

Behçet's syndrome incidence and prevalence in Sardinia: implications of a latent class analysis combining administrative and clinical data

M. Piga^{1,2}, D. Corradini¹, P. Carboni¹, A. Floris^{1,2}, M.A. Palmas³,
S. Sanna⁴, G. Porru⁵, G.L. Erre^{6,7}, S. Savasta⁸, A. Cauli^{1,2}

¹Department of Medicine and Public Health, University of Cagliari; ²Rheumatology Unit, AOU Cagliari; ³Regional Epidemiological Observatory, Department of Health and Hygiene, Sardinian Regional Government, Cagliari; ⁴SSD Reumatologia, ASL Sulcis Iglesiente, Iglesias; ⁵Ambulatorio di Reumatologia, ASL Nuoro; ⁶Dipartimento di Medicina, Chirurgia e Farmacia, Università degli Studi di Sassari; ⁷SSD Reumatologia, Azienda Ospedaliero-Universitaria di Sassari; ⁸Paediatric Clinic and Rare Diseases, Microcitamico Hospital, Department of Medicine and Public Health, University of Cagliari, Italy.

Abstract

Objective

To estimate the epidemiology of Behçet's syndrome (BS) in Sardinia using a combined administrative and clinical data latent class analysis (LCA).

Methods

Cases with a diagnosis of BS were retrieved from 2006-2016 in the Hospital Discharge Records (HDRs), 2006-2016 Rare Diseases Regional Register (RDRR), and 2014-2016 in the Specialist Outpatient Database (SOD). Medical records from regional rheumatology clinics were reviewed and classified by the International Criteria for Behçet's Disease (ICBD) and International Study Group (ISG) criteria. Statistical analysis involved cross-referencing databases and applying LCA to estimate the probability of a BS diagnosis. Prevalence and incidence of BS in Sardinia were calculated, as well as the sensitivity and specificity of each database source.

Results

Administrative databases analysis retrieved 271 unique cases. A medical record review of 133 patients identified 116 BS cases, of which 107 matched administrative records and 9 were new. After excluding two deaths, 280 cases were considered for analysis (164 administrative-only, 107 mixed, 9 clinical-only). Using ICBD, the LCA confirmed 193 (68.9%) as BS cases (68.3% female, mean age 47). This yielded a 2016 Sardinian BS prevalence of 11.7 per 100,000 inhabitants (15.7 for women and 7.6 for men), with annual incidence rates ranging from 0.24 to 0.48 per 100,000 inhabitants (2014-2016). Prevalence and incidence were lower using the ISG criteria. Database sensitivity varies from 40.6% to 70.5%, while specificity ranges from 16.3% to 99.0%.

Conclusion

BS in Sardinia is a rare disease. Relying on a single data source to estimate the epidemiology of rare diseases can lead to inaccurate results.

Key words

Behçet's syndrome, epidemiology, prevalence, incidence, latent class analysis.

Matteo Piga, MD
Davide Corradini, MD*
Pietro Carboni, MD*
Alberto Floris, MD
Maria Antonietta Palmas
Silvia Sanna, MD
Giovanni Porru, MD
Gian Luca Erre, MD, PhD
Salvatore Savasta, MD, PhD
Alberto Cauli, MD, PhD

*Contributed equally.

Please address correspondence to:

Matteo Piga
Dipartimento di Scienze Mediche
e Sanità Pubblica,
Università di Cagliari, SS 554,
09042 Monserrato (CA), Italy.
E-mail: matteo.piga@unica.it

Received on September 19, 2025; accepted
in revised form on January 16, 2026.

© Copyright CLINICAL AND
EXPERIMENTAL RHEUMATOLOGY 2026.

Introduction

Behçet's syndrome (BS) is a chronic inflammatory multisystem vasculitis with unknown aetiology characterised by recurrent episodes of oral and genital aphthae, uveitis, and skin lesions. BS has a diverse geographical distribution, traditionally regarded as endemic in the "Silk Route" region, primarily due to its high prevalence in the Middle East and Far East (1). This is attributed mainly to the high prevalence of the HLA-B51 allele among individuals in these areas, which is widely recognised as the primary genetic factor contributing to susceptibility to BS (2). In Italy, BS is considered a rare disease, with an estimated prevalence of about 3 to 15 cases per 100,000 inhabitants, showing a north-south gradient in prevalence (3, 4). Sardinia is an Italian region with approximately 1.6 million inhabitants, where the regional healthcare system provides universal health coverage. Its insularity and unique genetic profile make it an ideal candidate for epidemiological studies (5, 6).

Administrative databases have increasingly been used in epidemiological research, demonstrating a promising role in some studies of rare diseases, but not in others (7). Relying solely on administrative databases for epidemiological studies may have limitations, as they often lack clinical data. Additionally, there is limited research on the accuracy of case ascertainment for BS using administrative data. Conversely, while clinical databases offer the opportunity for direct case ascertainment, they often fail to encompass large population segments and typically do not systematically collect data elements over time. Latent class analysis (LCA) is a statistical approach that estimates true disease status by integrating information from multiple imperfect data sources when no single gold standard is available.

The primary aim of the present work is to estimate the prevalence and incidence of BS in Sardinia using a latent class analysis that combines administrative databases and medical records. The secondary aim is to estimate the sensitivity and specificity of data sources to inform future epidemiologic studies on BS and other rare diseases.

Material and methods

Study design and population

Data were collected anonymously from three regional administrative databases: Hospital Discharge Records (HDRs), the Rare Diseases Regional Register (RDRR), and the Specialist Outpatient Database (SOD). Additionally, information was obtained from medical records from second and third-level rheumatology clinics operating in the region.

The HDRs include all patients and residents in Sardinia who were discharged from the hospital with a primary or non-primary ICD-9-CM diagnosis code for BS (136.1) between 2006 and 2016. HDRs for this study were obtained from the Regional Epidemiological Observatory of the Regione Autonoma della Sardegna (Sardinian Regional Government), a division of the Regional Department of Health and Hygiene responsible for managing healthcare information for epidemiological purposes. The retrieved data included repeated hospitalisations, and the number of individual patients hospitalised. The data were provided anonymously under Italian privacy and data protection laws. The RDRR includes all patients with BS registered by multidisciplinary regional referral centres from 2006 to 2016. The SOD covers patients needing outpatient visits, tests, and treatment for specific diseases. Each procedure must be coded for disease identification, the BS code is RC0120. Patients who used outpatient services with this code between 2014 and 2016 are included in the register.

The Rheumatology Units at AOU of Cagliari and AOU of Sassari are regional referral centres for rheumatic diseases, including BS. They are advanced clinics that manage the most severe cases with major organ involvement, requiring inpatient admission, while also providing outpatient services. Second-level Rheumatology clinics, known as Local Health Units (LHUs), are distributed throughout the region and generally manage mild BS cases, particularly those without major organ involvement. The medical records of every potential BS case from both second and third-level Rheumatology

Competing interests: none declared.

clinics were reviewed by experienced rheumatologists to verify whether the International Criteria for Behçet's Disease (ICBD) (8) and International Study Group (ISG) (9) criteria for BS were met.

All data underwent thorough reliability and consistency checks. This study obtained the necessary permissions (No. 9968/14) from the Regional Government and was conducted following the principles of Good Clinical Practice and the Declaration of Helsinki.

Statistical analysis

Cases were retrieved by cross-analysing administrative databases. For every case, the following information was collected: the year of the diagnosis, the patient's age at the time of the diagnosis, the patient's sex, the centre where the diagnosis was made (outpatient clinic or hospital), and the patient's urban or rural residence. Urban locations refer to the Sardinian capitals, their hinterland, and towns with more than 15,000 inhabitants. All the other locations were classified as rural. Data obtained from the three administrative databases were cross-analysed with data from the second- and third-level Rheumatology clinics operating in the region. This procedure enabled us to confirm a proportion of cases extracted from the three administrative databases that were affected by BS. To estimate the probability that the remaining cases will be confirmed as affected by BS, we used an LCA, a hierarchical regression model developed using a Bayesian approach. LCA is used when there is no gold standard to confirm cases, such as in cases ascertained from an administrative database. This approach enables the evaluation of a patient's probability of having a disease, given a set of variables. In the present study, we used the following variables: a) having been retrieved in more than one administrative database, b) having been found in an administrative database and medical records, c) the database(s) from which the case was retrieved, d) age, e) sex, and f) urban or rural residence. The LCA model was constructed using a Bayesian hierarchical framework, assuming two latent classes (BS vs. non-

Table I. Data retrieved from various databases, along with the sensitivity and specificity for each database.

	HDR	SOD	RDRR	Clinical records
Records	177	138	108	133
LCA ascertained cases	90	136	107	116
Age				
0-24	14	6	5	3
25-44	71	59	47	52
45-64	74	63	49	70
>65	18	8	6	8
Sex				
Male	53	44	34	43
Female	124	92	73	90
Residence				
Urban	107	76	60	82
Rural	70	60	47	51
Sensitivity	46.63%	70.47%	55.44%	60.10%
Specificity	16.35%	98.08%	99.04%	83.65%

HDR: hospital discharge records; RDRR: rare diseases regional register; SOD: specialist outpatient database; LCA: latent class analysis.

BS). Posterior probabilities of class membership were estimated for each included patient, and cases with a posterior probability ≥ 0.90 were classified as BS. Sensitivity analyses were performed at thresholds of 0.80 and 0.95 to assess the robustness of the classification. All analyses were performed using the XLSTAT statistical software extension for Microsoft Excel (version 2024, Addinsoft, Paris, France).

The total prevalence of BS in Sardinia in 2016 was calculated by dividing the number of confirmed cases by the population of 1,653,135, as reported by ISTAT. The annual incidence for 2014-2016 was calculated by dividing the number of new BS cases (registered in at least two of the three databases) by the population for each year (1,639,591 in 2014, 1,643,098 in 2015, and 1,653,135 in 2016). Patients who died were excluded. The sensitivity and specificity of data sources for BS case detection were also calculated.

Results

The HDRs database included 177 records, the RDRR 108, and the SOD 138. After cross-checking administrative databases and excluding duplicates, 271 cases were retrieved, of which 111 were recorded in more than one database. Medical records analysis retrieved 133 patients (102 AOU Cagliari, 25 AOU Sassari, and 6 from LHU) of whom 116 were confirmed

as affected by BS using the ICBD and 73 using ISG criteria after case ascertainment by experienced rheumatologists (MP, AF). In total, 107 of 116 for ICBD and 62 of 73 for ISG matched administrative records, while 9 were not retrieved from administrative databases. Two patients died before 2016 and were excluded from the analysis. In total, 280 cases were retrieved; 164 cases were retrieved only in administrative databases, 107 cases in both administrative and clinical records, and 9 cases only in clinical records. Table I presents data from various databases, along with the sensitivity and specificity of each database for BS diagnosis using the ICBD.

LCA identified 193 patients with a high probability of BS, whereas 87 cases (31.1%) were unlikely to be classified as BS and were excluded from the prevalence estimate. Of the 193 patients estimated to be affected by BS, 132 (68.3%) were female, with a mean age of 47 years at the time of case extraction. A mild predominance of BS diagnosis was observed in patients from urban areas (57.0%) compared to those from rural areas (43.0%). Having been retrieved in more than one administrative database and having been found in an administrative database and in a medical record conferred 60.3% and 92.2% probability of being classified as BS by LCA, respectively. Having been retrieved only in HDR

conferred no probability (0%) of being classified as BS by LCA. Figure 1 shows the probability for each indicator based on the LCA results.

Accordingly, using the ICBBD, the 2016 prevalence of BS in Sardinia was estimated at 11.7 cases per 100,000 inhabitants, with a 95% confidence interval (CI) of 10.2–13.3. Differentiating by patients' sex, the 2016 estimated prevalence of BS was 15.7 per 100,000 women (95% CI: 13.2–18.6) and 7.6 per 100,000 men (95% CI: 5.8–9.7). The estimated incidence rates in 2014, 2015, and 2016 were 0.48 new cases per 100,000 inhabitants (95% CI: 0.35–0.63), 0.24 new cases per 100,000 inhabitants (95% CI: 0.15–0.35), and 0.36 new cases per 100,000 inhabitants (95% CI: 0.25–0.49), respectively. Using the ISG criteria, the 2016 prevalence estimate of BS in Sardinia was 7.4 cases per 100,000 inhabitants (95% CI: 2.1–9.5). The estimated incidence rates in 2014, 2015, and 2016 were 0.30 new cases per 100,000 inhabitants (95% CI: 0.21–0.40), 0.15 new cases per 100,000 inhabitants (95% CI: 0.09–0.23), and 0.23 new cases per 100,000 inhabitants (95% CI: 0.15–0.32), respectively.

Discussion

This study reports on the prevalence and incidence of BS in Sardinia, showing figures at the high end of the range observed in other European regions, but lower than those reported from regions endemic for BS. Indeed, the prevalence of BS is 320 cases per 100,000 inhabitants in Turkey, 80 cases per 100,000 inhabitants in Iran, and 20 cases per 100,000 inhabitants in Saudi Arabia (1). In Europe, BS prevalence increases from north to south and has increased recently due to genetic changes associated with migration and more sensitive ICBBD criteria (10). ICBBD (8) are more sensitive but less specific than the ISG criteria (9), leading to more BD diagnoses and a higher observed prevalence compared to previous studies using ISG criteria (3, 4). Our study confirmed the generally higher sensitivity of the new criteria by showing lower incidence and prevalence with the ISG criteria. In Italy, BS is rare, and our data support this, but prevalence varies by region.

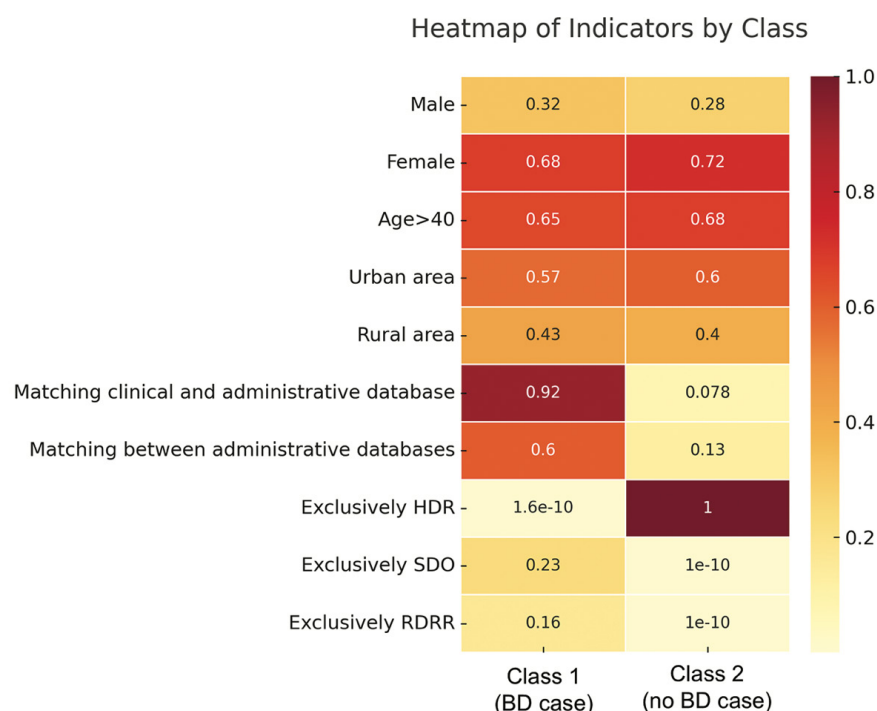


Fig. 1. Heatmap displaying the conditional probabilities for each indicator in class 1 (BS case) and class 2 (no BS case). Light colours represent lower probabilities, while darker colours represent higher probabilities.

Potenza in Southern Italy has the highest prevalence in Europe at 15.9 per 100,000 (95% CI 8.9–28.5) (3), while Reggio Emilia in Northern Italy reports a lower prevalence of 3.8 per 100,000 (95% CI 2.0–5.8) (4). The number of cases identified according to ISG criteria (9) was 11 and 18, respectively, which is a low number compared to the number of cases we identified. Indeed, thanks to our mixed approach, we were able to cover a larger area and population, which allowed us to identify a total of 193 BS cases. The prevalence of HLA-B51, a widely recognised genetic predisposing factor for BS, is approximately 14% in the general population of Sardinia (11), slightly higher than the Italian prevalence of around 10% (21). In the present study, we found a higher prevalence of BS in females, with a male-to-female ratio of 1:2, which contrasts with data from regions where BS is endemic, reporting that BS diagnosis is more common in males than in females (1), and suggests that European cohorts may have distinct phenotypic and gender distributions. A further possible reason for the higher prevalence among women is the mixed-methods approach used in our analysis, which

may have detected less severe cases more accurately. These cases often affect females (13) and may be missed in studies that rely on a single data source (14). Epidemiology studies estimate disease burden and provide insights into pathogenesis by highlighting genetic and environmental factors (15). This is particularly important in conditions such as BS, which exhibit geographical variability and diverse clinical features (16). However, differences in epidemiological methods across studies complicate comparisons (14). Specifically, a single clinical-based approach may be more accurate and provide a better description of the prevalence of clinical signs and genetic patterns. However, it could be counterproductive for epidemiological purposes because the larger the area of interest, the higher the likelihood of missing patients, especially in a heterogeneous disease such as BS. Alternatively, using administrative databases for research enables the study of large population samples. However, these databases often lack clinical data and may underestimate the prevalence of mild disease. This is highlighted in a meta-analysis (17), which estimated a 22-fold higher prevalence of BS in pop-

ulation surveys compared to census surveys. Census surveys may not capture BS patients who do not exhibit major organ involvement. In contrast, population surveys may overestimate the number of patients with a mild phenotype. This is also more evident in studies based on hospitalisation records, which may underestimate prevalence and overestimate the age of incidence (18). To address these issues, we adopted a mixed approach that combined clinical data and three administrative databases, which is a strength of this study. A proportion of cases in the administrative databases were clinically confirmed. LCA was used to estimate the probability of true BS cases. Indeed, this method offers a flexible framework for prevalence studies that integrates administrative and clinical data to estimate true disease prevalence despite the absence of a gold standard, while accounting for measurement errors and misclassification. The choice of variables included in the model, such as demographic characteristics, concordance between administrative and clinical records, and the type of database of origin, was guided by previous methodological studies that have applied LCA in the scientific literature (19). The hierarchical structure enables data clustering and covariate effect modelling, while the Bayesian approach incorporates prior knowledge and quantifies uncertainty. This study has some limitations. The absence of longitudinal data limits the ability to evaluate changes in disease prevalence and incidence over time, monitor disease progression, and understand the dynamic interaction of risk factors and outcomes. Another limitation, related to administrative databases, is that they cannot determine whether there is variability in the severity of BS manifestations or differences in organ involvement between or within patient subgroups. Finally, we found a very low sensitivity for HDR. Although this data should be interpreted with caution, some possible explanations include the

fact that BS is a complex and diverse disease, with no gold-standard diagnostic biomarkers that may mimic various other conditions and often require clinical expertise and extended observation before an accurate diagnosis can be made. This means that a patient's hospital stay might not be the point at which a definitive diagnosis is reached.

In conclusion, BS in Sardinia is a rare disease. Our approach demonstrated low accuracy in estimating rare disease prevalence from a single database, underscoring that reliance on a single source can lead to inaccuracies. Future studies should consider these findings.

Acknowledgements

The authors thank Gigi Marras for his valuable informatics support.

References

1. YAZICI H, SEYAHİ E, HATEMI G, YAZICI Y: Behçet syndrome: A contemporary view. *Nat Rev Rheumatol* 2018; 14: 107-19. <https://doi.org/10.1038/nrrheum.2017.208>
2. PIGA M, MATHIEU A: Genetic susceptibility to Behçet's disease: role of genes belonging to the MHC region. *Rheumatology (Oxford)* 2011; 50: 299-310. <https://doi.org/10.1093/rheumatology/keq331>
3. OLIVIERI I, LECCESI P, PADULA A *et al.*: High prevalence of Behçet's disease in southern Italy. *Clin Exp Rheumatol* 2013; 31 (Suppl. 77): S28-31.
4. SALVARANI C, PIPITONE N, CATANOSO MG *et al.*: Epidemiology and clinical course of Behçet's disease in the Reggio Emilia area of northern Italy: A seventeen-year population-based study. *Arthritis Care Res* 2007; 57: 171-78. <https://doi.org/10.1002/art.22500>
5. PIGA M, CASULA L, SANNA S *et al.*: Population-based analysis of hospitalizations for patients with systemic sclerosis in a West-European region over the period 2001-2012. *Rheumatol Int* 2016; 36: 73-81. <https://doi.org/10.1007/s00296-015-3330-1>
6. PIGA M, CASULA L, PERRA D *et al.*: Population-based analysis of hospitalizations in a West-European region revealed major changes in hospital utilization for patients with systemic lupus erythematosus over the period 2001-2012. *Lupus* 2016; 25: 28-37. <https://doi.org/10.1177/0961203315596597>
7. DESAI RJ, SOLOMON DH: Use of large healthcare databases for rheumatology clinical research. *Curr Opin Rheumatol* 2017; 29: 138-43. <https://doi.org/10.1097/bor.0000000000000363>
8. DAVATCHI F, ASSAAD-KHALIL S, CALAMIA KT *et al.*: The International Criteria for Behçet's Disease (ICBD): A collaborative study of 27 countries on the sensitivity and specificity of the new criteria. *J Eur Acad Dermatol Venereol* 2014; 28: 338-47. <https://doi.org/10.1111/jdv.12107>
9. Criteria for diagnosis of Behçet's disease. International Study Group for Behçet's Disease. *Lancet* 1990; 335(8697): 1078-80. [https://doi.org/10.1016/0140-6736\(90\)92643-v](https://doi.org/10.1016/0140-6736(90)92643-v)
10. LORENTSEN JLL, BRUNBORG C, AASEBØ AT *et al.*: Incidence and prevalence of Behçet's disease in Oslo Norway: a two-decade population-based analysis. *Rheumatology (Oxford)* 2025; 64(8): 4703-12. <https://doi.org/10.1093/rheumatology/keaf224>
11. PIGA M, PALADINI F, LAI S *et al.*: Genetics of Behçet's disease in Sardinia: two distinct extended HLA haplotypes harbour the B*51 allele in the normal population and in patients. *Clin Exp Rheumatol* 2012; 30 (Suppl. 72): S51-56.
12. SALVARANI C, BOIARDI L, MANTOVANI V *et al.*: Association of MICA alleles and HLA-B51 in Italian patients with Behçet's disease. *J Rheumatol* 2001; 28: 1867-70.
13. BONITSIS NG, LUONG NGUYEN LB, LAVALLEY MP *et al.*: Gender-specific differences in Adamantiades-Behçet's disease manifestations: an analysis of the German registry and meta-analysis of data from the literature. *Rheumatology (Oxford)* 2015; 54: 121-33. <https://doi.org/10.1093/rheumatology/keu247>
14. HATEMI G, ESATOGLU SN, YURDAKUL S: Considerations in designing and interpreting prevalence studies for Behçet syndrome. *Polish Arch Intern Med* 2018; 128: 148-49. <https://doi.org/10.20452/pamw.4233>
15. PEARCE N: Classification of epidemiological study designs. *Int J Epidemiol* 2012; 41: 393-97. <https://doi.org/10.1093/ije/dys049>
16. EMMI G, BETTIOL A, HATEMI G, PRISCO D: Behçet's syndrome. *Lancet* 2024; 403(10431): 1093-108. [https://doi.org/10.1016/S0140-6736\(23\)02629-6](https://doi.org/10.1016/S0140-6736(23)02629-6)
17. MALDINI C, DRUCE K, BASU N, LAVALLEY MP, MAHR A: Exploring the variability in Behçet's disease prevalence: a meta-analytical approach. *Rheumatology (Oxford)* 2018; 57: 185-95. <https://doi.org/10.1093/rheumatology/kew486>
18. KURAL-SEYAHİ E, FRESKO I, SEYAHİ N *et al.*: The long-term mortality and morbidity of Behçet syndrome: a 2-decade outcome survey of 387 patients followed at a dedicated center. *Medicine (Baltimore)* 2003; 82: 60-76. <https://doi.org/10.1097/00005792-200301000-00006>
19. MICHELENA X, SEPRIANO A, ZHAO SS *et al.*: Exploring the unifying concept of spondyloarthritis: a latent class analysis of the REGISPONSER registry. *Rheumatology (Oxford)* 2024; 63(11): 3098-105. <https://doi.org/10.1093/rheumatology/keaf005>