. Incidence (Fig. 1a) and prevalence (Fig. 1b) of sarcoidosis country-by-country. Data obtained from the main epidemiological studies published between 1958 and 2017; wide heterogeneity in study designs was observed, with a predominant use of mass x-ray screening surveys in studies reported in the 1960s and 70s, and nationwide-based surveys in more recent studies. In countries with information from more than one study, a mean incidence was calculated.

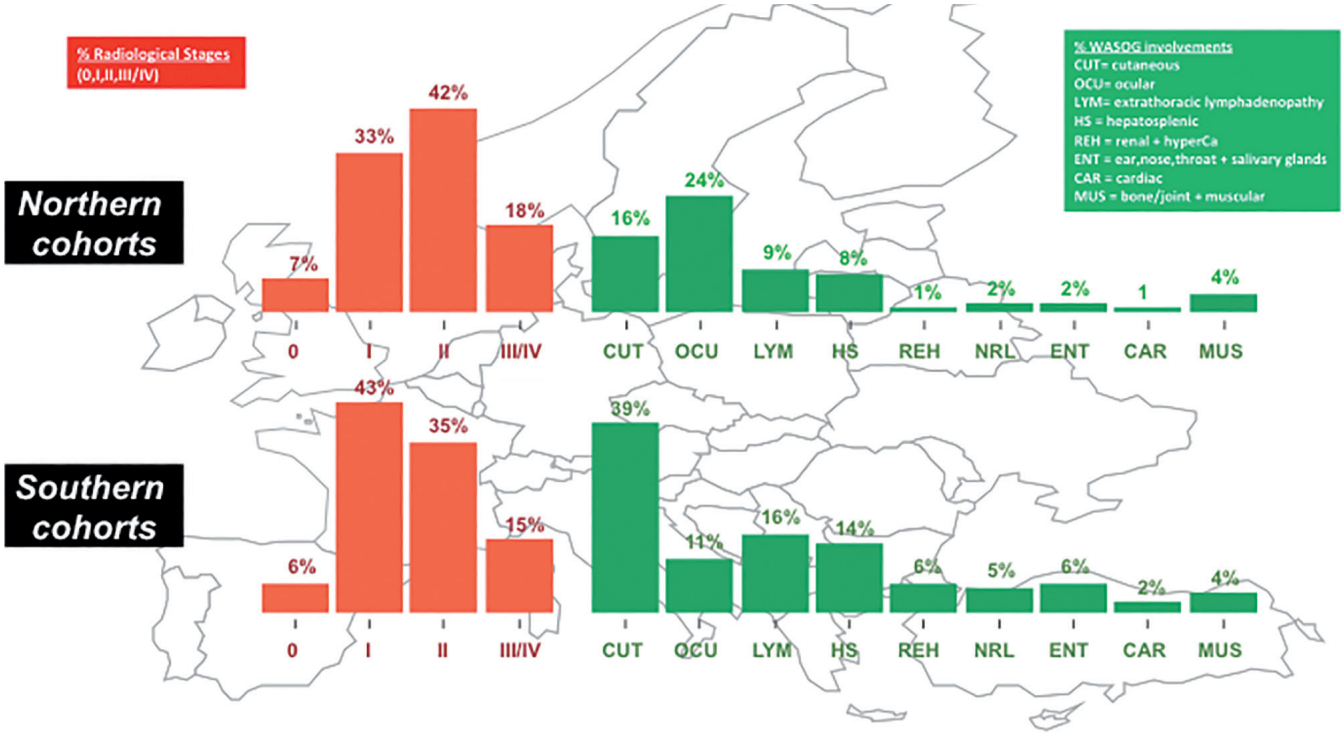

Fig. 2. Thoracic and extrathoracic phenotypic expression of sarcoidosis in Europe: comparison between studies in northern and southern cohorts. Coloured columns represent the figures obtained in each geographical area.

and Southern cohorts. Polar area charts were constructed to represent the association between ethnicity and extrathoracic WASOG organ-by-organ involvements. Pie charts were used to show differences between ethnic groups according to the radiological patterns. All significance tests were two tailed, and values of $p<0.05$ were considered significant. All analyses were conducted using the R v. 3.2.3 for Windows statistical software package.

Geographical determinants

Worldwide distribution

Our worldwide merged data shows that ≤1% of the nearly 130,000 patients collected are reported from Southern Hemisphere countries, a north-south ratio of 176:1, while the north-south ratio of people living in the two hemispheres is only 9:1; most series are reported from the US, Europe and Japan. The highest incidence rates (>10 new cases diagnosed yearly per 100,000 persons) are uniformly reported in countries located at the highest latitudes, *i.e.* Northern European countries (3, 9-12) (Fig. 1a). Geolocation plays a key role in understanding the epidemiology of autoimmune diseases. Worldwide, the preva-

Table I. Merged data for thoracic (Scadding radiological stages) and extrathoracic (WASOG classification) phenotypic expression of sarcoidosis in Europe: comparison between studies carried out in northern European and southern European cohorts.

	Northern European cohorts (n=71566)	Southern European cohorts (n=5902)
Mean age at diagnosis (years)	41.1 (36.4-51.2)	42.8 (39.8-47.3)
Gender (% of women)	33451/68239 (49)	3938/5901 (66.7)
Thoracic involvement (%)		
Radiological stage 0	302/4571 (6.6)	188/2951 (6.4)
Radiological stage I	1510/4571 (33)	1271/2951 (43.1)
Radiological stage II	1922/4571 (42)	1041/2951 (35.3)
Radiological stage III	634/4571 (13.9)	379/2951 (12.8)
Radiological stage IV	203/4571 (4.4)	72/2951 (2.4)
Extrathoracic involvement (%)		
Cutaneous involvement	610/3695 (16.5)	1126/2889 (39)
Ocular involvement	832/3465 (24)	294/2724 (10.8)
Extrathoracic lymph node involvement	300/3407 (8.8)	314/2027 (15.5)
Cardiac involvement	31/2909 (1.1)	45/2208 (2)
Liver involvement	129/3465 (3.7)	203/2208 (9.2)
Calcium/Vitamin D	15/409 (3.7)	76/1449 (5.2)
Splenic involvement	216/1940 (11.1)	107/2129 (5)
Neurological involvement	60/2704 (2.2)	129/2373 (5.4)
Salivary gland involvement	30/2061 (1.5)	96/2464 (3.9)
ENT involvement	7/207 (3.4)	43/1477 (2.9)
Bone/joint involvement	97/2704 (3.6)	67/1568 (4.3)
Renal involvement	6/2209 (0.3)	26/1477 (1.8)
Muscular involvement	3/1800 (0.2)	10/1477 (0.7)

Values are represented as means and range for continuous variables and numbers and percentages for categorical variables.
Variables are not detailed in all cases, and the prevalence of a specific feature has been stated as the number of cases with the feature (numerator)/number of cases in which the feature was detailed (denominator).
All comparisons were statistically significant ($p<0.05$) except for mean age at diagnosis ($p=0.185$), calcium/vitamin D ($p=0.240$), ENT involvement ($p=0.877$) and bone/joint involvement ($p=0.298$).

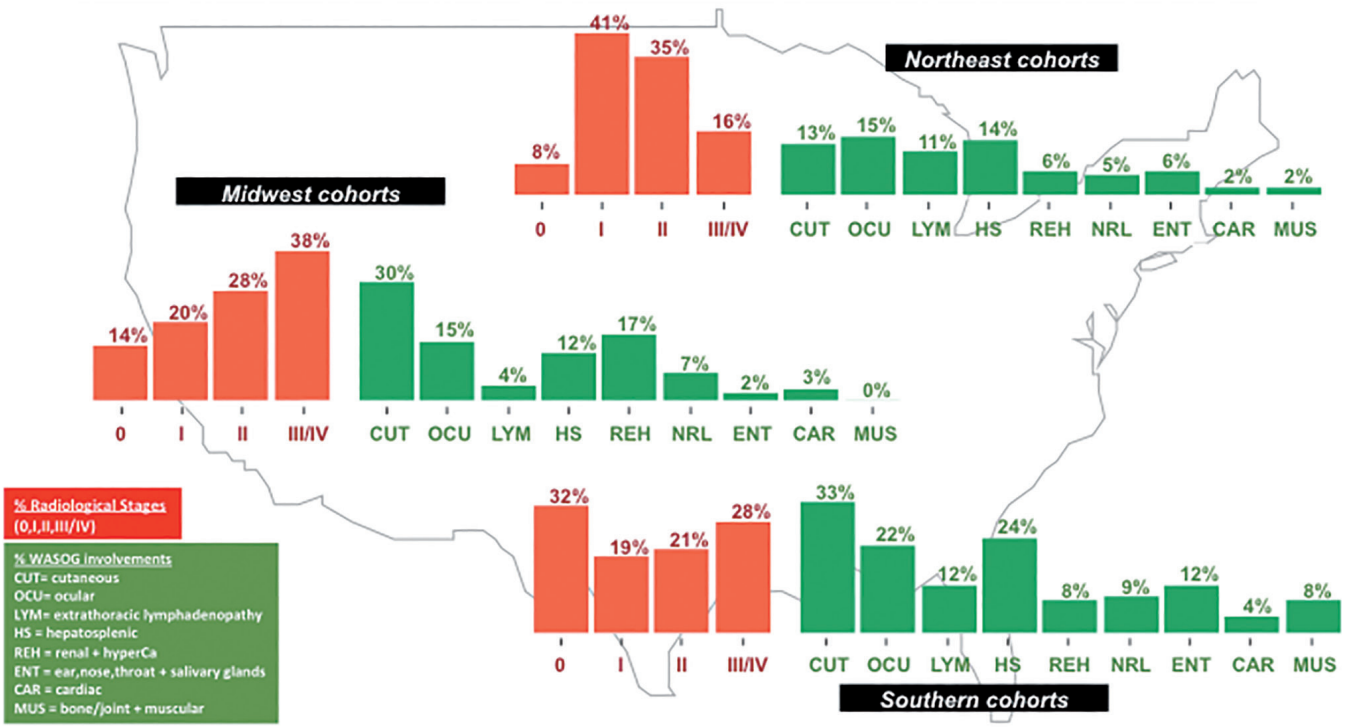

Fig. 3. Thoracic and extrathoracic phenotypic expression of sarcoidosis in the US: comparison between studies in the Midwest, Northeast and and southern cohorts. Coloured columns represent the figures obtained in each geographical area.

Table II. Merged data for thoracic (Scadding radiological stages) and extrathoracic (WASOG classification) phenotypic expression of sarcoidosis in US cohorts: comparison between studies carried out in Midwest, Northeast and South/West US.

	Midwest US cohorts (n=2510)	Northeast US cohorts (n=2296)	South/West US cohorts (n=2457)
Mean age at diagnosis (years)	39 (33.7-44.2)	38.9 (35.8-41.9)	41.5 (40.9-42)
Gender (% of women)	1571/2510 (62.6)	1381/2296 (60.1)	1344/1983 (67.8)
Thoracic involvement (%)			
Radiological stage 0	150/1096 (13.7)	93/1218 (7.6)	348/1083 (32.1)
Radiological stage I	222/1096 (20.3)	504/1218 (41.4)	208/1083 (19.2)
Radiological stage II	309/1096 (28.2)	431/1218 (35.4)	223/1083 (20.6)
Radiological stage III	233/1096 (21.3)	147/1218 (12.1)	114/1083 (10.5)
Radiological stage IV	182/1096 (16.6)	43/1218 (3.5)	190/1083 (17.5)
Extrathoracic involvement (%)			
Cutaneous involvement	392/1315 (29.8)	156/1180 (13.2)	614/1842 (33.3)
Ocular involvement	117/797 (14.7)	173/1180 (14.7)	426/1951 (21.8)
Extrathoracic lymph node involvement	12/345 (3.5)	135/1180 (11.4)	177/1508 (11.7)
Cardiac involvement	39/1315 (3)	27/1180 (2.3)	56/1248 (4.5)
Liver involvement	121/1315 (9.2)	104/1180 (8.8)	355/1842 (19.3)
Calcium/Vitamin D	46/345 (13.3)	47/843 (5.6)	89/1248 (7.1)
Splenic involvement	37/797 (4.6)	56/843 (6.6)	93/1248 (7.5)
Neurological involvement	93/1258 (7.4)	54/1180 (4.6)	129/1508 (8.6)
Salivary gland involvement	NAD	42/1180 (3.6)	34/1248 (2.7)
ENT involvement	6/345 (1.7)	29/843 (3.4)	146/1508 (9.7)
Bone/joint involvement	1/345 (0.3)	23/1180 (1.9)	105/1508 (7)
Renal involvement	12/345 (3.5)	5/843 (0.6)	11/1248 (0.9)
Muscular involvement	NAD	7/843 (0.8)	12/1248 (1)

Values are represented as means and range for continuous variables and numbers and percentages for categorical variables.
Variables are not detailed in all cases, and the prevalence of a specific feature has been stated as the number of cases with the feature (numerator)/number of cases in which the feature was detailed (denominator).
All comparisons were statistically significant ($p<0.05$) except for mean age at diagnosis ($p=0.867$), salivary gland involvement ($p=0.287$) and muscle involvement ($p=0.940$).
In case of no available data, p-value was computed excluding Midwest (†). NAD: No available data.

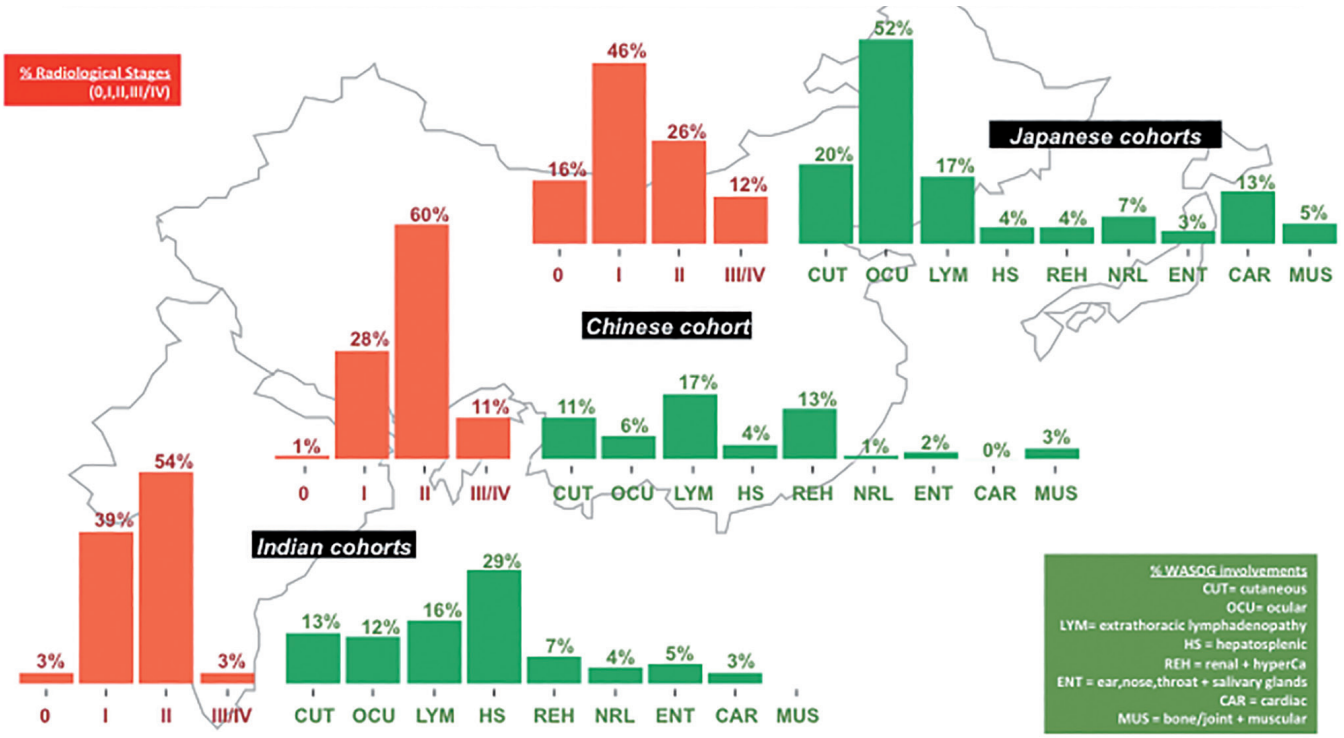

Fig. 4. Thoracic and extrathoracic phenotypic expression of sarcoidosis in Asia: comparison between studies in Indian, Chinese and Japanese cohorts. Coloured columns represent the figures obtained in each geographical area. There was no available data for bone/joint and muscular involvement for Indian cohorts.

Table III. Merged data for thoracic (Scadding radiological stages) and extrathoracic (WASOG classification) phenotypic expression of sarcoidosis in Asian cohorts: comparison between studies carried out in Middle East, Indian, Chinese and Japanese cohorts.

	Middle East cohorts (n=659)	Indian cohorts (n=264)	Chinese cohort (n=481)	Japanese cohorts (n=3315)
Mean age at diagnosis (years)	42.8 (42.8-42.8)	43.5 (43.5-43.5)	NAD	40 (30-45.1)
Gender (% of women)	396/659 (60.1)	126/264 (47.7)	341 (70.7)	1728/2629 (65.7)
Thoracic involvement (%)				
Radiological stage 0	17/336 (5.1)	7/264 (2.7)	5 (1)	358/2221 (16.1)
Radiological stage I	182/336 (54.2)	104/264 (39.4)	133 (27.7)	1024/2221 (46.1)
Radiological stage II	107/336 (31.8)	144/264 (54.5)	291 (60.5)	582/2221 (26.2)
Radiological stage III	23/336 (6.8)	8/264 (3)	28 (5.8)	169/2221 (7.6)
Radiological stage IV	7/336 (2.1)	1/264 (0.4)	24 (5)	88/2221 (4)
Extrathoracic involvement (%)				
Cutaneous involvement	85/534 (15.9)	34/264 (12.9)	51 (10.6)	533/2688 (19.8)
Ocular involvement	40/534 (7.5)	33/264 (12.5)	30 (6.2)	1514/2896 (52.3)
Extrathoracic lymph node involvement	17/142 (12)	41/264 (15.5)	80 (16.6)	288/1713 (16.8)
Cardiac involvement†	NAD	5/164 (3)	1 (0.2)	392/3108 (12.6)
Liver involvement†	NAD	31/164 (18.9)	7 (1.5)	62/1713 (3.6)
Calcium/Vitamin D‡	8/142 (5.6)	12/164 (7.3)	63 (13.1)	NAD
Splenic involvement†	NAD	17/164 (10.4)	12 (2.5)	0/686 (0)
Neurological involvement	13/392 (3.3)	11/264 (4.2)	6 (1.2)	71/1027 (6.9)
Salivary gland involvement†	NAD	14/264 (5.3)	4 (0.8)	31/1027 (3)
ENT involvement	NAD	NAD	4 (0.8)	NAD
Bone/joint involvement§	NAD	NAD	10 (2.1)	7/1027 (0.7)
Renal involvement§	NAD	NAD	0 (0)	36/1027 (3.5)
Muscular involvement§	NAD	NAD	3 (0.6)	42/1027 (4.1)

Values are represented as means and range for continuous variables and numbers and percentages for categorical variables.
Variables are not detailed in all cases, and the prevalence of a specific feature has been stated as the number of cases with the feature (numerator)/number of cases in which the feature was detailed (denominator). All comparisons were statistically significant ($p<0.05$) except for extrathoracic lymph node involvement (p=0.492). *p*-value could not be computed for mean age at diagnosis and ENT involvement. In the case of no available data, *p*-value was computed excluding Middle East (†), Japan (‡) or Middle East & India (§). NAD: No available data.

Table IV. Phenotypic clinical patterns driven by the three main ethnicities most frequently reported in sarcoidosis by classifying the study cohorts according to the predominant ethnicity (defined as that representing more than 50% of the population studied).

	White-predominant cohorts (n=94448)	BAA-predominant cohorts (n=4522)	Asian-predominant cohorts (n=4060)
Mean age at diagnosis (years)	41.4 (36-51.2)	39 (33.7-42)	40.7 (30-45.1)
Gender (% of women)	37814/74083 (51)	2683/4048 (66.3)	2194/3374 (65)
Thoracic involvement (%)			
Radiological stage 0	602/8924 (6.7)	528/2452 (21.5)	370/2966 (12.5)
Radiological stage I	3271/8924 (36.7)	595/2452 (24.3)	1261/2966 (42.5)
Radiological stage II	3382/8924 (37.9)	640/2452 (26.1)	1017/2966 (34.3)
Radiological stage III	1269/8924 (14.2)	365/2452 (14.9)	205/2966 (6.9)
Radiological stage IV	400/8924 (4.5)	324/2452 (13.2)	113/2966 (3.8)
Extrathoracic involvement (%)			
Cutaneous involvement	2209/8721 (25.3)	815/2797 (29.1)	618/3433 (18)
Ocular involvement	1401/8326 (16.8)	491/2288 (21.5)	1577/3641 (43.3)
Extrathoracic lymph node involvement	803/7119 (11.3)	189/1845 (10.2)	409/2458 (16.6)
Cardiac involvement	118/7012 (1.7)	81/2103 (3.9)	398/3753 (10.6)
Liver involvement	553/7568 (7.3)	414/2697 (15.4)	100/2358 (4.2)
Calcium/Vitamin D	211/3443 (6.1)	89/1248 (7.1)	75/645 (11.6)
Splenic involvement	284/5503 (5.2)	93/1248 (7.5)	29/1331 (2.2)
Neurological involvement	292/6972 (4.2)	176/2306 (7.6)	88/1772 (5)
Salivary gland involvement	167/5702 (2.9)	41/1585 (2.6)	49/1772 (2.8)
ENT involvement	85/2872 (3)	146/1508 (9.7)	4/481 (0.8)
Bone/joint involvement	150/5350 (2.8)	115/1845 (6.2)	17/1508 (1.1)
Renal involvement	49/4874 (1)	11/1248 (0.9)	36/1508 (2.4)
Muscular involvement	20/4120 (0.5)	12/1248 (1)	45/1508 (3)

Values are represented as means and range for continuous variables and numbers and percentages for categorical variables.
Variables are not detailed in all cases, and the prevalence of a specific feature has been stated as the number of cases with the feature (numerator)/number of cases in which the feature was detailed (denominator).
All comparisons were statistically significant (p-values<0.001) except for mean age at diagnosis (p=0.461) and salivary gland involvement (p=0.753).
BAA, Black or African American.

lence of sarcoidosis follows the same predominantly north-south gradient reported for other autoimmune diseases (Fig. 1b), including multiple sclerosis (13), inflammatory bowel disease (14, 15), primary biliary cirrhosis (16), systemic sclerosis (17), and Sjögren syndrome (18). This geographically-driven autoimmune gradient has been considered an epidemiological inverse mirror of the south-north gradient reported for infectious diseases (19).

Geographical determinants in Europe

More than 60 studies have estimated the incidence and prevalence of sarcoidosis in European countries and, although the figures differ, the higher the latitude, the higher the reported rates, with incidence rates greater than 10 cases per 100,000 persons in Sweden, Norway, Finland, Denmark, the Netherlands, the UK and Germany (3, 9-12); in contrast, incidence rates of <5 cases are reported in Croatia, Italy and Spain (20–22) (Fig. 1a). Several studies have identified a north-to-south geographical gradient of incidence rates between regions of a country, with a higher frequency of sarcoidosis observed in the northern regions of Sweden (23), Norway (24), Italy (25) and London (26), and the western regions of Denmark (27); in contrast, no clear predominant geo-gradient has been reported in Central European countries, including Germany (28), Switzerland (29) and Poland (30).

The clinical phenotype of sarcoidosis differs when studies in northern or southern European countries are compared (Table I). Sarcoidosis in southern countries is characterised by female involvement in two thirds of cases, radiological stage I as the predominant thoracic involvement (43% of cases) and with extrathoracic disease clearly dominated by cutaneous (39%) involvement, followed by extrathoracic lymph node involvement (16%), conforming a phenotypic cluster of lymphadenopathic-skin involvement typical of Löfgren syndrome. In contrast, sarcoidosis in northern European countries is characterised by an equal gender ratio, stage II as the predominant thoracic radiological stage II (42% of cases), and ocular involvement (24%) as the predominant extrathoracic feature (Fig. 2).

Geographical determinants in the US

The incidence rate of sarcoidosis in the US ranges between 5 and 10 cases per 100,000 persons (31-34). However, the largest US studies have reported geo-driven differences across regions, with a higher frequency of sarcoidosis found in the Midwest and Northeast compared with the Southwest (34, 35). Phenotypically, sarcoidosis in the southern US is characterised by female involvement in two thirds of cases, the highest rate of pulmonary fibrosis (stage IV in nearly 20% of cases), and extrathoracic disease dominated by cutaneous (33%), ocular (22%) and liver (19%) involvements. Midwest patients had the highest rate of thoracic stage III (21%) and predominantly cutaneous (30%), ocular (15%) and calcium/vitamin D (13%) involvements, while nearly 80% of Northeast patients presented with milder thoracic involvement (stages I–II) and predominantly

Table V. Influence of ethnicity on the phenotypic expression of sarcoidosis.

	Mean age at dx (Years)	Mean gender ratio (W:M)	Pulmonary disease (%)†	Pulmonary fibrosis (%)‡	Cutaneous involv (%)§	Ocular involv (%)§	Lymp node involv (%)§	Cardiac involv (%)§	Liver involv (%)§	NRL involv (%)§
Number of cohorts	22	53	35	34	29	28	20	21	21	21
Number of patients included	10,885	47,473	13,139	12,658	10,438	9,972	7,859	9,323	8,956	8,903
Whites (%)	-0.30 (0.179)	**-0.55 (<0.001)**	0.15 (0.386)	0.05 (0.796)	-0.06 (0.755)	**-0.59 (0.001)**	0.03 (0.895)	-0.40 (0.069)	-0.18 (0.443)	0.04 (0.853)
African Americans (%)	-0.06 (0.805)	**0.36 (0.008)**	0.29 (0.087)	**0.58 (<0.001)**	**0.37 (0.048)**	-0.02 (0.907)	-0.36 (0.115)	-0.33 (0.146)	0.26 (0.251)	0.26 (0.254)
Asian (%)	0.37 (0.089)	0.16 (0.256)	-0.23 (0.188)	-0.32 (0.063)	-0.01 (0.956)	0.37 (0.053)	0.06 (0.815)	0.23 (0.323)	0.09 (0.695)	-0.33 (0.144)

Values are represented as Spearman correlation coefficients and *p*-values in brackets between the ethnicity and the different phenotypic expression covariates. The number of cohorts with available data that were considered to compute correlations was detailed as well as the number of patients included. The correlation coefficients were computed considering the percentages of different ethnicity groups and the percentages of phenotypic expression covariates, except for age at diagnosis and gender ratio where mean values were considered.
Percentage of patients with radiological stages II+III+IV (†), stage IV (‡) and extrathoracic involvements defined according to the WASOG classification (§).
dx: Diagnosis; W: Women; M: Men; Involv: Involvement; NRL: Neurological.

extrathoracic disease, consisting of ocular (15%), cutaneous (13%) and peripheral lymph node (11%) involvements (Table II). The higher frequency of BAA patients in southern cohorts (70% *vs*. 50%) could influence the differentiated phenotypic patterns found across US regions, but not the differences between Northeast and Midwest patients, due to the similar regional ethnic distribution (Fig. 3).

Geographical determinants in Asia

Available epidemiological data on the reported incidence rates of sarcoidosis in Asian countries show homogeneous results in the four main geographical areas, with very low figures (often <1 per 100,000 persons) in Middle and Far Eastern countries. Although there is no available data on the incidence in South Asia, the prevalence of sarcoidosis reported in Indian studies is among the highest in the world, ranging between 60 and 150 cases per 100,000 persons (36) (Fig. 1b). Some Japanese studies have reported a north-to-south geographical gradient within Japan, with sarcoidosis more frequently reported in northern regions (37, 38).

Epidemiologically, the principal differences between the main Asian regions with available data (Middle East, India, China and Japan) were found in gender distribution, which was equal in India while, in Japan, two thirds of cases are female (Table III). With respect to thoracic presentation, patients from the Middle East presented mainly with stage I (54%) and those from India and China with stage II (55–60%), while Japanese patients had the highest frequency of stages III/IV (12%). The pattern of extrathoracic disease also varied widely, with a significant predominance of ocular involvement in Japanese patients (52%, the highest rate worldwide), extrathoracic lymph node involvement as the predominant involvement in China (16%) and liver involvement in India (19%) (Fig. 4).

Ethnic determinants

Ethnicity is one of the key epidemiological factors in the frequency and phenotypic expression of sarcoidosis. Although sarcoidosis affects all ethnicities, it is more frequently reported in Whites: our merged big-data analysis showed 60% of reported cases were classified as White, 23 % as BAA, 13% as Asian and 4% as other ethnicities (Table IV). However, variations between ethnic groups have been well documented since the 1960s (39), with most studies made in multi-ethnic populations inhabiting the same geographical area (predominantly the US).

Sarcoidosis in White people

Even though sarcoidosis is mainly reported in White people, the incidence rates are largely influenced by the previously-mentioned geographical determinants. The available epidemiological data shows a north-to-south gradient for incidence rates, with >10 cases per 100,000 persons in northern European countries, between 5 and 10 cases in Central Europe, Australia and White people living in the US or South Africa, and ≤5 cases in Mediterranean countries (Fig. 1a).

Comparison of international cohorts grouped according to the predominant ethnicity (White, BAA or Asian) found that sarcoidosis in predominantly White cohorts is characterised by an equal gender distribution, the oldest mean age at diagnosis among the three ethnic groups, radiological stage II as the predominant thoracic pattern (38%), and lower frequencies of the most-severe WASOG extrathoracic involvements (Table IV) (Fig. 5). The higher the frequency of patients classified as White in a cohort, the higher the reported frequency of males affected and the lower the reported frequency of ocular and cardiac involvements (Table V).

Sarcoidosis in Black/African-American people

The highest incidence rates of sarcoidosis are uniformly reported in BAA people, independently of the geographical location, with rates between 2 and 10-fold higher than those reported in White people living in the same geographical area. In US epidemiological studies, the lowest and highest incidence rates in BAA people are 17.8 (34) and 81.8 (40) per 10^6 persons, respectively, compared with 19.8 (41) in BAA people living in London and 35.5 in South-Africans (42). The main studies characterising the clinical presentation of sarcoidosis in BAA people have always used a face-to-face comparison with White people living in US and have found that BAA people develop the first sarcoidosis-related symptoms at an earlier age, with a diagnostic biopsy reported ten years earlier, have a two-fold higher median granuloma density, more frequently re-

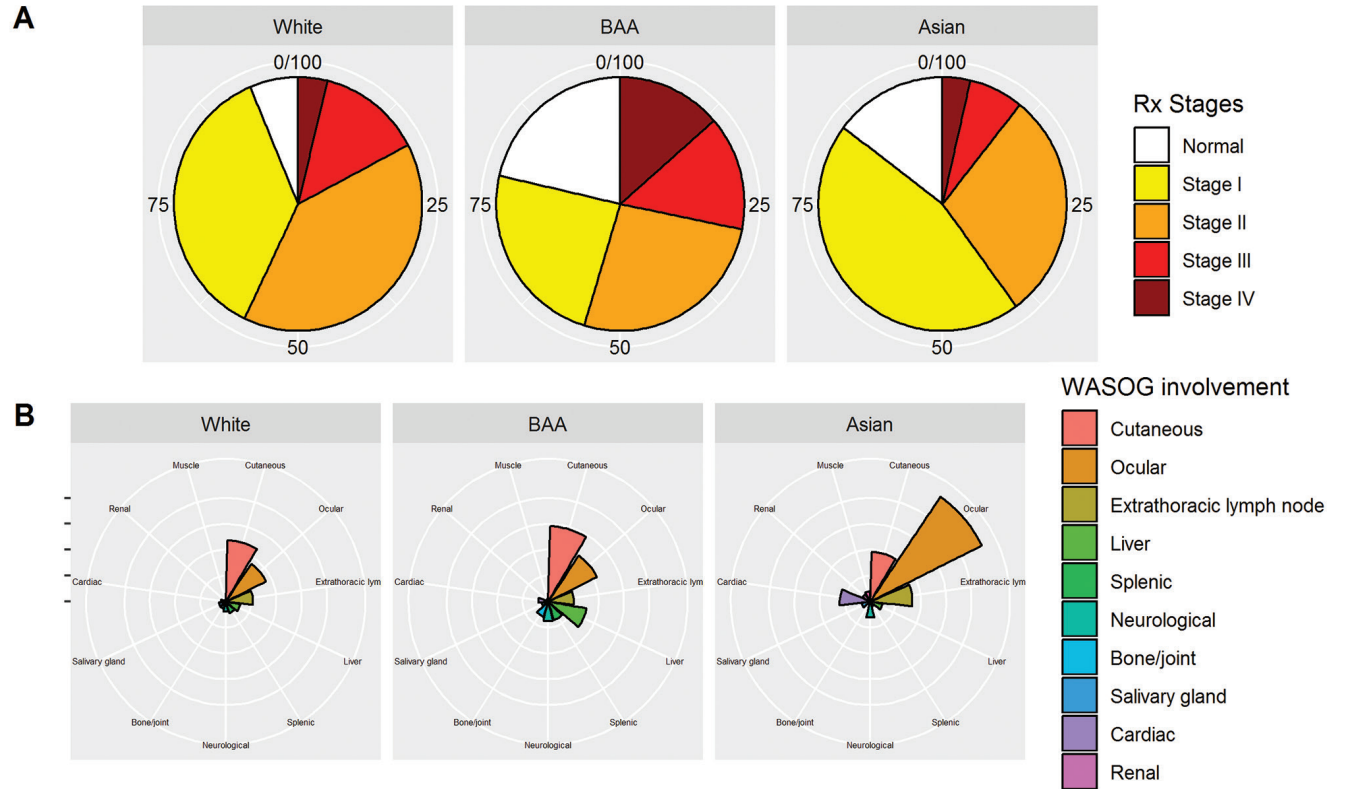

Fig. 5. Radiological patterns and extrathoracic WASOG organ-by-organ involvements in the three ethnic groups most frequently reported in sarcoidosis; ethnic groups were defined according to the predominant ethnicity (as representing more than 50% of the population studied) of each cohort.

quire sarcoidosis-specific therapies and have a poorer prognosis with respect to both disease severity and mortality, with increased hospitalisation rates and a 12-fold higher age-adjusted mortality rate (32, 43-54).

The phenotype of sarcoidosis in predominantly BAA cohorts is characterised by the highest frequency of female involvement and the lowest mean age at diagnosis among the three main ethnic groups, a higher frequency of the more severe thoracic stages (III-IV in around 30% of cases), and with cutaneous (29%), ocular (22%) and liver (15%) involvements being the predominant extrathoracic features (Table IV, Fig. 5). The higher the frequency of patients classified as BAA in a cohort, the higher the reported frequency of female involvement, pulmonary fibrosis and cutaneous involvement (Table V).

Sarcoidosis in Asian people

The incidence rates of sarcoidosis in Asian countries are the lowest reported worldwide, with figures of 0.5–2 cases per 100,000 persons in Israel (55, 56), 0.3-1.7 in Japan (57), 0.6 in Singapore (58) and 0.1 in Korea (59) (Fig. 1a). In the largest reported US series, the incidence rate in Asians living in the US was 3.2 (34), while Anantham *et al.* (58) reported a differentiated yearly incidence in people born in India (4.57), Malaysia (1.30) or China (0.23) living in Singapore.

The phenotype of sarcoidosis in predominantly-Asian cohorts is principally characterised by a higher frequency of thoracic stage I (43%) and of many WASOG extrathoracic involvements including ocular (43%), extrathoracic lymph node (17%), hypercalcemia (12%) and cardiac (11%) involvements compared with the other main ethnic groups (Table IV, Fig. 5). The higher the frequency of patients classified as Asian included in a cohort, the higher the reported frequency of ocular involvement (Table V). Comparative ethnicity studies including Asian cohorts are very limited. Pietinalho *et al.* (60) compared Japanese and Finnish patients and found a lower mean age at diagnosis, a higher frequency of ocular involvement and normal radiographic stage, and a lower frequency of Löfgren syndrome (with no cases of erythema nodosum) in Japanese patients. Behbehani *et al.* (61) found more advanced radiological stages in Arabs compared with Asians living in Kuwait, while Yigla *et al.* (62) found radiographic stage II was the most frequent in Jewish Israeli patients and stage I was predominant in Israeli Arabs.

Sarcoidosis in other ethnic groups

There are few studies of sarcoidosis in other ethnic groups. Only isolated cases have been reported in patients having origins in any of the original American peoples (Alaska Native, American Indians). With respect to Hispanic people, the largest US series have reported an incidence of 4.3 (34), similar to that reported for White US people.

A recent New Zealand study compared 337 White patients with 69 Maori/Pacific Islanders (PI), and found that only people of Indian ethnicity were over-represented with respect to the current New Zealand census data (2013), with lung involvement and erythema nodosum being more commonly reported in White people, while ocular

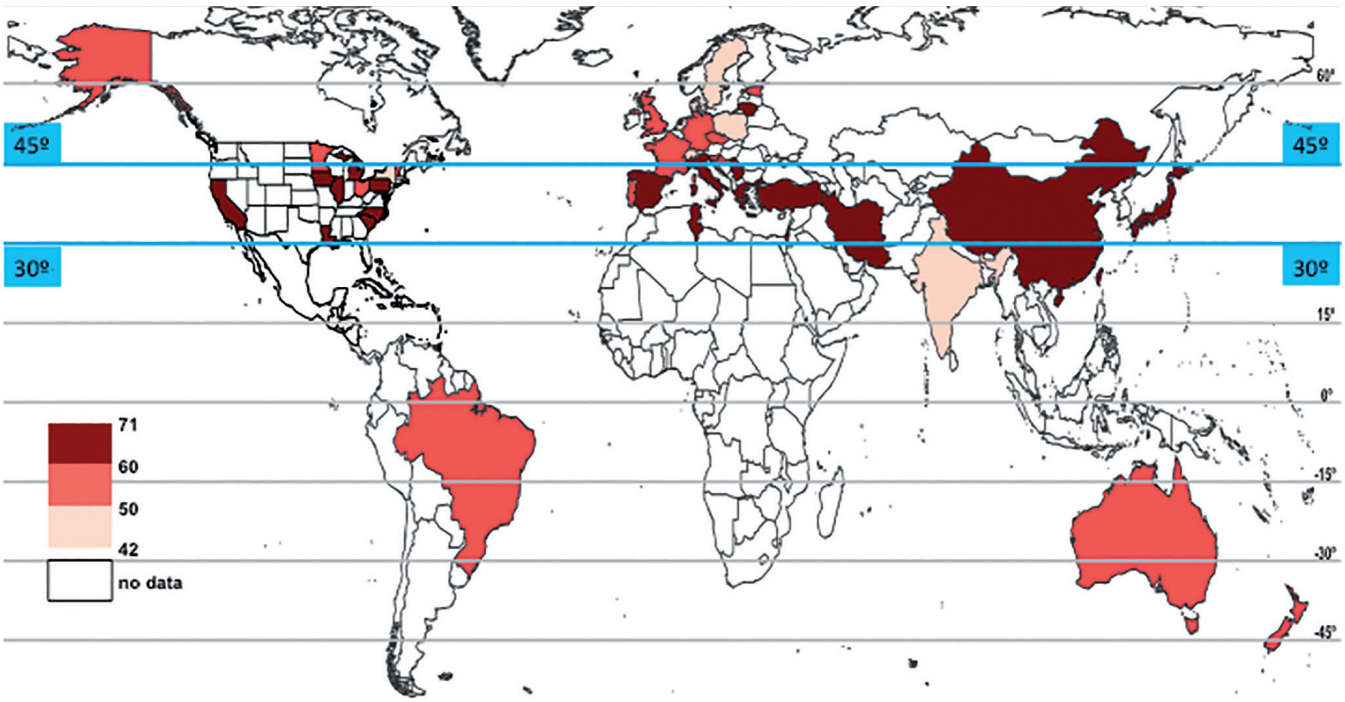

Fig. 6. Geographically-driven distribution of the frequency of women affected by sarcoidosis: geographical cluster of greater female involvement in countries located between parallels 30 and 45.

involvement and other cutaneous features were more commonly reported in Maori/PI people. The study concluded that people of Polynesian ancestry appear to have less pulmonary and more extra-pulmonary manifestations of sarcoidosis (63).

Combining geographic and ethnic determinants

Due to the influence of ethnicity on the prevalence and incidence rates of sarcoidosis, geographically-driven differences may be influenced by the specific ethnic distribution in each country. However, there are phenotypic geographically-driven similarities between countries with different ethnicities. With respect to the worldwide gender distribution, we have identified a geographical cluster of greater female involvement (more than three-fifths of cases) in countries located between parallels 30 and 45 (Mediterranean countries, China, Japan and Southern US) (Fig. 6); in contrast, an equal gender distribution is found in Northern Europe and India. With respect to extrathoracic involvement, there is a clearly-differentiated phenotypic pattern between northern European and Mediterranean cohorts (Fig. 2) and between Northeast and Midwest US cohorts (Fig. 3), even though these geographical regions have a similar ethnic distribution. Investigation of the ancestral influences on geographically-driven patterns in areas with a similar ethnic profile might be of interest, since there are significant variations in ancestry between northern and southern European White people (64), and in White US people, in whom Scandinavian and German ancestry is predominantly reported in people from the northern and western US.

Additionally, geoepidemiology might underscore the complex combination of genes and environment entailed in the pathogenesis of sarcoidosis by comparing ethnic migrant populations with native populations and with the same ethnic group in their region of origin. Outside the US, people from India living in London or Singapore had the highest rates of incidence and those from East Asia living in the US or Singapore the lowest rates in comparison with the other ethnicities living in the same geographical area. In multiethnic cohorts from London, with a differing ethnic distribution than the US cohorts, the highest rates were reported in West Indian (incidence of 58, prevalence of 183 cases per 100,000 persons) and Irish (incidence 21, prevalence 155) people in comparison with UK-born people (incidence 4, prevalence 27) (65, 66). Benatar found a prevalence among Blacks of 23.2 per 10^6 inhabitants compared with Whites (3.7) and mixed race (11.6) (42) people in Cape Town, while Anantham *et al*. (58) reported a yearly incidence of 0.56 per 100,000 persons in Singapore, with clearly different figures for Indian (4.57), Malaysian (1.30) and Chinese (0.23) people.

Conclusions

Sarcoidosis is probably one of the systemic diseases with the greatest influence of geoepidemiological factors on the frequency and phenotypic expression. Geographically, sarcoidosis is principally reported in the Northern Hemisphere, and among international studies including >100 cases, 96% of cohorts (including 99.6% of reported cases) were from northern countries. Although a potential bias concerning unreported cases in southern countries cannot be ruled out, the numbers seem to be too large to be attributed to bias after adjusting for the total population of each hemisphere. By continent, there is a clear north-south gradient in the incidence and prevalence in Europe,

while in other continents the influence of ethnicity may be greater than that of geographical determinants. The best examples are the US studies: the incidence rates in BAA people are 2–7-fold higher than the global US rate, while the rates in Hispanics and Asians are half as much. In Asia, the highest prevalence rate is reported in India, while Middle East countries have similar figures to Mediterranean countries, and East Asian countries have the lowest worldwide incidence and prevalence rates.

There are large ethnicity-driven variations in the frequency, epidemiology, clinical expression and outcomes of sarcoidosis. In addition to the strong influence of ethnicity on modulating the disease frequency around the world, ethnicity heavily influences the clinical phenotype by modifying the age at diagnosis (younger ages in BAA, older ages in East Asians) and the rates of thoracic (highest rates for stage IV in BAA) and extrathoracic (highest rates in BAA and East Asians) involvements. Ethnicity also plays a role in modulating the predominant organ-by-organ extrathoracic involvements among the three main ethnic groups: Asian patients have the highest rates of ocular (2–3-fold higher than other ethnicities), cardiac (3–6-fold higher) and renal and muscular (4-fold higher) sarcoidosis, while BAA patients have the highest frequencies of liver (3-fold) and bone/joint (2–4-fold) sarcoidosis. Ethnically-driven differences should be evaluated taking into account socioeconomic disparities that could influence the level of exposure to potential environmental factors (3).

Big data-based approaches will be essential to analyse the huge number of factors potentially involved and their complex relationships (different exposures at different degrees in different geographical locations). A new era in the collection, management and interpretation of large amounts of geoepidemiological medical data is opening up for diseases which, like sarcoidosis, are rare and have a complex etiopathogenic scenario. However, the results of big data analysis should be interpreted with caution, because statistical analysis in very large cohorts could detect differences which, although significant, could be clinically irrelevant, with specific studies being required to confirm relevance. In addition, the predominant presence of European, US and Japanese patients, due to the largest cohorts being overwhelmingly reported from these regions, could limit generalisation of the results in specific populations of patients. The management of aggregate data also limits the standardisation of the results for some variables such as ethnicity (cohorts should be classified according to the predominant ethnicity, not by the crude figures corresponding to each ethnicity). In spite of these limitations, offering a picture of nearly 130,000 patients with sarcoidosis may be considered of interest for a better understanding of how the disease is expressed worldwide. Geoepidemiological studies enhanced by big data could yield important clues to a better understanding of the aetiopathogenesis of sarcoidosis, helping to design strategies to reduce its development (primary prevention) and severity (secondary prevention).

Members of the Autoimmune Big Data Study Group

The members of the Autoimmune Big Data Study Group involved in this project (serving as scientific advisors, critically reviewing the study proposal, helping to collect data, and participating in writing or technical editing of the manuscript) are:

Pilar Brito-Zerón[1,2,17], Roberto Pérez-Alvarez[3,17], Daphne Superville[4], Belchin Kostov[5], Nihan Acar-Denizli[6], Antoni Sisó-Almirall[5], Alejandra Flores-Chavez[2,8], Courtney Grant[15], Soledad Retamozo[16], Xavier Bosch[7], Lucio Pallarés[9,17], David Buss[2], Yehuda Shoenfeld[12], Munther A. Khamashta[11], Robert P. Baughman[14], Manuel Ramos-Casals[2,13,17] (coordinator).

Affiliations

[1]Autoimmune Diseases Unit, Dept. of Medicine, Hospital CIMA - Sanitas, Barcelona, Spain.
[2]Laboratory of Autoimmune Diseases Josep Font, IDIBAPS-CELLEX, Department of Autoimmune Diseases, ICMiD, Hospital Clínic, Barcelona, Spain.
[3]Dept. of Internal Medicine, Hospital Alvaro Cunqueiro, Vigo, Spain.
[4]Massachusetts Institute of Technology (MIT), Cambridge, MA, USA.
[5]Primary Healthcare Transversal Research Group, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), CAPSBE, Barcelona, Spain.
[6]Dept. of Statistics, Faculty of Science and Letters, Mimar Sinan Fine Arts University, Istanbul, Turkey.
[7]Dept. of Internal Medicine, ICMiD, Hospital Clínic, Barcelona, Spain;
[8]Biomedical Research Unit 02, Clinical Epidemiology Research Unit, UMAE, Specialties Hospital, Western Medical Center, Mexican Institute for Social Security (IMSS), Guadalajara, Mexico; Postgraduate Program of Medical Science, University Center for Biomedical Research (CUIB), University of Colima, Colima, Mexico.
[9]Systemic Autoimmune Diseases Unit, Dept. of Internal Medicine, Hospital de Son Espases, Palma de Mallorca, Spain.
[11]Lupus Research Unit, The Rayne Institute, St Thomas' Hospital, King's College University, London, UK.
[12]Zabludowicz Center for Autoimmune Diseases, Chaim Sheba Medical Center, Tel Hashomer, Israel Incumbent of the Laura Schwarz-Kipp Chair for Research of Autoimmune Diseases, Sackler Faculty of Medicine, Tel-Aviv University, Tel Aviv, Israel.
[12]Dept. of Medicine, University of Barcelona, Barcelona, Spain.
[14]University of Cincinnati Medical Center, Dept. of Medicine, Cincinnati, OH, USA.
[15]Glasgow University, UK.
[16]Instituto de Investigaciones en Ciencias de la Salud (INICSA), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Cordoba, Argentina.
[17]SarcoGEAS Study Group, Grupo de Estudio de Enfermedades Autoinmunes Sistémicas (GEAS) de la Sociedad Española de Medicina Interna (SEMI).

References

1. BAUGHMAN RP, LOWER EE, DU BOIS RM: Sarcoidosis. *Lancet* 2003; 361: 1111-8.
2. CULVER DA: Sarcoidosis. *Immunol Allergy Clin North Am* 2012; 32: 487-511.
3. DUBREY S, SHAH S, HARDMAN T, SHARMA R: Sarcoidosis: the links between epidemiology and aetiology. *Postgrad Med J* 2014; 90: 582-9.
4. PEREIRA CAC, DORNFELD MC, BAUGHMAN R, JUDSON MA: Clinical phenotypes in sar-

coidosis. *Curr Opin Pulm Med* 2014; 20: 496-502.
5. RAMOS-CASALS M, BRITO-ZERÓN P, KOSTOV B *et al*.: Google-driven search for big data in autoimmune geoepidemiology: Analysis of 394,827 patients with systemic autoimmune diseases. *Autoimmun Rev* 2015; 14: 670-9.
6. HUNNINGHAKE GW, COSTABEL U, ANDO M *et al*.: ATS/ERS/WASOG statement on sarcoidosis. American Thoracic Society/European Respiratory Society/World Association of Sarcoidosis and other Granulomatous Disorders. *Sarcoidosis Vasc Diffus lung Dis* 1999; 16: 149-73.
7. JUDSON MA, COSTABEL U, DRENT M *et al*.; ORGAN ASSESSMENT INSTRUMENT INVESTIGATORS TWS: The WASOG Sarcoidosis Organ Assessment Instrument: An update of a previous clinical tool. Sarcoidosis, Vasc. Diffus. lung Dis. Off. J. WASOG / World Assoc. Sarcoidosis Other Granulomatous Disord 2014; 31: 19-27.
8. WICKHAM H: ggplot2: elegant graphics for data analysis. New York: Springer, 2009.
9. HOSODA Y, YAMAGUCHI M, HIRAGA Y: Global epidemiology of sarcoidosis. What story do prevalence and incidence tell us? *Clin Chest Med* 1997; 18: 681-94.
10. SHARMA OP: Sarcoidosis around the world. *Clin Chest Med* 2008; 29: 357-63, vii.
11. BAUER HJ, LOFGREN S: International study of pulmonary sarcoidosis in mass chest radiography. *Acta Med Scand Suppl* 1964; 425: 103-5.
12. LEVINSKY L, CUMMISKEY J, ROMER FK *et al*.: Sarcoidosis in Europe: a cooperative study. *Ann N Y Acad Sci* 1976; 278: 335-46.
13. GRANT WB, MASCITELLI L: Evidence that the north-south gradient of multiple sclerosis may not have disappeared. *J Neurol Sci* 2012; 315: 178-9.
14. JUSSILA A, VIRTA LJ, SALOMAA V, MAKI J, JULA A, FARKKILA MA: High and increasing prevalence of inflammatory bowel disease in Finland with a clear North-South difference. *J Crohns Colitis* 2013; 7: e256-62.
15. LOGAN I, BOWLUS CL: The geoepidemiology of autoimmune intestinal diseases. *Autoimmun Rev* 2010; 9: A372-8.
16. METCALF J, JAMES O: The geoepidemiology of primary biliary cirrhosis. *Semin Liver Dis* 1997; 17: 13-22.
17. SHAPIRA Y, AGMON-LEVIN N, SHOENFELD Y: Geoepidemiology of autoimmune rheumatic diseases. *Nat Rev Rheumatol* 2010; 6: 468-76.
18. BRITO-ZERON P, ACAR-DENIZLI N, ZEHER M *et al*.: Influence of geolocation and ethnicity on the phenotypic expression of primary Sjögren's syndrome at diagnosis in 8310 patients: a cross-sectional study from the Big Data Sjögren Project Consortium. *Ann Rheum Dis* 2017; 76: 1042-50.
19. YOUINOU P, PERS J-O, GERSHWIN ME, SHOENFELD Y: Geo-epidemiology and autoimmunity. J Autoimmun 2010; 34: J163-7.
20. MISE K, JURCEV-SAVICEVIC A, GOIC-BARISIC I, KOZUL K, GUDELJ I: Sarcoidosis and tuberculosis in South Croatia: are there epidemiological similarities or not? *Public Health* 2011; 125: 734-7.
21. FAZZI P, SOLFANELLI S, DI PEDE F, BEGLIOMINI E, DIVIGGIANO E, GIUNTINI C: Sarcoidosis in Tuscany. A preliminary report. *Sarcoidosis* 1992; 9: 123-6.
22. MANA J, BADRINAS F, MORERA J, FITE E, MANRESA F, FERNANDEZ-NOGUES F: Sarcoidosis in Spain. *Sarcoidosis* 1992; 9: 118-22.
23. ARKEMA E V, GRUNEWALD J, KULLBERG S, EKLUND A, ASKLING J: Sarcoidosis incidence and prevalence: a nationwide register-based assessment in Sweden. *Eur Respir J England* 2016; 48: 1690-9.
24. GLENNAS A, KVIEN TK, MELBY K *et al*.: Acute sarcoid arthritis: occurrence, seasonal onset, clinical features and outcome. *Br J Rheumatol* 1995; 34: 45-50.
25. BLASI A, PEZZA A: Epidemiology of sarcoidosis in Italy. *Sarcoidosis* 1989; 6: 55-6.
26. JAMES DG, BRETT GZ: Prevalence of intrathoracic sarcoidosis in Britain. *Acta Med Scand Suppl Sweden* 1964; 425: 115-7.
27. BYG K-E, MILMAN N, HANSEN S: Sarcoidosis in Denmark 1980-1994. A registry-based incidence study comprising 5536 patients. *Sarcoidosis Vasc Diffus lung Dis* 2003; 20: 46-52.
28. BUSS J, DORKEN H: Epidemiology of intrathoracic sarcoidosis in the Federal Republic of Germany. *Pneumonologie* 1975; 153: 1-20.
29. DEUBELBEISS U, GEMPERLI A, SCHINDLER C, BATY F, BRUTSCHE MH: Prevalence of sarcoidosis in Switzerland is associated with environmental factors. *Eur Respir J* 2010; 35: 1088-97.
30. KOWALSKA M, NIEWIADOMSKA E, ZEJDA JE: Epidemiology of sarcoidosis recorded in 2006-2010 in the Silesian voivodeship on the basis of routine medical reporting. *Ann Agric Environ Med* 2014; 21: 55-8.
31. HENKE CE, HENKE G, ELVEBACK LR, BEARD CM, BALLARD DJ, KURLAND LT: The epidemiology of sarcoidosis in Rochester, Minnesota: a population-based study of incidence and survival. *Am J Epidemiol* 1986; 123: 840-5.
32. UNGPRASERT P, CARMONA EM, UTZ JP, RYU JH, CROWSON CS, MATTESON EL: Epidemiology of Sarcoidosis 1946-2013: A Population-Based Study. *Mayo Clin Proc* 2016; 91: 183-8.
33. REICH JM, JOHNSON RE: Incidence of clinically identified sarcoidosis in a northwest United States population. *Sarcoidosis Vasc Diffus lung Dis* 1996; 13: 173-7.
34. BAUGHMAN RP, FIELD S, COSTABEL U *et al*.: Sarcoidosis in America. Analysis based on health care use. *Ann Am Thorac Soc* 2016; 13: 1244-52.
35. DUMAS O, ABRAMOVITZ L, WILEY AS, COZIER YC, CAMARGO CAJ: Epidemiology of sarcoidosis in a prospective cohort study of U.S. women. *Ann Am Thorac Soc* 2016; 13: 67-71.
36. GUPTA SK, MITRA K, CHATTERJEE S, CHAKRAVARTY SC: Sarcoidosis in India. *Br J Dis Chest* 1985; 79: 275-83.
37. HOSODA Y, HIRAGA Y, ODAKA M *et al*.: A cooperative study of sarcoidosis in Asia and Africa: analytic epidemiology. *Ann N Y Acad Sci* 1976; 278: 355-67.
38. YAMAGUCHI M, HOSODA Y, SASAKI R, AOKI K: Epidemiological study on sarcoidosis in Japan. Recent trends in incidence and prevalence rates and changes in epidemiological features. *Sarcoidosis* 1989; 6: 138-46.
39. RYBICKI BA, IANNUZZI MC: Epidemiology of sarcoidosis: recent advances and future prospects. *Semin Respir Crit Care Med* 2007; 28: 22-35.
40. SARTWELL PE, EDWARDS LB: Epidemiology of sarcoidosis in the U.S. Navy. *Am J Epidemiol* 1974; 99: 250-7.
41. WESTNEY GE, JUDSON MA: Racial and ethnic disparities in sarcoidosis: from genetics to socioeconomics. *Clin Chest Med* 2006; 27: 453-62, vi.
42. BENATAR SR: Sarcoidosis in South Africa. A comparative study in Whites, Blacks and Coloureds. *S Afr Med J* 1977; 52: 602-6.
43. SAWAHATA M, SUGIYAMA Y, NAKAMURA Y *et al*.: Age-related and historical changes in the clinical characteristics of sarcoidosis in Japan. *Respir Med* 2015; 109: 272-8.
44. MIRSAEIDI M, MACHADO RF, SCHRAUFNAGEL D, SWEISS NJ, BAUGHMAN RP: Racial difference in sarcoidosis mortality in the United States. *Chest* 2015; 147: 438-49.
45. BURKE RR, STONE CH, HAVSTAD S, RYBICKI BA: Racial differences in sarcoidosis granuloma density. *Lung* 2009; 187: 1-7.
46. 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: 119-27.
47. 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: 1885-9.
48. RABIN DL, THOMPSON B, BROWN KM *et al*.: Sarcoidosis: social predictors of severity at presentation. *Eur Respir J* 2004; 24: 601-8.
49. 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: 609-18.
50. GILLUM RF: Sarcoidosis in the United States--1968 to 1984. Hospitalization and death. *J Natl Med Assoc* 1988; 80: 1179-84.
51. GIDEON NM, MANNINO DM: Sarcoidosis mortality in the United States 1979-1991: an analysis of multiple-cause mortality data. *Am J Med* 1996; 100: 423-7.
52. KAJDASZ DK, JUDSON MA, MOHR LCJ, LACKLAND DT: Geographic variation in sarcoidosis in South Carolina: its relation to socioeconomic status and health care indicators. *Am J Epidemiol* 1999; 150: 271-8.
53. GERKE AK, YANG M, TANG F, CAVANAUGH JE, POLGREEN PM: Increased hospitalizations among sarcoidosis patients from 1998 to 2008: a population-based cohort study. *BMC Pulm Med* 2012; 12: 19.
54. GORHAM ED, GARLAND CF, GARLAND FC, KAISER K, TRAVIS WD, CENTENO JA: Trends and occupational associations in incidence of hospitalized pulmonary sarcoidosis and other lung diseases in Navy personnel: a 27-year historical prospective study, 1975-2001. *Chest* 2004; 126: 1431-8.

55. RAKOWER J: Epidemiology of sarcoidosis in Israel. *Acta Med Scand Suppl* 1964; 425: 163-4.
56. YIGLA M, BADARNA-ABU-RIA N, TOV N, RAVELL-WEILLER D, RUBIN A-HE: Sarcoidosis in northern Israel; clinical characteristics of 120 patients. *Sarcoidosis Vasc Diffus lung Dis* 2002; 19: 220-6.
57. NOBECHI K: Epidemiologic features of sarcoidosis in Japan. *Acta Med Scand Suppl* 1964; 425: 165-6.
58. ANANTHAM D, ONG SJ, CHUAH KL, FOOK-CHONG S, HSU A, ENG P: Sarcoidosis in Singapore: epidemiology, clinical presentation and ethnic differences. *Respirology* 2007; 12: 355-60.
59. KIM DS: Sarcoidosis in Korea: report of the Second Nationwide Survey. *Sarcoidosis Vasc Diffus lung Dis* 2001; 18: 176-80.
60. PIETINALHO A, OHMICHI M, HIRASAWA M, HIRAGA Y, LOFROOS AB, SELROOS O: Familial sarcoidosis in Finland and Hokkaido, Japan--a comparative study. *Respir Med* 1999; 93: 408-12.
61. BEHBEHANI N, JAYKRISHNAN B, KHADADAH M, HAWA H, FARAH Y: Clinical presentation of sarcoidosis in a mixed population in the Middle East. *Respir Med* 2007; 101: 2284-8.
62. YIGLA M, BADARNA-ABU-RIA N, GORALNIK L, RUBIN A-HE, WEILER-RAVELL D: Sarcoidosis in residents of northern Israel of Arabic and Jewish origin: a comparative study. *Respirology* 2006; 11: 586-91.
63. WILSHER ML, YOUNG LM, HOPKINS R, CORNERE M: Characteristics of sarcoidosis in Maori and Pacific Islanders. *Respirology* 2017; 22: 360-3.
64. MOSKVINA V, SMITH M, IVANOV D *et al*.: Genetic differences between five European populations. *Hum Hered* 2010; 70: 141-9.
65. MCNICOL MW, LUCE PJ: Sarcoidosis in a racially mixed community. *J R Coll Physicians Lond* 1985;19: 179-83.
66. BRETT GZ: Epidemiological trends in tuberculosis and sarcoidosis in a district of London between 1958 and 1963. *Tubercle* 1965; 46: 413-6.