

Historical Silk Road Connectivity and the Geographic Distribution of Behçet's Disease: A Cross-National Ecological Analysis

Hanying Wang^{1, *}

¹ Hangzhou Foreign Languages School, 309 Liuhe Road, Xihu District, Hangzhou, Zhejiang, China

* Corresponding Author Email: Avoria_wong@hotmail.com

Abstract. Behçet's disease (BD) is a systemic inflammatory disorder with a distinct geographic distribution along the Silk Road. Although HLA-B51 is regarded as the strongest known genetic susceptibility factor, differences in BD prevalence among populations cannot be explained solely by genetics, and the influence of long-term historical connectivity remains unclear. To address this gap, we constructed a cross-national ecological dataset of 20 countries and developed a Silk Road Exposure Index (SREI) based on distance to the reconstructed Silk Road and territorial overlap with a 500-km corridor buffer. Layered ordinary least squares (OLS) regression models evaluated the association between SREI and BD prevalence, sequentially adjusting for HLA-B51 frequency, latitude, Human Development Index, and infectious burden. Robustness analyses included Cook's distance, leave-one-out analyses, Negative Binomial regression, and Spearman correlation. In the minimally adjusted model, SREI was positively associated with log-transformed BD prevalence ($\beta = 0.3703$, $P = 0.021$; $R^2 = 0.264$), but the association weakened after full adjustment ($\beta = 0.1880$, $P = 0.458$). Leave-one-out analyses showed a consistently positive effect direction, while Spearman correlation remained significant ($\rho = 0.568$, $P = 0.009$). These findings provide a quantitative framework for evaluating the Silk Road hypothesis in BD epidemiology and suggest that historical connectivity may reflect long-term interactions among population movement, genetic susceptibility, and environmental influences rather than acting as an independent determinant.

Keywords: Behçet's disease; Silk Road; Spatial epidemiology.

1. Introduction

Behçet's disease is a multisystem inflammatory disorder characterized by recurrent oral and genital aphthosis, uveitis, and diverse skin ulcers [1]. It shows a distinctive geographic distribution along the Silk Road. Higher BD prevalence is reported along the historical Silk Road corridor from the Mediterranean region to East Asia. That is why it is called the "Silk Road Disease" [2,3]. For years, the mechanisms of this remarkable geographic clustering and the decline of prevalence outside this region remain poorly understood [2].

Among known genetic factors, the human leukocyte antigen (HLA)-B51 allele is the strongest established susceptibility factor for BD [4]. Its frequency shows consistent associations across diverse populations, but global epidemiological patterns indicate that HLA-B51 frequency alone cannot fully explain disease distribution [1, 4]. Some populations with relatively high HLA-B51 frequency show limited BD prevalence, while some Mediterranean and Middle Eastern populations show disease burdens exceeding predictions based simply on genetic susceptibility. These observations suggest that additional factors like environmental, historical, and demographic factors may contribute to the formation of BD geographic hotspots.

Currently, there are several environmental hypotheses, including infectious exposure, microbial stimulation, solar radiation, and climatic factors represented by latitude [5, 6]. However, these existing geographic models focus mainly on contemporary environmental variables and ignore long-term human connectivity. For example, the Silk Road is a unique historical network that shaped migration, trade, and population interactions across Eurasia for more than two millennia. Also, despite the Silk Road exposure is frequently described as a characteristic feature of BD epidemiology, it is a qualitative concept rather than a measurable epidemiological variable. Furthermore, historical

connectivity's independent contribution beyond genetic susceptibility, socioeconomic development, and modern infectious burden remains unclear [1].

To address this gap, we developed the Silk Road Exposure Index. It is quantitative, and it integrates geographic proximity to the Silk Road and the territorial overlap with historical trade routes. This study constructed a cross-national dataset of 20 countries. The aims are to (1) evaluate the association between BD prevalence and SREI after adjusting for HLA-B51 frequency, latitude, HDI, and infectious burden; (2) assess the robustness and interpretative implications of historical connectivity using sensitivity analyses.

2. Methods

2.1. Data Acquisition

We constructed a cross-national ecological dataset containing 20 countries to represent diverse geographic relationships with the historical Silk Road corridor. In the choice of countries, we considered both countries' Silk Road proximity and their territorial exposure. The final country set and their corresponding data sources are summarized in Table 1.

Table 1. Baseline characteristics and covariates of the 20 study nations included in the ecological analysis

Country	BD Prevalence	HLA-B*51 Frequency	Infectious Burden	Latitude	Reference
China	10	0.0692	1190.4	36.904	[12]
Israel	15.2	0.0422	1189.5	31.446	[13]
France	7.1	0.0771	1337.3	46.706	[3]
South Korea	26.2	0.098	1356.2	36.46	[14]
Germany	4.7	0.0619	1255.6	51.085	[15]
Sweden	4.9	0.0533	1141.8	62.195	[16]
Turkey	119.8	0.1473	2637.4	38.961	[6]
Spain	0.66	0.0801	1163.8	39.899	[17]
United Kingdom	19.07	0.054	1496	54.314	[18]
Iraq	17	0.1345	5066.2	33.221	[19]
Italy	15.9	0.104	1035.9	42.492	[20]
Switzerland	4.03	0.0753	1327.8	46.812	[21]
Iran	80	0.121	2181.9	32.422	[22]
Netherlands	1	0.057	1216.1	52.106	[23]
Saudi Arabia	20	0.1647	2204.4	24.244	[24]
Portugal	2.4	0.1059	1640.1	39.565	[25]
Jordan	660	0.099	2668.9	31.27	[26]
United States of America	5.2	0.0493	1494.1	37.246	[27]
Egypt	3.6	0.0605	4153.9	26.824	[8, 28]
Japan	13.5	0.1061	1633.1	43.462	[29]

Notes: BD prevalence is reported as cases per 100,000 population. HLA-B51 frequency represents population-level allele frequency; HLA-B51:01 frequency was used as a proxy when general HLA-B51 data were unavailable. Infectious burden represents CMNN disease-related DALYs from the Global Burden of Disease Study 2019. Latitude represents the latitude of country representative points.

To obtain Behçet's disease prevalence estimates for different countries, we systematically searched and compiled published epidemiological studies. For countries with multiple estimates, we prioritize population-based studies considering study design and population representativeness [1, 5]. When national estimates were unavailable, the most appropriate population-based or regional estimates were retained. All prevalence estimates were expressed as cases per 100,000 population for cross-country comparison.

Population-level HLA_B51 allele frequencies were mainly from Allele Frequency Net Database (AFND) [7]. When population-level HLA-B51 estimates were unavailable, HLA-B51:01 frequency

was used as a proxy. For Egypt, HLA-B51 frequency was obtained from a published Egyptian HLA typing study due to the lack of available AFND data [8].

The Human Development Index (HDI) was included as a socioeconomic variable [9]. It was used to explain the differences in socioeconomic development and healthcare access that may influence disease detection and reporting across countries. Apart from that, we used disability-adjusted life years due to communicable, maternal, neonatal, and nutritional diseases from the Global Burden of Disease Study 2019 to represent contemporary infectious disease burden [10, 11]. Latitude was calculated in QGIS (version 4.0.0). We used the feature “Points on Surface” to obtain points representing each country. Latitude gained was the latitude of these points. These variables were processed as covariates instead of components of the Silk Road Exposure Index, allowing historical connectivity measures to be evaluated independently from contemporary socioeconomic and disease-burden differences.

2.2. Geospatial Construction: Distance and Route Density (QGIS)

We firstly used QGIS to calculate two geospatial predictors, Distance to Silk Road and Route Density [30]. Then we integrated them into the composite Silk Road Exposure Index described in Section 2.3. Distance quantified the geographic proximity of countries to the Silk Road, while Route Density measured the proportion of national territory that overlapped with the defined corridor.

The silk road corridor was generated by connecting historical trade hubs along the principal overland network, including Xi'an, Dunhuang, Kashgar, Samarkand, Tehran, Baghdad, Ankara, and Istanbul. We also obtained the country boundary polygons from Natural Earth administrative boundary data [31]. We then used the reconstructed route as the spatial reference for calculating distance and Silk Road exposure metrics.

Distance calculations used an azimuthal equidistant (AEQD) projection centered on the Eurasian landmass to minimize distortion in distance measurements across the transcontinental study area. We measured the shortest planar distance between the centroids for each country and the reconstructed Silk Road in kilometers.

To calculate Route Density, we constructed a 500-km buffer surrounding the reconstructed Silk Road route [32]. Then we intersected the buffer region with each country's boundary polygon. The overlapping area was quantified using area-preserving measurements. The Route Density was defined as:

$$\text{Route Density} = \frac{\text{Area}(\text{Country} \cap \text{Silk Road Corridor})}{\text{Area}(\text{Country})} \quad (1)$$

where $\text{Area}(\text{Country} \cap \text{Silk Road Corridor})$ represents the intersection area of country and the Silk Road corridor buffer, and $\text{Area}(\text{Country})$ represents the area of each country.

This definition produced a normalized metric ranging from 0 to 1, allowing comparison among countries of different geographic sizes. Higher values represent higher overlapping between territory and the reconstructed Silk Road, while lower values indicated a lower spatial overlap.

2.3. SREI Construction

The Silk Road Exposure Index was constructed as a composite measure that integrate two complementary dimensions: geographic proximity to the Silk Road corridor and the territorial overlap with the corridor network. Because no external gold-standard measure exists, equal weighting was applied based on a theory-driven approach.

Distance to Silk Road and Route Density were regarded as complementary factors instead of hierarchical components. Distance reflected that interaction probability decreased with increasing geographic separation, while Route Density showed the extent to which a country's territory overlapped with the reconstructed trade corridor [33]. Therefore, we assigned equal weights to both

dimensions to avoid assuming a disproportionate distribution of both factors to historical exposure [34].

Considering nonlinear spatial decay, Distance to Silk Road was first log-transformed:

$$D' = \ln(\text{Distance} + 1) \quad (2)$$

The inverse distance component was then calculated:

$$-D' \quad (3)$$

Both Route Density and inverse log-distance were standardized using z-score transformation:

$$Z(x) = \frac{x - \bar{x}}{\sigma_x} \quad (4)$$

where $\bar{x}$ represents the sample mean and σ_x represents the standard deviation.

The final SREI was calculated as:

$$\text{SREI} = 0.5 \times Z(\text{RouteDensity}) + 0.5 \times Z[-\ln(\text{Distance} + 1)] \quad (5)$$

Higher SREI values indicated greater estimated historical connectivity with the Silk Road corridor.

2.4. Statistical Analysis and Model Architecture

All statistical analysis were conducted in Python 3.9 within a Jupyter Notebook environment. Before regression modeling, we $\log_{10}(x)$ transformed BD prevalence to correct right-skewness [35].

To separate the independent contribution of historical Silk Road connectivity from established genetic and geographic predictors of BD prevalence, we performed a hierarchical sequence of Ordinary Least Squares models. Model 1 included SREI alone; Model 2 was additionally adjusted for HLA-B51; Model 3 added latitude; and Model 4 included all covariates (HLA-B51, latitude, HDI, and infectious burden).

2.5. Outlier Diagnostics, Sensitivity Analysis, and Confounder Control

To examine whether the observed association between Silk Road exposure and Behçet's disease prevalence was influenced by specific individual countries or modelling assumptions, we performed a series of diagnostic analyses and sensitivity analyses. Because the dataset contained only 20 countries, and BD prevalence varied substantially between populations, we performed additional analyses to evaluate the stability of the main regression results.

First, we used Cook's Distance to evaluate the influential observations. This measures the contribution of individual observations to the whole coefficient estimates [36]. For countries showing high influence, we further performed leave-one-out analyses [37]. This checked whether the SREI-BD association was driven by individual high-prevalence countries.

BD is a rare event, and its prevalence reported varies substantially across different countries. Therefore, we performed additional sensitivity analyses to determine whether the direction of associations remained consistent under an alternative statistical assumption. We used negative binomial regression as a complementary sensitivity analysis [38]. It was not replacing the primary regression analysis.

Finally, to assess the monotonic relationship between SREI ranking and BD prevalence ranking, we calculated Spearman rank correlation. This approach provided an additional evaluation that was both less dependent on linearity assumptions and less influenced by extreme observations [39].

The overall workflow is shown in Figure 1.

Analytical workflow for evaluating the association between historical Silk Road connectivity and Behçet's disease prevalence

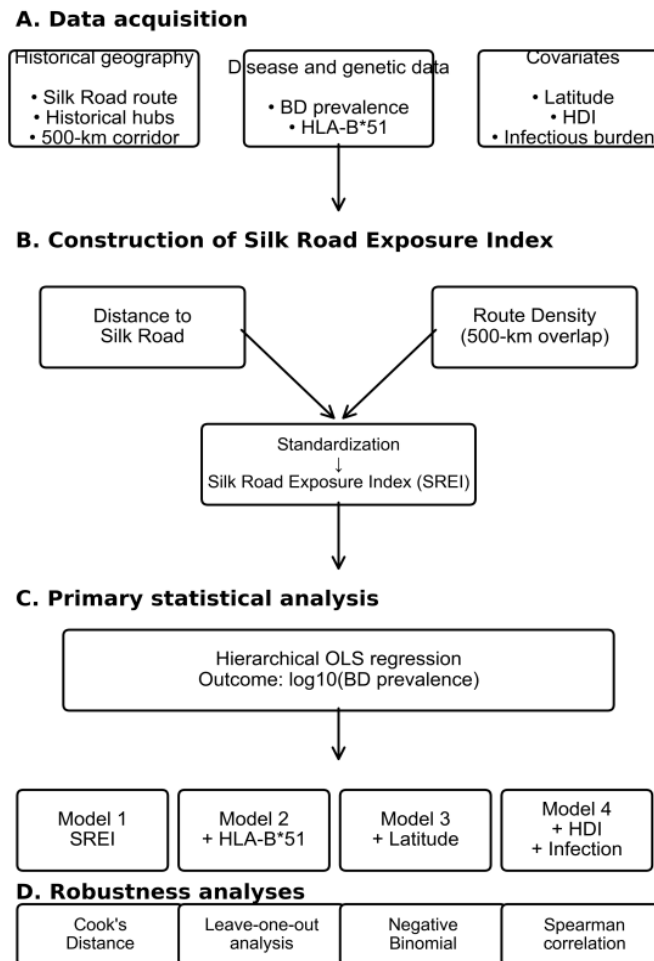


Figure 1. Analytical workflow for evaluating the association between historical Silk Road connectivity and Behçet's disease prevalence. Analytical workflow showing data integration, SREI construction, regression modeling, and sensitivity analyses. (Picture credit: Original)

All statistical analyses were performed using Python (version 3.9) with the statsmodels and scipy libraries [40, 41]. Statistical significance was defined as $P < 0.05$, and values between 0.05 and 0.10 were considered as suggestive associations.

3. Results

3.1. Cohort Selection and Geospatial Distribution

There were 20 representative countries selected in the primary analysis. The geographic distribution of Behçet's disease prevalence showed a clustered pattern along the historical Silk Road region. Higher prevalence estimates were observed in Jordan (660 per 100,000), Turkey (119.8 per 100,000), and Iran (80 per 100,000), whereas lower prevalence was observed in countries distant from the Silk Road corridor (Figure 2). This provides a preliminary spatial basis for the historical connectivity hypothesis [34].

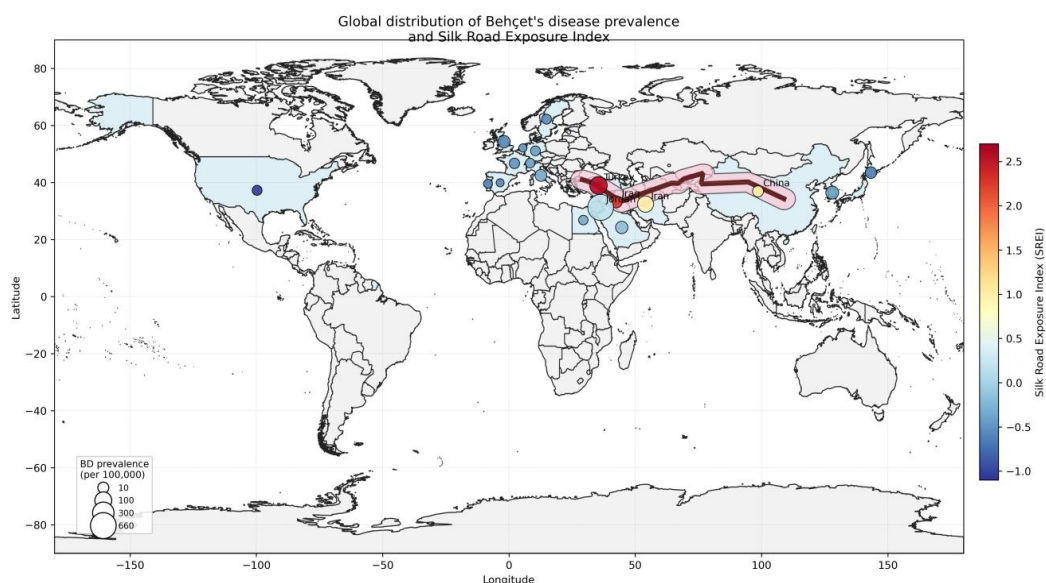


Figure 2. Global geospatial distribution of BD prevalence mapped against SREI gradients. Red circles represent national prevalence magnitude, while the shaded region indicates the 500 km historical trade corridor buffer (Picture credit: Original)

3.2. Geographic Characterization of Silk Road Exposure

Countries located near the central Eurasian corridor showed the highest levels of historical exposure. For example, Turkey, Iran, China, and Jordan were close to the Silk Road, and they showed high Route Density. In contrast, countries that were relatively far from the Silk Road, including Sweden and the United States, had greater distances to the Silk Road and almost zero territorial overlap with the corridor.

Detailed Distance to Silk Road, Route Density, and SREI values for all countries are provided in Table 2.

Table 2. Geographic characterization of Silk Road exposure metrics across the 20 study countries.

Country	Distance to Silk Road (km)	Route Density	SREI
China	185.863	0.277	0.930297266
Israel	708.307	0	-0.05397311
France	2230.231	0	-0.488659594
South Korea	1748.533	0	-0.396411897
Germany	1806.852	0	-0.408849363
Sweden	2536.564	0	-0.537457073
Turkey	30.768	0.856	2.602188219
Spain	2737.386	0	-0.566346193
United Kingdom	2716.009	0	-0.563373592
Iraq	120.651	0.955	2.263852701
Italy	1359.939	0	-0.301141385
Switzerland	1763.4	0	-0.399621462
Iran	432.096	0.47	0.94473035
Netherlands	2163.585	0	-0.477157374
Saudi Arabia	1008.518	0.02	-0.153308031
Portugal	3146.691	0	-0.619182982
Jordan	638.376	0.13	0.209877199
United States of America	9890.078	0	-1.053468802
Egypt	1409.755	0	-0.314777296
Japan	3130.423	0	-0.617217583

Notes: Distance to Silk Road represents the shortest geographic distance between each country centroid and the reconstructed Silk Road corridor, measured in kilometers. Route Density represents the proportion of national territory overlapping with the 500-km Silk Road corridor buffer. The Silk Road Exposure Index (SREI) integrates standardized Route Density and inverse log-transformed Distance to Silk Road using equal weighting. Higher SREI values indicate greater estimated historical connectivity with the Silk Road corridor.

3.3. Primary Multivariable Regression and Predictor Stability

To evaluate whether historical Silk Road exposure had independent association with the geographic distribution of BD prevalence, we performed a series of OLS regression models. Log-transformed BD prevalence was the outcome variable (Table 3).

Table 3. Hierarchical multivariable regression analysis of factors associated with Behçet's Disease prevalence.

Variables	Model 1	Model 2	Model 3	Model 4
SREI	0.3703 (0.1458) / 0.021	0.2588 (0.1765) / 0.161	0.2494 (0.1795) / 0.1840.1880 (0.2460) / 0.458	
HLA-B51 Frequency	---	5.4299 (4.9110) / 0.284	3.9776 (5.3730) / 0.4705.6441 (5.8818) / 0.354	
Latitude	---	---	-0.0120 (0.0167) / 0.481	-0.0056 (0.0223) / 0.804
Infectious_burden	---	---	---	-0.0002 (0.0003) / 0.347
HDI	---	---	---	-3.7440 (4.2052) / 0.388
R ²	0.264	0.313	0.335	0.382
Adjusted R ²	0.223	0.232	0.210	0.161
F-statistic	6.453	3.878	2.686	1.727

Abbreviations: SREI, Silk Road Exposure Index; SE, Standard Error; HLA, Human Leukocyte Antigen.

Note: All independent variables and the dependent variable (Prevalence) underwent log-transformation to minimize skewness. Bold values indicate significance at $P < 0.05$; italic values indicate a suggestive trend at $P < 0.10$.

In the unadjusted model (Model 1), SREI showed a significant positive association with BD prevalence ($\beta = 0.3703$, $SE = 0.1458$, $P = 0.021$). This model explained 26.4% of the variance in log-transformed BD prevalence ($R^2 = 0.264$; adjusted $R^2 = 0.223$).

After sequential adjustment for HLA-B51 frequency and latitude (Models 2–3), the association between SREI and BD prevalence remained positive but became statistically non-significant (Model 2: $\beta = 0.259$, $P = 0.161$; Model 3: $\beta = 0.249$, $P = 0.184$). Neither HLA-B51 frequency nor latitude showed significant associations with BD prevalence.

In the fully adjusted model (Model 4), the association between SREI and BD prevalence was further weakened and remained non-significant ($\beta = 0.188$, $P = 0.458$). None of the additional covariates showed significant associations. Although this model explained the highest proportion of variance ($R^2 = 0.382$), the adjusted R^2 decreased to 0.161, suggesting reduced model stability due to the limited sample size.

Overall, hierarchical models indicated that the association between SREI and BD prevalence was strongest in the unadjusted model but attenuated after adjustment for genetic, geographic, socioeconomic, and infectious disease-related factors.

3.4. Influence Diagnostics and Sensitivity Analysis of Silk Road Exposure Association

According to Cook's distance analyses, Iraq and Jordan showed the highest influence values and exceeded the conventional threshold for influential observations ($4/n = 0.20$). Thus, they were identified as influential observations and were further evaluated in sensitivity analyses (Figure 3)

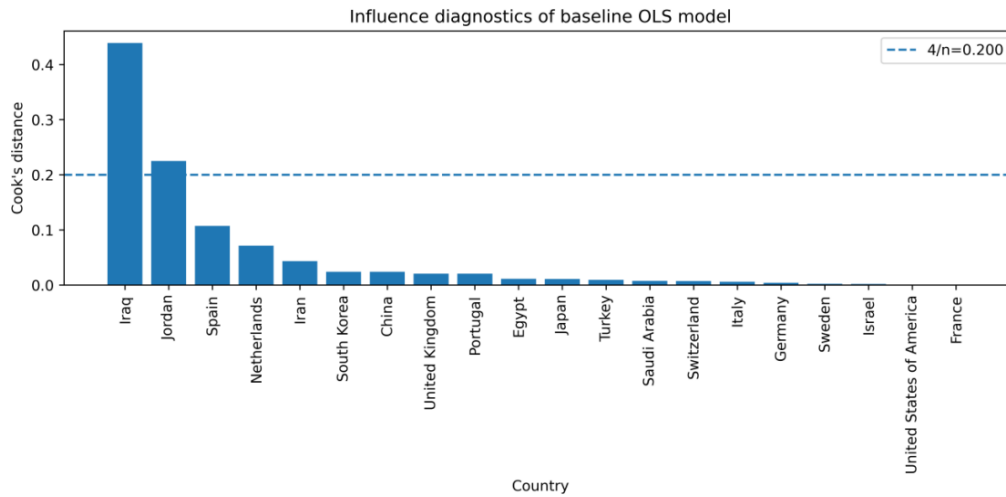


Figure 3. Cook’s distance diagnostics identifying influential observations in the fully adjusted OLS model. Cook’s distance was calculated from the final multivariable ordinary least squares (OLS) regression model including SREI, HLA-B51 frequency, latitude, HDI, and Infectious burden as predictors. The dashed line represents the conventional influence threshold ($4/n$, where $n = 20$). Countries with relatively higher Cook’s distance values were further evaluated in sensitivity analyses to determine whether the observed SREI–Behçet’s disease prevalence association depended on individual high-leverage observations (Picture credit: Original)

We subsequently conducted a layered sensitivity analysis to assess the stability of the SREI–BD association (Table 4). In the baseline model (Tier 1, $n=20$), the association between SREI and BD prevalence was positive and significant ($\beta = 0.370$, $P = 0.021$; adjusted $R^2 = 0.223$). After removing Iraq (Tier 2A, $n = 19$), the association was not weakened, instead, it was strengthened ($\beta = 0.468$, $P = 0.002$; adjusted $R^2 = 0.238$). Furthermore, after removing both Iraq and Jordan (Tier 2B, $n = 18$), we still got a positive association ($\beta = 0.450$, $P = 0.003$; adjusted $R^2 = 0.391$). This finding indicated that the observed association was not solely driven by these high-leverage observations.

Table 4. Results of Tiered Sensitivity Analysis and Influential Observation Exclusion Tests

Model Tier	Sample Description	Sample Size (n)	SREI Coefficient (β)	SREI P-value	Adjusted R^2	Key Findings
Tier 1	Full Cohort (OLS)	20	$\beta = 0.370$	0.020512	0.223000	Baseline association
Tier 2A	Excluding Iraq (OLS)	19	$\beta = 0.468$	0.019664	0.238302	Iraq influence tested
Tier 2B	Excluding Iraq and Jordan (OLS)	18	$\beta = 0.450$	0.003279	0.391043	Iraq and Jordan influence tested
Tier 3	Full Cohort (Spearman)	20	$\rho = 0.623$	0.003372	NaN	Non-parametric validation

Notes: Tiered sensitivity analyses were performed to evaluate whether the association between Silk Road Exposure Index and Behçet’s disease prevalence was influenced by individual high-leverage observations. Tier 1 represents the OLS model with full cohort. Tier 2A represents exclusion analyses removing Iraq, and Tier 2B represents the OLS model removing both Iraq and Jordan. β represents the estimated SREI coefficient from OLS regression models, and ρ represents Spearman’s rank correlation coefficient. Adjusted R^2 was reported only for OLS models and was not applicable to Spearman correlation analysis.

3.5. Robustness and Sensitivity Check

To evaluate whether the association between Silk Road exposure and Behçet’s disease was consistent under different analytical assumptions, we conducted several sensitivity analyses, including leave-one-out analysis, Negative Binomial regression, and Spearman correlation analysis.

First, we conducted a leave-one-out sensitivity analysis based on the unadjusted OLS regression model. Then, we removed each country one by one and calculated the SREI coefficient. This process aimed to determine whether the observed association relied on any individual geographic observation. Across all the 20 leave-one-out iterations, the direction of the SREI association always kept positive, showing that the geographic gradient observed was not directly influenced by a single country. The estimated SREI coefficient ranged from $\beta = 0.351$ after removing of Turkey to $\beta = 0.496$ after removing

Iraq. When Iraq was removed, we obtained the strongest positive association ($\beta = 0.496$, $P = 0.010$; adjusted $R^2 = 0.291$). While when Turkey was removed, the association was weakened significantly ($\beta = 0.351$, $P = 0.088$; adjusted $R^2 = 0.112$). These results suggested that although countries located along the central Silk Road corridor contributed substantially to the observed spatial pattern, the positive relationship between SREI and BD prevalence remained directionally consistent after removal of individual countries.

Next, we performed series of Negative Binomial models (Table 5). In the unadjusted Negative Binomial model (NB Model 1), SREI and BD case counts showed a positive association ($\beta = 1.4893$, $SE = 0.7496$, $P = 0.0469$). When HLA-B51 frequency (NB Model 2), latitude (NB Model 3), and HDI and infectious disease burden (NB Model 4) were sequentially adjusted, the estimated effect of SREI decrease gradually. In the fully adjusted model including all the factors above, SREI coefficient became negative and non-significant ($\beta = -0.8516$, $SE = 0.5579$, $P = 0.1269$). Taking together, these findings suggested that the positive association observed in the unadjusted model was sensitive to high covariate adjustment.

Table 5. Comparison of ordinary least squares and Negative Binomial regression models for evaluating the association between Silk Road exposure and Behçet's disease burden

Model	Framework	Outcome	Predictors	$\beta(SE)/P$
OLS Model 1	OLS	$\log_{10}(\text{BD prevalence})$	SREI	0.3703 (0.1458) / 0.0205
NB Model 1	Negative Binomial	BD case counts	SREI	1.4893 (0.7496) / 0.0469
NB Model 2	Negative Binomial	BD case counts	SREI + HLA-B51	0.8631 (0.8952) / 0.335
NB Model 3	Negative Binomial	BD case counts	SREI + HLA-B51 + Latitude	0.4916 (0.684) / 0.4723
NB Model 4	Negative Binomial	BD case counts	SREI + HLA-B51 + Latitude + Infection + HDI	-0.8516 (0.5579) / 0.1269

Notes: Negative Binomial models used country-level BD case counts as the outcome with log-transformed population size included as an offset term. Models were sequentially adjusted for genetic background (HLA-B51 frequency), geographic latitude, and contemporary environmental and socioeconomic indicators. β represents the estimated SREI coefficient.

Finally, to evaluate whether the relationship between historical Silk Road exposure and BD prevalence persisted without assuming linearity, we performed Spearman rank correlation analysis across the whole cohort (Table 6). The result showed that SREI showed a significant positive monotonic correlation with BD prevalence ($\rho = 0.568$, $P = 0.009$). This association provided additional evidence for a geographic exposure gradient independent of assumptions of linear regression structure.

Table 6. Spearman rank correlation analysis between Silk Road Exposure Index and Behçet's disease prevalence

Analysis	Variables	Sample size (n)	Spearman rho	P-value
Spearman rank correlation	SREI vs BD prevalence	20	0.568	0.009

Notes: Spearman rank correlation was performed to assess monotonic relationships between SREI and BD prevalence without assuming linearity or normality. The analysis included all 20 countries. ρ represents Spearman's rank correlation coefficient.

4. Discussion

4.1. Historical connectivity as a geographic framework for Behçet's disease distribution

This study identified a positive geographic gradient between Silk Road exposure and BD prevalence across 20 countries by integrating spatial indicators of historical connectivity, population-level HLA-B51 frequency, socioeconomic factors, and infectious disease burden.

According to the results, the spatial distribution pattern observed was consistent with previous epidemiological observations. BD distributed disproportionately along the Silk Road, particularly across the Mediterranean basin, the Middle East, and East Asia [42]. Unlike previous qualitative descriptions, we quantified Silk Road exposure using distance decay and territorial overlap. This

suggested that historically connected regions may share population-level genetic, demographic, or environmental characteristics associated with BD distribution.

Notably, after adding covariates like HLA-B51, latitude, HDI, and infectious burden, the effect of SREI was weakened. This claimed that Silk Road exposure itself was unlikely to be an independent causal factor. Instead, historical connectivity may represent a composite indicator reflecting long-term interactions among population movement, genetic variation, and environmental exposure. This interpretation was consistent with the evolutionary and epidemiological perspective: instead of being driven by a single determinant, BD distribution reflected interactions between genetic susceptibility and environmental exposures [43].

4.2. Interaction between genetic susceptibility and historical population movement

HLA-B51 is still the strongest recognized genetic susceptibility factor of BD [4]. In this study, HLA-B51 frequency showed positive association with BD prevalence. However, when other covariates were controlled, the association was no longer statistically significant. This might reflect the limited statistical power caused by limited number of countries and the ecological nature of the dataset.

The relationship between HLA-B51 distribution and Silk Road exposure provided us with an interesting evolutionary perspective. The historical trade corridors created conditions for repeated population interactions across Eurasia [37]. They therefore might influence the distribution of immune-related genetic variants through gene admixture and selection pressures [38]. Previous evolutionary hypotheses have proposed that immune-related alleles may persist because of balancing effects between pathogen defense and autoimmune susceptibility [45].

Also, current study cannot determine whether the distribution of HLA-B51 was derived directly from population movement along the Silk Road. Maybe both patterns reflect broader demographic processes. Future studies should incorporate population genetic approaches, such as ancestry-aware analyses and allele frequency modeling, to further examine this hypothesis.

4.3. Robustness of the geographic association and implications for disease ecology

To evaluate whether the observed association was driven by specific countries or statistical assumptions, we performed several sensitivity analyses. The positive direction of the SREI association persisted after sequential country removal. This finding suggested that the geographic pattern was not depending on a single observation. Apart from that, Spearman correlation analysis also supported that there was a monotonic relationship between Silk Road exposure ranking and BD prevalence ranking.

However, Negative Binomial models showed that after incorporating genetic and socioeconomic covariates, the association became weaker. This may reflect the complexity of interpreting historical connectivity within cross-national ecological analyses. Particularly, current healthcare access, disease recognition, population structure, and differences in epidemiological surveillance may influence reported prevalence estimates across countries [10, 46].

Therefore, Silk Road exposure should not be interpreted as a direct causal determinant of BD distribution. Instead, a more reasonable interpretation is that historical connectivity may be an integrative geographic indicator. It reflects long-term interactions among migration, microbial exposure, genetic variation, and environmental conditions [44].

4.4. Limitations and future directions

This study has several limitations. First, the ecological design determines that the results cannot be inferred to individual level [47]. Country-level associations cannot be directly extrapolated to individual risk.

Second, BD prevalence estimates were obtained from published epidemiological studies, so difference in diagnostic criteria, healthcare systems, and case ascertainment may lead to heterogeneity between countries [10].

Third, population-level HLA-B51 estimates may not fully capture national genetic diversity, and proxy estimates were used when direct measurements were unavailable. Finally, the small sample size limited statistical power and multivariable adjustment.

Future studies should integrate individual-level genetic data, ancestry-aware analyses, and spatial epidemiological models to clarify interactions between historical connectivity, genetic susceptibility, and environmental exposure [37]. Spatial Bayesian models and ancestry-adjusted analyses may help understand how genetic and environmental factors interact across geographic gradients [48].

5. Summary

In conclusion, this study quantitatively evaluated the association between historical Silk Road connectivity and the global distribution of BD. By developing SREI, we found that countries with high historical connectivity generally had a higher BD prevalence. Although the association became weaker after incorporating genetic, geographic, socioeconomic, and infectious disease-related factors, the geographic pattern remained consistent in multiple sensitivity analyses. This observation indicated a broader historical-environmental gradient behind BD distribution.

These findings suggest that the modern epidemiology of Behçet's disease may reflect the combined influence of ancient population connectivity, immune-related genetic variation, and changing environmental exposures [4, 5]. This study does not propose the Silk Road as a direct cause of disease. Instead, we proposed it as a useful framework for understanding how evolutionary processes and human movement may shape the geographic landscape of complex immune-mediated disorders.

References

- [1] Sakane, T., Takeno, M., Suzuki, N., & Inaba, G. (1999). Behçet's disease. *New England Journal of Medicine*, 341(17), 1284–1291. <https://doi.org/10.1056/NEJM199910213411707>
- [2] Davatchi, F., Chams-Davatchi, C., Shams, H., Shahram, F., Nadji, A., Akhlaghi, M., ... Masoumi, M. (2017). Behçet's disease: epidemiology, clinical manifestations, and diagnosis. *Expert Review of Clinical Immunology*, 13(1), 57–65. <https://doi.org/10.1080/1744666X.2016.1205486>
- [3] Mahr, A., Belarbi, L., Wechsler, B., Jeanneret, D., Dhote, R., Fain, O., Lhote, F., Ramanoelina, J., Coste, J. and Guillevin, L. (2008), Population-based prevalence study of Behçet's disease: Differences by ethnic origin and low variation by age at immigration. *Arthritis & Rheumatism*, 58: 3951–3959. <https://doi.org/10.1002/art.24149>
- [4] de Menthon, M., Lavalley, M. P., Maldini, C., Guillevin, L., & Mahr, A. (2009). HLA-B51/B5 and the risk of Behçet's disease: a systematic review and meta-analysis of case-control genetic association studies. *Arthritis and rheumatism*, 61(10), 1287–1296. <https://doi.org/10.1002/art.24642>
- [5] Gül, A. Pathogenesis of Behçet's disease: autoinflammatory features and beyond. *Semin Immunopathol* 37, 413–418 (2015). <https://doi.org/10.1007/s00281-015-0502-8>
- [6] Carla Maldini, Katie Druce, Neil Basu, Michael P LaValley, Alfred Mahr, Exploring the variability in Behçet's disease prevalence: a meta-analytical approach, *Rheumatology*, Volume 57, Issue 1, January 2018, Pages 185–195, <https://doi.org/10.1093/rheumatology/kew486>
- [7] Faviel F Gonzalez-Galarza, Antony McCabe, Eduardo J Melo dos Santos, James Jones, Louise Takeshita, Nestor D Ortega-Rivera, Glenda M Del Cid-Pavon, Kerry Ramsbottom, Gurpreet Ghattaoraya, Ana Alfirevic, Derek Middleton, et.al, Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools, *Nucleic Acids Research*, Volume 48, Issue D1, 08 January 2020, Pages D783–D788, <https://doi.org/10.1093/nar/gkz1029>
- [8] ElNahass, Y., Mekky, N., Abdelfattah, N.M. et al. HLA alleles, haplotypes frequencies, and their association with hematological disorders: a report from 1550 families whose patients underwent allogeneic bone marrow transplantation in Egypt. *Immunogenetics* 76, 243–260 (2024). <https://doi.org/10.1007/s00251-024-01343-x>
- [9] UNDP (United Nations Development Programme). 2020. Human Development Report 2020: The Next Frontier: Human Development and the Anthropocene. New York.

- [10] Vos T, Lim S, Abbafati C et al., Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019, *The Lancet*, 396, 1204-1222
- [11] Global Burden of Disease Collaborative Network. (2020). Global Burden of Disease Study 2019 (GBD 2019) results. Institute for Health Metrics and Evaluation (IHME). <https://ghdx.healthdata.org/gbd-2019>
- [12] Ru Li, Jian Sun, Li-Min Ren, Hong-Yu Wang, Wen-Hong Liu, Xue-Wu Zhang, Shi Chen, Rong Mu, Jing He, Yi Zhao, et al., Epidemiology of eight common rheumatic diseases in China: a large-scale cross-sectional survey in Beijing, *Rheumatology*, Volume 51, Issue 4, April 2012, Pages 721–729, <https://doi.org/10.1093/rheumatology/ker370>
- [13] Krause, I., Yankevich, A., Fraser, A. et al. Prevalence and clinical aspects of Behçet's disease in the north of Israel. *Clin Rheumatol* 26, 555–560 (2007). <https://doi.org/10.1007/s10067-006-0349-4>
- [14] Lee, Y.B., Lee, S.Y., Choi, J.Y., Lee, J.H., Chae, H.S., Kim, J.W., Han, K.D., Park, Y.G. and Yu, D.S. (2018), Incidence, prevalence, and mortality of Adamantiades-Behçet's disease in Korea: a nationwide, population-based study (2006–2015). *J Eur Acad Dermatol Venereol*, 32: 999-1003. <https://doi.org/10.1111/jdv.14601>
- [15] Zouboulis CC, Borchert J, Diesing J, Heinrich R, Galetzka W, Medelnik J-P, et al. Epidemiology and treatment of Adamantiades-Behçet's disease in Germany: A healthcare claims database study. *J Eur Acad Dermatol Venereol*. 2025; 39: 1017–1027. <https://doi.org/10.1111/jdv.20489>
- [16] Aladdin Mohammad, Thomas Mandl, Gunnar Sturfelt, Mårten Segelmark, Incidence, prevalence and clinical characteristics of Behçet's disease in southern Sweden, *Rheumatology*, Volume 52, Issue 2, February 2013, Pages 304–310, <https://doi.org/10.1093/rheumatology/kes249>
- [17] González-Gay MA, García-Porrúa C, Brañas F, López-Lázaro L, Olivieri I. Epidemiologic and clinical aspects of Behçet's disease in a defined area of Northwestern Spain, 1988-1997. *The Journal of Rheumatology*. 2000 Mar;27(3):703-707. PMID: 10743812.
- [18] Chandratte P, Chandan JS, Hammond B, Thayakaran R, Cusworth S, Adderley N, et al. Behçet's syndrome epidemiology in England: a real-world retrospective cohort analysis with a nested case-control study using linked UK primary and secondary care data. *RMD Open*. 2026; 12: e006594. <https://doi.org/10.1136/rmdopen-2025-006594>
- [19] Al-Rawi, Z., Neda, A. (2004). Prevalence of Behçet's Disease among Iraqis. In: Zouboulis, C.C. (eds) Adamantiades-Behçet's Disease. *Advances in Experimental Medicine and Biology*, vol 528. Springer, Boston, MA. https://doi.org/10.1007/0-306-48382-3_6
- [20] Olivieri, I., Leccese, P., Padula, A., Nigro, A., Palazzi, C., Gilio, M., & D'Angelo, S. (2013). High prevalence of Behçet's disease in southern Italy. *Clin Exp Rheumatol*, 31(3 Suppl 77), 28-31.
- [21] Villiger, R. A., Stefanski, A. L., Grobety, V., Adler, S., & Villiger, P. M. (2019). Behçet's syndrome: clinical presentation and prevalence in Switzerland. *Swiss medical weekly*, 149(2930), w20072-w20072.
- [22] Davatchi, F., Shahram, F., Chams-Davatchi, C., Shams, H., Abdolahi, B. S., Nadji, A., ... & Mousavi, M. (2019). Behçet's disease in Iran: analysis of 7641 cases. *Modern rheumatology*, 29(6), 1023-1030.
- [23] Kappen, J. H., van Dijk, E. H., Baak-Dijkstra, M., van Daele, P. L., Lam-Tse, W. K., van Hagen, P. M., & van Laar, J. A. (2015). Behçet's disease, hospital-based prevalence and manifestations in the Rotterdam area. *Neth J Med*, 73(10), 471-7.
- [24] al-Dalaan, A. N., Al Balaa, S. R., el Ramahi, K. A. M. A. L., al-Kawi, Z. U. H. E. I. R., Bohlega, S. A. E. E. D., Bahabri, S. U. L. T. A. N., & Al Janadi, M. A. (1994). Behçet's disease in Saudi Arabia. *The journal of rheumatology*, 21(4), 658-661.
- [25] Crespo, J., & Grupo Nacional para o Estudo da Doença de Behçet. (1997). Behçet's disease-a national casuistic. *Grupo Nacional Para o Estudo Da Doença de Behçet. Med Int*, 4, 225-232.
- [26] Madanat, W. Y., Alawneh, K. M., Smadi, M. M., Saadeh, S. S., Omari, M. M., & Bani Hani, A. B. (2017). The prevalence of Behçet's disease in the north of Jordan: a hospital-based epidemiological survey. *Clin Exp Rheumatol*, 35(Suppl 108), 51-54.
- [27] Calamia, K. T., Wilson, F. C., Icen, M., Crowson, C. S., Gabriel, S. E., & Kremers, H. M. (2009). Epidemiology and clinical characteristics of Behçet's disease in the US: a population-based study. *Arthritis Care & Research: Official Journal of the American College of Rheumatology*, 61(5), 600-604.
- [28] Gheita, T. A., El-Latif, E. A., El-Gazzar, I. I., Samy, N., Hammam, N., Abdel Noor, R. A., ... & Egyptian College of Rheumatology-Behçet's Disease Study Group (ECR-BDSG). (2019). Behçet's disease in Egypt: a multicenter nationwide study on 1526 adult patients and review of the literature. *Clinical rheumatology*, 38(9), 2565-2575.
- [29] Nakae, K. (1993). Recent epidemiological features of Behçet's disease in Japan. *Behçet's disease*.
- [30] QGIS.org, 2026. QGIS Geographic Information System. QGIS Association. <http://www.qgis.org>
- [31] Natural Earth. (n.d.). Admin 0 – Countries (1:10m scale), `ne_10m_admin_0_countries.shp`. Natural Earth. <https://www.naturalearthdata.com/downloads/10m-cultural-vectors/10m-admin-0-countries/>

- [32] John R. Nuckols, Mary H. Ward, Lars Jarup; Using Geographic Information Systems for Exposure Assessment in Environmental Epidemiology Studies. *Environ. Health Perspect.* 1 June 2004; 112 (9): 1007–1015. <https://doi.org/10.1289/ehp.6738>
- [33] Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46, 234–240. <https://doi.org/10.2307/143141>
- [34] OECD/European Union/EC-JRC (2008), Handbook on Constructing Composite Indicators: Methodology and User Guide, OECD Publishing, Paris, <https://doi.org/10.1787/9789264043466-en>.
- [35] Feng, C., Wang, H., Lu, N. and Tu, X.M. (2013), Log transformation: application and interpretation in biomedical research. *Statist. Med.*, 32: 230-239. <https://doi.org/10.1002/sim.5486>
- [36] Cook, R. D. (1977). Detection of Influential Observation in Linear Regression. *Technometrics*, 19(1), 15–18. <https://doi.org/10.1080/00401706.1977.10489493>
- [37] Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). Regression diagnostics: Identifying influential data and sources of collinearity. John Wiley & Sons. <https://doi.org/10.1002/0471725153>
- [38] Hilbe JM. Negative Binomial Regression. 2nd ed. Cambridge University Press; 2011.
- [39] Spearman, C. (1904). The Proof and Measurement of Association between Two Things. *The American Journal of Psychology*, 15(1), 72–101. <https://doi.org/10.2307/1412159>
- [40] Seabold, S., & Perktold, J. (2010). Statsmodels: econometric and statistical modeling with python. *scipy*, 7(1), 92–96.
- [41] Virtanen, P., Gommers, R., Oliphant, T.E. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 17, 261–272 (2020). <https://doi.org/10.1038/s41592-019-0686-2>
- [42] Verity, D. H., Marr, J. E., Ohno, S., Wallace, G. R., & Stanford, M. R. (1999). Behçet's disease, the Silk Road and HLA-B51: historical and geographical perspectives. *Tissue antigens*, 54(3), 213-220.
- [43] Sezgin E and Kaplan E (2022) Diverse selection pressures shaping the genetic architecture of behçet disease susceptibility. *Front. Genet.* 13:983646. doi: 10.3389/fgene.2022.983646
- [44] Takeuchi, M., Kastner, D. L., & Remmers, E. F. (2015). The immunogenetics of Behçet's disease: A comprehensive review. *Journal of autoimmunity*, 64, 137–148. <https://doi.org/10.1016/j.jaut.2015.08.013>
- [45] Krummel M. Matthew Krummel: visions enumerated. Interview by Amy Maxmen. *J Exp Med.* 2009; 206(6): 1212-1213. doi:10.1084/jem.2066pi
- [46] Murray C, Aravkin A, Zheng P et al., Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019, *The Lancet*, 396, 1223-1249
- [47] Robinson, W. S. (1950). Ecological Correlations and the Behavior of Individuals. *American Sociological Review*, 15(3), 351–357. <https://doi.org/10.2307/2087176>
- [48] Besag, J., York, J. & Mollié, A. Bayesian image restoration, with two applications in spatial statistics. *Ann Inst Stat Math* 43, 1–20 (1991). <https://doi.org/10.1007/BF00116466>