Calvo Pascual Luis Ángel (Orcid ID: 0000-0002-8818-5542)
Castro-Corredor David (Orcid ID: 0000-0001-7315-6274)
López-Medina Clementina (Orcid ID: 0000-0002-2309-5837)

Title: The impact of social factors on VAS pain in spondyloarthritis patients: a Propensity Score Matching analysis

Running head: Social factors and VAS pain in SpA

Authors:
Luis Ángel Calvo Pascual, B.Sc., M. Sc., Ph.D. (0000-0002-8818-5542) Universidad Pontificia de Comillas Madrid. Spain. lacalvo@comillas.edu

David Castro Corredor, M.D., Ph.D. (0000-0001-7315-6274) Hospital General Universitario de Ciudad Real. Spain. d.castrocorredor@gmail.com

Eduardo Collantes-Estévez, M.D., Ph.D. (0000-0002-7647-6289) Maimonides Institute for Research in Biomedicine of Cordoba (IMIBIC), Reina Sofia University Hospital, University of Cordoba, Córdoba, Spain. educollantes@yahoo.es

Clementina López-Medina, M.D., Ph.D. (0000-0002-2309-5837) Maimonides Institute for Research in Biomedicine of Cordoba (IMIBIC), Reina Sofia University Hospital, University of Cordoba, Córdoba, Spain. clementinalopezmedina@gmail.com

Correspondence to:
Luis Ángel Calvo Pascual, ICADE. Universidad Pontificia de Comillas, C/Alberto Aguilera, 23, 28015, Madrid. Spain. lacalvo@comillas.edu

Disclosure statements:
a) Funding statement: The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The REGISPONSER registry was funded by the Spanish Society of Rheumatology. There was no funding for this subanalysis.
b) Informed consent/ Patient consent: All patients gave their written consent to participate in the REGISPONSER and RESPONDIA registries.
c) Ethical Statement: The study was conducted in compliance with the recommendations of the Declaration of Helsinki and was centrally approved by the Ethics Committee of the Reina Sofia University Hospital from Córdoba (Spain) on 21 April 2006 ("Comisión de Ética e Investigación Sanitaria, REGISPON-2004 y RESPONDIA-2006").

Word count: 3638

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/acr.25675

Abstract: [word count: 248/250]

Background: Pain in spondyloarthritis reflects not only inflammatory activity but also social context. We quantified the association between a composite Social Factors Index (SFI) and patient-reported pain severity on the Visual Analogue Scale (VAS).
Methods: We analyzed 2,042 patients from REGISPONSER (n=1,514) and RESPONDIA (n=528). Missing data were handled with multiple imputation (m=20). The SFI combined education, employment, housing, income source, and Graffar class using penalized regression (sparse-group lasso). Mutual information (MI) quantified the SFI's relative importance. Propensity scores (logistic regression) supported 1:1 caliper matching (0.2 SD on the logit), comparing the highest (Q4) vs lowest (Q1) SFI quartiles.
Results: Before matching, Q1 and Q4 included on average 918 and 829 patients per imputation, differing mainly in race (White 80.8% vs 86.8%; SMD=−0.16) and country (Spain 66.2% vs 78.2%; SMD=−0.27). After matching, a mean of 745 pairs (range 731–761) were retained with excellent covariate balance (all |SMD|<0.10). In the matched sample, patients in Q4 reported higher pain (Average Treatment effect on the Treated (ATT) +0.38 VAS points; 95% CI: 0.05–0.71; p=0.023). Post-matching regression yielded an identical estimate (β=+0.37; 95% CI: 0.04–0.70; p=0.030). MI indicated the SFI carried more information about VAS than any single social variable. Stratified analyses suggested heterogeneity: Spain +0.31 (95% CI: 0.08–0.71) and Latin America +0.62 (95% CI: −0.22 to 1.45).
Conclusions: Social disadvantage, summarized by the SFI, is independently associated with greater pain severity in spondyloarthritis after adjustment. These findings support incorporating social determinants into clinical assessment and pain management.

Keywords: Mutual Information, Propensity Score Matching, VAS, Spondyloarthritis.

Significance and Innovations

- We developed a Social Factors Index (SFI) using sparse-group lasso with bootstrap across multiply imputed datasets; mutual information quantified the SFI's informativeness for patient-reported VAS pain.

- The SFI captured more mutual information with VAS than any single social variable, including Graffar Scale class, employment status, housing conditions, or income.

- Using propensity-score matching (>700 matched pairs per imputation), patients in the highest SFI quartile reported +0.38 VAS points vs the lowest (95% CI: 0.05–0.71; p=0.023), supporting an association independent of measured disease activity

- Findings underscore the clinical relevance of social vulnerability in SpA and support integrating social determinants into comprehensive care and management.

INTRODUCTION

Spondyloarthritis (SpA) encompasses a group of chronic inflammatory rheumatic diseases that primarily affect the axial skeleton, leading to significant physical limitations and diminished quality of life. It often strikes individuals in early adulthood, causing chronic back pain, stiffness, and progressive loss of mobility. Consequently, SpA is associated with severe physical limitations, functional impairment, and decreased quality of life[1].

Pain is a central and enduring symptom of SpA, traditionally attributed to inflammatory processes. However, a substantial subset of patients continues to experience significant pain despite effective control of inflammation through pharmacological interventions. This persistent pain suggests the involvement of non-inflammatory mechanisms, notably central sensitization, an increased responsiveness of nociceptive neurons in the central nervous system, leading to heightened pain sensitivity and perception[2].

Beyond biological mechanisms, social determinants have emerged as important contributors to pain severity in SpA[3,4]. Lower educational attainment and socioeconomic disadvantage are independently associated with higher patient-reported pain, even after adjustment for disease activity and treatment access[5-7]. In a cluster analysis, patients in the low-socioeconomic group experienced longer diagnostic delays, higher body mass index, and more severe structural damage despite adjustment for disease duration; although access to biologic disease-modifying anti-rheumatic drugs (DMARDs) was comparable across groups, disadvantaged patients had a substantially higher prevalence of permanent work disability, defined as official certification of long-term incapacity for work, independent of disease activity and duration[8]. Moreover, patients with chronic low back pain reported significantly worse pain severity, disability, and health-related quality of life than those with axial spondyloarthritis, despite similar pain intensity scores, underscoring that pain experience extends beyond inflammatory burden[9].

The present study investigates whether social determinants contribute to patient-reported pain in spondyloarthritis (SpA) using data from two registries (REGISPONSER and RESPONDIA). We examined education, employment status, housing conditions, and Graffar social class, and aggregated them into a Social Factors Index (SFI) whose component weights were learned via lasso regularization (sparse-group lasso) with bootstrap resampling across multiply imputed datasets. We then quantified the importance of the SFI for Visual Analogue Scale (VAS) pain using mutual information, and applied propensity-score matching to balance key demographic and clinical covariates, contrasting patients in the highest versus lowest SFI quartiles.

PATIENTS AND METHODS

Study design:

This study was based on data obtained from two multicenter observational registries: REGISPONSER and RESPONDIA, both specifically designed to collect standardized clinical information on patients with spondyloarthritis (SpA).

REGISPONSER is a national, multicenter registry from Spain that prospectively enrolled consecutive patients diagnosed with SpA according to the European Spondyloarthropathy Study Group (ESSG) criteria. Recruitment was conducted between March 2004 and March 2007 across 31 rheumatology departments. Detailed information regarding the design, sampling methodology, and recruitment process can be found in Collantes et al.[10].

RESPONDIA, a parallel registry with an identical study design, was conducted between 2006 and 2007 in Latin America. It included 33 centers across eight countries: Argentina, Chile, Colombia, Costa Rica, Mexico, Uruguay, and Venezuela. Like REGISPONSER, RESPONDIA included consecutive patients who met the ESSG classification criteria for SpA and used the same case report forms, clinical definitions, and variable collection procedures[11].

For this study, data from both sources were merged to generate a large, multinational cohort of SpA patients.

Inclusion and exclusion criteria:

All patients aged 18 years or older with a diagnosis of spondyloarthritis (SpA) according to the European Spondyloarthropathy Study Group (ESSG) criteria were eligible. Participants were drawn from the REGISPONSER registry (Spain, 2004–2007) and the RESPONDIA registry (Latin America, 2006–2007). Patients with incomplete clinical information in key exposure or outcome variables, or those with duplicate records, were excluded. No additional exclusion criteria were applied.

Outcomes:

The primary outcome of interest was pain intensity, measured using the Visual Analogue Scale (VAS), which ranged from 0 (no pain) to 10 (worst imaginable pain), referring to overall pain experienced during the previous week. No formal secondary outcomes were prespecified.

Variables:

This study analyzed data from 2042 patients diagnosed with spondyloarthritis. A total of 66 variables were collected and organized into four conceptual domains: personal variables, social variables, treatment-related variables, and disease activity. All variables were assessed at the baseline visit of each registry, with laboratory tests performed within 15 days before inclusion. Detailed information is provided in the supplementary material (*Descriptive.docx*), which includes six tables: three with categorical variables and three with numerical variables, each presented for the full sample, for REGISPONSER, and for RESPONDIA.

The personal variables include age, weight, height, physical activity habits, smoking status, race/ethnicity and country of residence across Ibero-American regions such as Spain, Argentina, Mexico, and others. Genetic and familial factors were captured through HLA-B27 status (positive or negative) and the presence or absence of a family history of spondyloarthritis.

The social variables domain was composed of educational attainment, ranging from illiterate to university level. Employment status included categories such as employed, unemployed, student, retired, or homemaker. Socioeconomic stratification was measured using the Graffar scale[12], while housing conditions ranged from inadequate to excellent, with luxury. Marital status, sources of income (salary, profits, donations, rents), and professional background (e.g., skilled or unskilled labour, technician, employee) were also recorded. All social variables were systematically collected for all participants at the baseline visit.

Therapeutic exposure variables consisted of 15 measures of pharmacological treatment and management, including corticosteroids, NSAIDs (regular, on-demand, or none), biologic therapies (infliximab, etanercept, adalimumab), and conventional synthetic DMARDs (methotrexate, sulfasalazine, leflunomide).

The disease activity variables included measures of clinical status, functional impairment, and patient-reported outcomes. Disease activity was assessed with the Ankylosing Spondylitis Disease Activity Score (ASDAS). Functional impact and quality of life were measured using the Ankylosing Spondylitis Quality of Life (ASQoL) questionnaire and the physical and mental component scores of the SF-12 survey. Axial and peripheral involvement was evaluated through the number of swollen joints, chest expansion, and Schober's test. The dataset also recorded the clinical form of the disease (axial, peripheral, enthesitic, or mixed), extra-articular manifestations (iritis, uveitis, enthesitis, dactylitis, pustulosis, acne conglobata, balanitis, prostatitis, and nail involvement), comorbidities (cardiac, renal, pulmonary, and neurological) and disease duration (years from symptom onset to the baseline visit). Pain was assessed with the Visual Analogue Scale (VAS), where patients rated their overall pain (not restricted to back pain) from 0 (no pain) to 10 (worst imaginable pain) based on the previous week.

Data analysis:

We merged REGISPONSER and RESPONDIA, removed outliers and data errors. We summarized numerical and categorical variables to characterize the sample; baseline demographic, clinical, and social characteristics are shown in Table 1. More detailed distributions and missing-data summaries by cohort, together with boxplots of VAS across categorical variables and kernel density plots for numerical variables, are provided in the Supplementary Material (Descriptive.docx; Boxplot.docx).

Given non-trivial missingness, we used multiple imputation to create m = 20 completed datasets (13). We rejected MCAR via Little's test and proceeded under a Missing-At-Random assumption (14). To study social determinants of pain, we built a Social Factors Index (SFI) from education, employment/occupation, income source, housing, and Graffar. In each imputation, models used penalization with out-of-fold predictions to avoid optimism; when group-lasso was unavailable, we used elastic net with internal cross-validation (15). Individual SFI scores were averaged across imputations. We evaluated SFI performance via correlation and discrimination for high pain (AUC for VAS ≥ 7), reporting per-imputation and pooled metrics (Fig. 1).

To quantify associations with VAS beyond linear effects, we computed mutual information (MI) between VAS and candidate predictors—including SFI—within each imputed dataset, following recent applications in rheumatology (16–19). We used the standard definition

$$\mathrm{MI}(X, VAS) = H(VAS) - H(VAS \mid X),$$

where $H(\cdot)$ is the Shannon entropy function. Practically, we discretized VAS into five quantile-based bins, encoded categorical variables, scaled continuous predictors, and estimated MI with bootstrap uncertainty (100 resamples per imputation), then pooled ranks across imputations (Fig. 2). MI guided the choice of pre-exposure covariates for causal analyses.

We estimated the average treatment effect on the treated (ATT) by contrasting high (Q4) vs. low (Q1) SFI using propensity score matching with logistic-model scores, overlap trimming (20), 1:1 nearest-neighbor matching without replacement on the logit scale, and a 0.2 SD caliper (with 0.3 SD as a sensitivity check) (21). Balance was assessed using standardized mean differences, and ATT estimates with standard errors were pooled across imputations via Rubin's rules. We additionally fit doubly robust outcome regressions for covariates with residual imbalance. The pooled ATT (95% CIs) is shown in Figure 3 (forest plot), covariate balance before and after matching in Figure 4 (love plot), and paired post-matching VAS differences in Figure 5 (distribution of matched differences). Full specifications, diagnostics, and sensitivity analyses are provided in the Supplementary Methods (dataanalysis.docx); Supplementary Material also includes: SFI validation outputs and metrics (social_index_eval.xlsx), full PSM results with per-imputation and pooled ATT (psm_results.xlsx), detailed balance tables (balance_tables.xlsx), doubly robust regression outputs (regression_psm_dr.xlsx), caliper sensitivity (0.2 vs 0.3 SD; caliper_sensitivity.xlsx), and stratified Spain vs non-Spain analyses (country_meta.xlsx).

RESULTS

In total, 2,042 patients were included (REGISPONSER n = 1,514; RESPONDIA n = 528). A concise summary of baseline demographic, clinical, and social characteristics is provided in Table 1. Patients from REGISPONSER were on average older (48.1 years, SD 12.9) than those from RESPONDIA (44.7 years, SD 14.7), and the proportion of males was higher (74.7% vs. 67.0%). Regarding disease activity and pain, the mean VAS score was 3.20 (SD 2.13) in REGISPONSER and 3.72 (SD 2.58) in RESPONDIA, while ASDAS was 2.58 (SD 1.12) and 2.03 (SD 1.01), respectively. Socioeconomic differences were also observed: for example, university education and employment were more common in RESPONDIA. Comparative boxplots of all variables can be consulted in the supplementary file boxplots.xlsx, where it can also be observed that patients in the Graffar upper class reported lower VAS levels. Detailed descriptive statistics (means, SD, quartiles, n and %, missing data) by registry are provided in descriptive.docx.

The Social Factor Index (SFI) was defined as a weighted sum of nine social factors each compared to the middle baseline of its domain:

$$
\begin{aligned}
SFI = & +0.0467\,[Employment\ Status\_student] \\
& + 0.0177\,[Employment\ Status\_Unemployed] \\
& - 0.0175\,[Employment\ Status\_Employed] \\
& + 0.0179\,[Education\_Illiterate] - 0.0236\,[Education\_University] \\
& + 0.0414\,[Housing\ Conditions\_Inadequate] \\
& - 0.0224\,[Housing\ Conditions\_Deficients] \\
& + 0.0246\,[Graffar\ Scale\_Lower\ class] \\
& - 0.0211\,[Graffar\ Scale\_Upper\ class].
\end{aligned}
$$

Positive coefficients indicate higher expected pain (VAS) relative to baseline. The index showed a modest but consistent association with VAS (Pearson r≈0.17) and modest discrimination for high pain (VAS ≥ 7, AUC ≈ 0.56). However, the boxplot of VAS across SFI quartiles (Fig. 1) showed an upward shift in both medians and interquartile ranges, suggesting higher pain levels with increasing SFI.

Mutual-information ranking (Fig. 2) showed that the Social Factors Index captured more information about VAS than any single social variable. The pooled MI for the SFI index was 0.0239, placing it 20th overall and ahead of other social indicators such as the Graffar scale (0.0124), Employment status (0.0134), Housing conditions (0.0144), Source of income (0.0102), Race (0.0151), and Country (0.0257). As expected, clinical/activity measures dominated the top ranks (e.g., ASDAS 0.579, ASQoL 0.237, Weight 0.104, SF-12 Physical 0.156). As we described in the previous section, we also used these MI results to select those variables high ranked and plausibly relate to both social risk and pain. This led to the final set of baseline covariates: Age (0.074), Diagnosis Age (0.066), Disease Duration (0.059), Diagnostic Delay (0.044), Weight (0.104), Height (0.051), Race (0.015), and Country (0.026).

Regarding the results of the propensity score matching analysis, we compared pain outcomes between patients with high social factor scores (Q4 of the SFI) and those with low scores (Q1). Before matching, Q1 and Q4 included, on average, 918 and 829 patients per imputation, respectively, reflecting roughly one quarter of the cohort. The pooled average treatment effect on the treated (ATT) was +0.38 VAS points (SE = 0.17), with a 95% confidence interval of 0.05 to 0.71 and a p value of 0.023 across the 20 imputations (Fig. 3). After overlap trimming and 1:1 nearest-neighbour caliper matching, an average of 745 matched pairs per imputation was retained (range 731–761), corresponding to about 90% of eligible patients in the extreme quartiles. Matching quality was excellent, with a mean absolute standardized difference of 0.041 across covariates. Figure 4 displays the pooled standardized mean differences (SMDs) across imputations for all variables in the propensity model, weighted by matched pairs. Before matching (grey dots), some covariates showed imbalance, particularly Race (SMD = –0.16) and Country (SMD = –0.27). After matching, these imbalances were reduced to –0.09 and –0.04, respectively, while continuous covariates such as age, diagnosis age, disease duration, delay, weight, and height all achieved post-match SMDs between –0.07 and 0.04. Thus, every covariate met the conventional threshold of |SMD| < 0.10. In the matched sample, the distribution of pairwise differences in VAS remained shifted to the right, with a pooled mean difference of +0.38 points (95% CI: 0.05–0.71), confirming that pain severity was consistently higher in the high-SFI group.

The results of the caliper sensitivity analysis (see caliper_sensitivity.xlsx, Supplementary Material) were highly consistent across specifications. When widening the caliper to 0.3 SD, results were almost identical to those obtained with 0.2 SD (ATT = +0.41; SE = 0.16; 95% CI: 0.09–0.72; p = 0.011), with a similar number of matched pairs (802) and a mean absolute standardized mean difference after matching of 0.022.

Finally, in the stratified analysis to observe differences between Spain and Latin American countries, we obtained an average treatment effect on the treated of +0.31 (SE = 0.20; 95% CI: 0.08 to 0.71) across 20 imputations for Spain, and +0.62 (SE = 0.43; 95% CI: −0.22 to 1.45) across 20 imputations for Latin America.

DISCUSSION

Some previous studies have primarily attributed the impact of socio-economic factors not to pain, but to disease activity. For example, Capelusnik et al.[7] suggest that lower educational levels and single marital status were associated with higher ASDAS. Other studies[23, 24] found that lower levels of education and income were associated with a higher likelihood of receiving a chronic pain diagnosis.

Our Social Factors Index (SFI) shows a clear gradient: students and unemployed individuals, those living in inadequate housing, and patients from lower social classes have higher predicted pain, whereas those with university education and from upper social classes have lower predicted pain. These patterns align with evidence that social determinants are linked to worse outcomes in axial spondyloarthritis[7]. Mendelian-randomization work indicates that higher educational attainment causally reduces multisite chronic pain risk, plausibly via healthier behaviours and greater psychological resources[25]. Daily diary research also shows that financial hardship and unemployment amplify day-to-day pain reactivity, particularly among women with chronic musculoskeletal disease who worry about finances when unemployed[26]. Taken together, this supports aggregating correlated social exposures to reflect cumulative disadvantages.

Despite the modest performance of the SFI in predicting VAS (Pearson r ≈0.17; AUC ≈ 0.56), a low overall correlation does not rule out meaningful group differences. When we group SFI into quartiles, Figure 1 shows that median VAS (and its spread) increases stepwise from Q1 to Q4. In addition, the SFI ranked first among all social indicators in the mutual-information analysis, ahead of Education, Graffar scale, Employment status, Housing conditions, and Source of income (Figure 2). This means that, while social context explains less variance in pain than clinical activity, the combined index captures more information about VAS than any single social variable, supporting its use as a compact summary of cumulative social factors.

Using propensity-score matching to compare patients in the highest SFI quartile (Q4) with those in the lowest (Q1), the pooled average treatment effect on the treated (ATT) for VAS was +0.38 points (SE = 0.17; 95% CI: 0.05–0.71; p = 0.023) across 20 imputations (Figure 3). Covariate balance was excellent after matching (mean absolute SMD ≈ 0.041) with ~745 matched pairs per imputation (Figures 4, 5), and results were stable to caliper changes (0.2–0.3 SD), consistent with prior methodological work[27-29]. In plain terms, patients with high SFI reported ~0.4 points more VAS pain than comparable patients with low SFI. Although modest, this effect was consistent and clinically relevant. The gradient was more pronounced in Latin America (ATT +0.62; 95% CI: −0.22 to 1.45) than in Spain (ATT +0.31; 95% CI: −0.08 to 0.71), suggesting stronger social disparities in the former.

Our methodology has several strengths. We used multiple imputations (20 datasets) and derived a stable, interpretable SFI via sparse-group lasso with bootstrap resampling. Mutual information was applied only to assess the SFI's informativeness for VAS, not for confounder selection. Prespecified baseline covariates (age, age at diagnosis, disease duration, diagnostic delay, weight, height, race, country) achieved excellent post-match balance (mean absolute SMD $\approx$ 0.041). Propensity scores were estimated with logistic regression, prioritizing balance over prediction; effects were robust to caliper variation on the logit scale (0.2–0.3 SD) and were larger in Latin America than in Spain.

However, several limitations should be noted. First, the study is observational and cross-sectional rather than longitudinal, which limits the ability to establish causal relationships and raises the possibility of reverse causation. Second, although we adjusted for a wide range of demographic and clinical covariates, residual confounding cannot be excluded, particularly from unmeasured psychosocial factors such as depression, coping styles, health literacy, or differences in healthcare access. Third, the SFI was developed primarily in a Spanish cohort, and while the effect appeared stronger in Latin America, where social inequalities are more pronounced, external validation and larger datasets from these countries are needed to confirm and generalize these findings.

CONCLUSIONS

This study suggests that social factors could contribute independently to pain severity in patients with spondyloarthritis (SpA). We summarized education, employment status, Graffar scale and housing conditions into a single Social Factors Index (SFI), showing a clear gradient in pain: VAS increased stepwise from the lowest to the highest SFI quartile. With propensity-score matching, patients in the highest SFI quartile had modest yet consistent increases in VAS compared with otherwise comparable patients in the lowest quartile (ATT = +0.38; 95% CI: 0.05–0.71; $p = 0.023$), supporting an effect beyond measured disease activity. Although the individual-level effect size is small, such gradients can be meaningful at the population level and have policy relevance, helping identify groups who may benefit from targeted support. Clinically, patients with lower education, unstable employment, or inadequate housing may benefit from targeted assessment and support (e.g., pain self-management support, social work referral). Future work should externally validate the SFI and matched findings, use longitudinal designs to establish temporality, and test interventions that address the identified social pathways.

REFERENCES

1. Mease, P. J., Bhutani, M. K., Hass, S., Yi, E., Hur, P., Kim, N. (2022). Comparison of Clinical Manifestations in Rheumatoid Arthritis vs. Spondyloarthritis: A Systematic Literature Review. Rheumatology and Therapy, 9(2), 331–378. https://doi.org/10.1007/s40744-021-00407-8

2. Lee, Y. C., Malfait, A. M., Ogdie, A. R. (2024). Unmet needs in spondyloarthritis: Understanding and managing chronic pain. The Journal of Rheumatology, 51(12), 1235–1240. https://doi.org/10.3899/jrheum.2024-0936

3. Kiltz, U., Baraliakos, X., Regel, A., Bühring, B., & Braun, J. (2017). Causes of pain in patients with axial spondyloarthritis. Clin Exp Rheumatol, 35(Suppl 107), 102-107.

4. Pathan, E. M., Inman, R. D. (2017). Pain in spondyloarthritis: A neuro–immune interaction. *Best Practice & Research Clinical Rheumatology, 31*(6), 830–845. https://doi.org/10.1016/J.BERH.2018.07.003

5. Capelusnik, D., Boonen, A., Ramiro, S., Nikiphorou, E. (2025). The role of social determinants of health on disease outcomes in axial spondyloarthritis: A narrative review. *Autoimmunity Reviews, 24*(5), Article 103762. https://doi.org/10.1016/j.autrev.2025.103762

6. Santos, H., Henriques, A. R., Branco, J., Machado, P. M., Canhão, H., Pimentel-Santos, F. M., Rodrigues, A. M. (2023). Health-related quality of life among spondyloarthritis and chronic low back pain patients: Results from a nationwide population-based survey. Quality of Life Research, 32(2), 383–399. https://doi.org/10.1007/s11136-022-03274-0

7. Capelusnik, D., Zhao, S. S., Boonen, A., Ziade, N., Medina, C. L., Dougados, M., Nikiphorou, E., Ramiro, S. (2022). Individual-level and country-level socio-economic factors and health outcomes in spondyloarthritis: Analysis of the ASAS-perSpA study. Rheumatology (Oxford, England), 61(5), 2043–2053. https://doi.org/10.1093/rheumatology/keab638

8. Ruiz-Vilchez, D., Ladehesa-Pineda, L., Puche-Larrubia, M. Á., Ábalos-Aguilera, M. C., Font-Ugalde, P., Escudero-Contreras, A., López-Medina, C. (2024). The socioeconomic status of patients with ankylosing spondylitis and its association with the burden of the disease and permanent disability: A cross-sectional cluster analysis. Therapeutic Advances in Musculoskeletal Disease, 16, Article 1759720X241272947. https://doi.org/10.1177/1759720X241272947

9. Cheung, J. P. Y., Cheung, P. W. H., Wong, C. K. H., Chung, H. Y., Tsang, H. H. L. (2023). Propensity-matched comparison between chronic nonspecific low back pain and axial spondyloarthritis: impact on patient-perceived quality of life. Spine, 48(8), 577-589.

10. Collantes, E., Zarco, P., Muñoz, E., et al. (2007). Disease pattern of spondyloarthropathies in Spain: Description of the first national registry (REGISPONSER) – extended report. Rheumatology (Oxford), 46(8), 1309–1315. https://doi.org/10.1093/rheumatology/kem084

11. Vazquez-Mellado J, Font Ugalde P, Muñoz Gomariz E, et al. Ibero-American Spondylarthritis Registry (RESPONDIA): what is, how it came about, who we are and what we do. Reumatología Clínica 2008; 4:S17–22.

12. Graffar, M. (1956). *Une méthode de classification sociale d'échantillons de population*. *Courrier de l'Office International d'Hygiène Publique*.

13. Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: John Wiley & Sons.

14. Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. Journal of the American Statistical Association, 83(404), 1198–1202. https://doi.org/10.1080/01621459.1988.10478722

15. Friedman, J., Hastie, T., Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, 33*(1), 1–22. https://doi.org/10.18637/jss.v033.i01

16. Castro Corredor, D., Calvo Pascual, L. Á., Collantes-Estévez, E., López-Medina, C. (2025). *Inflammatory activity levels in patients with anti-TNF therapy: Most important factors and a decision tree model. Data from REGISPONSER and RESPONDIA registries*. Therapeutic Advances in Musculoskeletal Diseases, 17, https://doi.org/10.1177/1759720X251332224

17. Calvo Pascual L. A., Castro Corredor D., Garrido Merchán E. C., Machine learning classification of vitamin D levels in spondyloarthritis patients. Intelligence-Based Medicine. Volume 9. 2024. 100125. ISSN 2666-5212. https://doi.org/10.1016/j.ibmed.2023.100125.

18. Castro Corredor D, Calvo Pascual L. A. (2023) Imbalanced machine learning classification models for removal biosimilar drugs and increased activity in patients with rheumatic diseases. PLoS ONE 18(11): e0291891. https://doi.org/10.1371/journal.pone.0291891

19. Curtis, J. R., Su, Y., Black, S., Xu, S., Langholff, W., Bingham, C. O., Kafka, S., Xie, F. (2022). Machine learning applied to patient-reported outcomes to classify physician-derived measures of rheumatoid arthritis disease activity. ACR Open Rheumatology, 4(12), 995–1003. https://doi.org/10.1002/acr2.11499

20. Crump, R. K., Hotz, V. J., Imbens, G. W., Mitnik, O. A. (2009). *Dealing with limited overlap in estimation of average treatment effects*. **Biometrika**, 96(1), 187–199. https://doi.org/10.1093/biomet/asn055

21. Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Statistics in Medicine, 28(25), 3083–3107. https://doi.org/10.1002/sim.3697

22. DerSimonian, R., Laird, N. (1986). Meta-analysis in clinical trials. Controlled Clinical Trials, 7(3), 177–188. https://doi.org/10.1016/0197-2456(86)90046-2

23. Jöud, A., Petersson, I. F., Jordan, K. P., Löfvendahl, S., Grahn, B., & Englund, M. (2014). Socioeconomic status and the risk for being diagnosed with spondyloarthritis and chronic pain: A nested case–control study. Rheumatology International, 34, 1291–1298. https://doi.org/10.1007/s00296-014-3039-6

24. Prego-Domínguez, J., Khazaeipour, Z., Mallah, N., & Takkouche, B. (2021). Socioeconomic status and occurrence of chronic pain: A meta-analysis. Rheumatology, 60(3), 1091–1105. https://doi.org/10.1093/rheumatology/keaa758

25. Wang, H., Liu, M., Li, H., Xu, S. (2025). Association between educational attainment and chronic pain: A mediation Mendelian randomization study. *Journal of Pain Research, 18*, 1793–1804. https://doi.org/10.2147/JPR.S515921

26. Rios, R., Zautra, A. J. (2011). Socioeconomic disparities in pain: The role of economic hardship and daily financial worry. *Health Psychology, 30*(1), 58–66. https://doi.org/10.1037/a0022025

27. Lee, B. K., Lessler, J., Stuart, E. A. (2010). Improving propensity score weighting using machine learning. Statistics in Medicine, 29(3), 337–346. https://doi.org/10.1002/sim.3782

28. Pirracchio, R., Petersen, M. L., van der Laan, M. J. (2015). Improving propensity score estimators' robustness to model misspecification using super learner. American Journal of Epidemiology, 181(2), 108–119. https://doi.org/10.1093/aje/kwv290

29. Zhao, P., Su, X. G., Ge, T., & Fan, J. (2016). Propensity score and proximity matching using random forest. Contemporary Clinical Trials, 47, 85–92. https://doi.org/10.1016/j.cct.2015.12.012

FIGURE LEGENDS

Figure 1. Distribution of pain scores (VAS) across quartiles of the Social Factors Index (SFI)

Figure 2. Mutual information with VAS for the 30 clinical and social features, computed across 100 bootstrap iterations. We highlight the SFI index in red to show its relative position with respect to the other variables.

Figure 3. Forest plot of ATT estimates for VAS across imputations and pooled results.

Figure 4. Love plot depicting covariate balance before and after PSM

Figure 5. Distribution of paired differences in pain scores after matching

TABLE LEGENDS

Table 1. Baseline characteristics of patients with spondyloarthritis in the joint pooled dataset, REGISPONSER and RESPONDIA cohorts.

SUPPLEMENTARY FILES

- **social_index_eval.xlsx –** Evaluation metrics and outputs of the Social Factors Index (SFI), including per-imputation performance, pooled summaries, out-of-fold predictions, and calibrated VAS scores.
- **psm_results.xlsx –** Full outputs of the propensity score matching analysis (Q4 vs. Q1 of the Social Factors Index), including covariate balance and ATT estimates across imputations.
- **Balance_tables.xlsx–** Detailed covariate balance tables before and after propensity score matching, reported separately for each imputation and pooled across imputations
- **Boxplot.docx –** Visualizations of VAS pain by categorical and numerical predictors, with boxplots and kernel density estimates.
- **caliper_sensitivity.xlsx –** Sensitivity analysis of propensity score matching using alternative caliper widths (0.2 vs. 0.3 SD) on the logit of the propensity score.
- **country_meta.xlsx –** Results of stratified propensity score matching analyses by region (Spain vs. non-Spain), including stratum-specific pooled ATT estimates.
- **dataanalysis.docx –** Extended Data Analysis methods, including multiple imputation, SFI construction/validation, mutual information setup, PSM specifications (trimming/calipers/matching), balance diagnostics, doubly robust models, and stratified analyses.
- **Descriptive.docx –** Summary tables of baseline demographic, social, and clinical characteristics by registry (REGISPONSER and RESPONDIA), including distributions and missingness.
- **regression_psm_dr.xlsx** – Results of multivariable regression analyses performed after propensity score matching. For each imputed dataset, the file reports the treatment effect estimate (β for Q4 vs. Q1 of the Social Factors Index) with cluster-robust standard errors by matched pair, 95% confidence intervals, and the list of covariates included when residual imbalance was observed

Acknowledgements: The Authors would like to thank all the investigators from the REGISPONSER and RESPONDIA study groups. We also like to acknowledge the support of technological tools such as ChatGPT, whose assistance has helped improve the fluency and clarity of English language expression in some sections of this document.

Table 1

Variable	Dataset (N=2042)	REGISPONSER (N=1514)	RESPONDIA (N=528)
Demographics			
Age, years	47.3 ± 13.4	48.1 ± 12.9	44.7 ± 14.7
Diagnostic delay, years	7.5 ± 9.3	7.7 ± 9.4	6.8 ± 9.1
Disease duration, years	19.5 ± 13.1	21.1 ± 13.1	14.8 ± 11.6
Symptom duration, years	11.7 ± 10.3	13.2 ± 10.4	7.4 ± 8.5
Age at diagnosis, years	35.4 ± 12.7	34.9 ± 11.9	37.0 ± 14.9
Weight, kg	73.8 ± 14.2	74.2 ± 14.2	72.4 ± 14.2
Height, cm	166.4 ± 9.3	166.7 ± 9.1	165.4 ± 10.0
Clinical measures			
Chest expansion, cm	3.6 ± 2.1	3.8 ± 2.2	3.3 ± 1.9
Schober test, cm	3.0 ± 1.9	3.0 ± 1.8	3.2 ± 2.1
Finger-to-floor distance, cm	19.0 ± 15.0	18.6 ± 14.2	20.2 ± 16.8
Occiput-to-wall distance, cm	4.5 ± 6.2	4.4 ± 6.0	4.9 ± 6.6
Swollen joint count	0.8 ± 2.9	0.4 ± 1.8	1.9 ± 4.8
VAS (pain 0–10)	3.3 ± 2.3	3.2 ± 2.1	3.7 ± 2.6
ASQoL (0–18 higher=worse)	6.6 ± 5.2	6.4 ± 5.1	7.4 ± 5.3
SF-12 Physical Component	37.0 ± 7.5	37.2 ± 7.5	36.3 ± 7.5
SF-12 Mental Component	50.2 ± 5.5	50.6 ± 5.5	49.3 ± 5.5
ASDAS-CRP	2.4 ± 1.1	2.6 ± 1.1	2.0 ± 1.0
Form			
Axial	1233 (60.9%)	1055 (69.8%)	178 (34.6%)
Mixed	744 (36.7%)	433 (28.7%)	311 (60.5%)
Peripheral	39 (1.9%)	19 (1.3%)	20 (3.9%)
Enthesitic	9 (0.4%)	4 (0.3%)	5 (1.0%)
Family history			
No	1513 (79.5%)	1117 (79.6%)	396 (79.2%)
Yes	390 (20.5%)	286 (20.4%)	104 (20.8%)
Gender			
Men	1485 (72.7%)	1131 (74.7%)	354 (67.0%)
Women	557 (27.3%)	383 (25.3%)	174 (33.0%)
Education			
High school	606 (41.6%)	359 (38.1%)	247 (47.9%)
Elementary school	556 (38.1%)	433 (46.0%)	123 (23.8%)
Illiterate	20 (1.4%)	13 (1.4%)	7 (1.4%)
University	276 (18.9%)	137 (14.5%)	139 (26.9%)
Marital status			
Married	1023 (70.2%)	713 (75.6%)	310 (60.2%)
Single	364 (25.0%)	194 (20.6%)	170 (33.0%)
Divorced/Separated	48 (3.3%)	25 (2.7%)	23 (4.5%)
Widowed	23 (1.6%)	11 (1.2%)	12 (2.3%)

Profession			
Unskilled worker	377 (25.9%)	238 (25.3%)	139 (27.0%)
Skilled worker	349 (24.0%)	283 (30.1%)	66 (12.8%)
Employee	309 (21.3%)	201 (21.4%)	108 (21.0%)
Technician	219 (15.1%)	129 (13.7%)	90 (17.5%)
University profession	200 (13.8%)	88 (9.4%)	112 (21.7%)
Source of income			
Salary	1022 (70.3%)	643 (68.5%)	379 (73.6%)
Donations	216 (14.9%)	150 (16.0%)	66 (12.8%)
Profits	113 (7.8%)	61 (6.5%)	52 (10.1%)
Rents	103 (7.1%)	85 (9.1%)	18 (3.5%)
Housing conditions			
Good housing	735 (50.5%)	463 (49.2%)	272 (53.0%)
Excellent (no luxury)	588 (40.4%)	411 (43.6%)	177 (34.5%)
Deficient	86 (5.9%)	47 (5.0%)	39 (7.6%)
Excellent with luxury	24 (1.6%)	9 (1.0%)	15 (2.9%)
Inadequate	22 (1.5%)	12 (1.3%)	10 (1.9%)
Graffar scale			
Middle-lower class	594 (41.0%)	411 (43.9%)	183 (35.7%)
Middle class	443 (30.6%)	288 (30.8%)	155 (30.2%)
Upper-middle class	285 (19.7%)	166 (17.7%)	119 (23.2%)
Lower class	95 (6.6%)	59 (6.3%)	36 (7.0%)
Upper class	32 (2.2%)	12 (1.3%)	20 (3.9%)
Race			
White	1193 (81.4%)	930 (98.3%)	263 (50.6%)
White-Indigenous	201 (13.7%)	2 (0.2%)	199 (38.6%)
White-Black	23 (1.6%)	3 (0.3%)	20 (3.8%)
Other	18 (1.2%)	7 (0.7%)	11 (2.1%)
Black	13 (0.9%)	1 (0.1%)	12 (2.3%)
Indigenous	10 (0.7%)	1 (0.1%)	9 (1.7%)
Black-Indigenous	5 (0.3%)	1 (0.1%)	4 (0.8%)
Indigenous-Asian	3 (0.2%)	1 (0.1%)	2 (0.4%)
Employment status			
Employed	1038 (51.6%)	738 (49.5%)	300 (57.6%)
At home	580 (28.8%)	525 (35.2%)	55 (10.6%)
Unemployed	309 (15.4%)	182 (12.2%)	127 (24.4%)
Student	19 (0.9%)	5 (0.3%)	14 (2.7%)
Disability			
No disability	1417 (72.0%)	1047 (71.0%)	370 (75.2%)
Permanent disability	444 (22.6%)	367 (24.9%)	77 (15.7%)
Temporary disability	106 (5.4%)	61 (4.1%)	45 (9.1%)
Country			
Spain	1514 (74.1%)	1514 (100.0%)	0 (0%)
Argentina	147 (7.2%)	0 (0%)	147 (27.8%)

Mexico	103 (5.0%)	0 (0%)	103 (19.5%)
Portugal	88 (4.3%)	0 (0%)	88 (16.7%)
Chile	76 (3.7%)	0 (0%)	76 (14.4%)
Venezuela	38 (1.9%)	0 (0%)	38 (7.2%)
Peru	32 (1.6%)	0 (0%)	32 (6.1%)
Uruguay	28 (1.4%)	0 (0%)	28 (5.3%)
Costa Rica	16 (0.8%)	0 (0%)	16 (3.0%)

*Values are presented as mean ± standard deviation or n (%). Percentages use the variable-specific denominator due to missing data. Abbreviations: VAS = Visual Analogue Scale; ASQoL = Ankylosing Spondylitis Quality of Life; ASDAS = Ankylosing Spondylitis Disease Activity Score.

Supporting Information

DataS1

Supinfo1

DataS2

DataS3

Supinfo2

Supinfo3

DataS4

DataS5

DataS6

DataS7

Title: The impact of social factors on VAS pain in spondyloarthritis patients: a Propensity Score Matching analysis

Running head: Social factors and VAS pain in SpA

Authors:
Luis Ángel Calvo Pascual, B.Sc., M. Sc., Ph.D. (0000-0002-8818-5542) Universidad Pontificia de Comillas Madrid. Spain. lacalvo@comillas.edu

David Castro Corredor, M.D., Ph.D. (0000-0001-7315-6274) Hospital General Universitario de Ciudad Real. Spain. d.castrocorredor@gmail.com

Eduardo Collantes-Estévez, M.D., Ph.D. (0000-0002-7647-6289) Maimonides Institute for Research in Biomedicine of Cordoba (IMIBIC), Reina Sofia University Hospital, University of Cordoba, Córdoba, Spain. educollantes@yahoo.es

Clementina López-Medina, M.D., Ph.D. (0000-0002-2309-5837) Maimonides Institute for Research in Biomedicine of Cordoba (IMIBIC), Reina Sofia University Hospital, University of Cordoba, Córdoba, Spain. clementinalopezmedina@gmail.com

Correspondence to:
Luis Ángel Calvo Pascual, ICADE. Universidad Pontificia de Comillas, C/Alberto Aguilera, 23, 28015, Madrid. Spain. lacalvo@comillas.edu

Disclosure statements:
a) Funding statement: The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The REGISPONSER registry was funded by the Spanish Society of Rheumatology. There was no funding for this subanalysis.
b) Informed consent/ Patient consent: All patients gave their written consent to participate in the REGISPONSER and RESPONDIA registries.
c) Ethical Statement: The study was conducted in compliance with the recommendations of the Declaration of Helsinki and was centrally approved by the Ethics Committee of the Reina Sofia University Hospital from Córdoba (Spain) on 21 April 2006 (“Comisión de Ética e Investigación Sanitaria, REGISPON-2004 y RESPONDIA-2006”).

Word count: 3638

Abstract: [word count: 248/250]

Background: Pain in spondyloarthritis reflects not only inflammatory activity but also social context. We quantified the association between a composite Social Factors Index (SFI) and patient-reported pain severity on the Visual Analogue Scale (VAS).

Methods: We analyzed 2,042 patients from REGISPONSER (n=1,514) and RESPONDIA (n=528). Missing data were handled with multiple imputation (m=20). The SFI combined education, employment, housing, income source, and Graffar class using penalized regression (sparse-group lasso). Mutual information (MI) quantified the SFI's relative importance. Propensity scores (logistic regression) supported 1:1 caliper matching (0.2 SD on the logit), comparing the highest (Q4) vs lowest (Q1) SFI quartiles.

Results: Before matching, Q1 and Q4 included on average 918 and 829 patients per imputation, differing mainly in race (White 80.8% vs 86.8%; SMD=−0.16) and country (Spain 66.2% vs 78.2%; SMD=−0.27). After matching, a mean of 745 pairs (range 731–761) were retained with excellent covariate balance (all |SMD|<0.10). In the matched sample, patients in Q4 reported higher pain (Average Treatment effect on the Treated (ATT) +0.38 VAS points; 95% CI: 0.05–0.71; p=0.023). Post-matching regression yielded an identical estimate (β=+0.37; 95% CI: 0.04–0.70; p=0.030). MI indicated the SFI carried more information about VAS than any single social variable. Stratified analyses suggested heterogeneity: Spain +0.31 (95% CI: 0.08–0.71) and Latin America +0.62 (95% CI: −0.22 to 1.45).

Conclusions: Social disadvantage, summarized by the SFI, is independently associated with greater pain severity in spondyloarthritis after adjustment. These findings support incorporating social determinants into clinical assessment and pain management.

__Keywords:__ Mutual Information, Propensity Score Matching, VAS, Spondyloarthritis.

Significance and Innovations

- We developed a Social Factors Index (SFI) using sparse-group lasso with bootstrap across multiply imputed datasets; mutual information quantified the SFI's informativeness for patient-reported VAS pain.

- The SFI captured more mutual information with VAS than any single social variable, including Graffar Scale class, employment status, housing conditions, or income.

- Using propensity-score matching (>700 matched pairs per imputation), patients in the highest SFI quartile reported +0.38 VAS points vs the lowest (95% CI: 0.05–0.71; p=0.023), supporting an association independent of measured disease activity

- Findings underscore the clinical relevance of social vulnerability in SpA and support integrating social determinants into comprehensive care and management.

INTRODUCTION

Spondyloarthritis (SpA) encompasses a group of chronic inflammatory rheumatic diseases that primarily affect the axial skeleton, leading to significant physical limitations and diminished quality of life. It often strikes individuals in early adulthood, causing chronic back pain, stiffness, and progressive loss of mobility. Consequently, SpA is associated with severe physical limitations, functional impairment, and decreased quality of life[1].

Pain is a central and enduring symptom of SpA, traditionally attributed to inflammatory processes. However, a substantial subset of patients continues to experience significant pain despite effective control of inflammation through pharmacological interventions. This persistent pain suggests the involvement of non-inflammatory mechanisms, notably central sensitization, an increased responsiveness of nociceptive neurons in the central nervous system, leading to heightened pain sensitivity and perception[2].

Beyond biological mechanisms, social determinants have emerged as important contributors to pain severity in SpA[3,4]. Lower educational attainment and socioeconomic disadvantage are independently associated with higher patient-reported pain, even after adjustment for disease activity and treatment access[5-7]. In a cluster analysis, patients in the low-socioeconomic group experienced longer diagnostic delays, higher body mass index, and more severe structural damage despite adjustment for disease duration; although access to biologic disease-modifying anti-rheumatic drugs (DMARDs) was comparable across groups, disadvantaged patients had a substantially higher prevalence of permanent work disability, defined as official certification of long-term incapacity for work, independent of disease activity and duration[8]. Moreover, patients with chronic low back pain reported significantly worse pain severity, disability, and health-related quality of life than those with axial spondyloarthritis, despite similar pain intensity scores, underscoring that pain experience extends beyond inflammatory burden[9].

The present study investigates whether social determinants contribute to patient-reported pain in spondyloarthritis (SpA) using data from two registries (REGISPONSER and RESPONDIA). We examined education, employment status, housing conditions, and Graffar social class, and aggregated them into a Social Factors Index (SFI) whose component weights were learned via lasso regularization (sparse-group lasso) with bootstrap resampling across multiply imputed datasets. We then quantified the importance of the SFI for Visual Analogue Scale (VAS) pain using mutual information, and applied propensity-score matching to balance key demographic and clinical covariates, contrasting patients in the highest versus lowest SFI quartiles.

PATIENTS AND METHODS

Study design:

This study was based on data obtained from two multicenter observational registries: REGISPONSER and RESPONDIA, both specifically designed to collect standardized clinical information on patients with spondyloarthritis (SpA).

REGISPONSER is a national, multicenter registry from Spain that prospectively enrolled consecutive patients diagnosed with SpA according to the European Spondyloarthropathy Study Group (ESSG) criteria. Recruitment was conducted between March 2004 and March 2007 across 31 rheumatology departments. Detailed information regarding the design, sampling methodology, and recruitment process can be found in Collantes et al.[10].

RESPONDIA, a parallel registry with an identical study design, was conducted between 2006 and 2007 in Latin America. It included 33 centers across eight countries: Argentina, Chile, Colombia, Costa Rica, Mexico, Uruguay, and Venezuela. Like REGISPONSER, RESPONDIA included consecutive patients who met the ESSG classification criteria for SpA and used the same case report forms, clinical definitions, and variable collection procedures[11].

For this study, data from both sources were merged to generate a large, multinational cohort of SpA patients.

Inclusion and exclusion criteria:

All patients aged 18 years or older with a diagnosis of spondyloarthritis (SpA) according to the European Spondyloarthropathy Study Group (ESSG) criteria were eligible. Participants were drawn from the REGISPONSER registry (Spain, 2004–2007) and the RESPONDIA registry (Latin America, 2006–2007). Patients with incomplete clinical information in key exposure or outcome variables, or those with duplicate records, were excluded. No additional exclusion criteria were applied.

Outcomes:

The primary outcome of interest was pain intensity, measured using the Visual Analogue Scale (VAS), which ranged from 0 (no pain) to 10 (worst imaginable pain), referring to overall pain experienced during the previous week. No formal secondary outcomes were prespecified.

Variables:

This study analyzed data from 2042 patients diagnosed with spondyloarthritis. A total of 66 variables were collected and organized into four conceptual domains: personal variables, social variables, treatment-related variables, and disease activity. All variables were assessed at the baseline visit of each registry, with laboratory tests performed within 15 days before inclusion. Detailed information is provided in the supplementary material (*Descriptive.docx*), which includes six tables: three with categorical variables and three with numerical variables, each presented for the full sample, for REGISPONSER, and for RESPONDIA.

The personal variables include age, weight, height, physical activity habits, smoking status, race/ethnicity and country of residence across Ibero-American regions such as Spain, Argentina, Mexico, and others. Genetic and familial factors were captured through HLA-B27 status (positive or negative) and the presence or absence of a family history of spondyloarthritis.

The social variables domain was composed of educational attainment, ranging from illiterate to university level. Employment status included categories such as employed, unemployed, student, retired, or homemaker. Socioeconomic stratification was measured using the Graffar scale[12], while housing conditions ranged from inadequate to excellent, with luxury. Marital status, sources of income (salary, profits, donations, rents), and professional background (e.g., skilled or unskilled labour, technician, employee) were also recorded. All social variables were systematically collected for all participants at the baseline visit.

Therapeutic exposure variables consisted of 15 measures of pharmacological treatment and management, including corticosteroids, NSAIDs (regular, on-demand, or none), biologic therapies (infliximab, etanercept, adalimumab), and conventional synthetic DMARDs (methotrexate, sulfasalazine, leflunomide).

The disease activity variables included measures of clinical status, functional impairment, and patient-reported outcomes. Disease activity was assessed with the Ankylosing Spondylitis Disease Activity Score (ASDAS). Functional impact and quality of life were measured using the Ankylosing Spondylitis Quality of Life (ASQoL) questionnaire and the physical and mental component scores of the SF-12 survey. Axial and peripheral involvement was evaluated through the number of swollen joints, chest expansion, and Schober's test. The dataset also recorded the clinical form of the disease (axial, peripheral, enthesitic, or mixed), extra-articular manifestations (iritis, uveitis, enthesitis, dactylitis, pustulosis, acne conglobata, balanitis, prostatitis, and nail involvement), comorbidities (cardiac, renal, pulmonary, and neurological) and disease duration (years from symptom onset to the baseline visit). Pain was assessed with the Visual Analogue Scale (VAS), where patients rated their overall pain (not restricted to back pain) from 0 (no pain) to 10 (worst imaginable pain) based on the previous week.

Data analysis:

We merged REGISPONSER and RESPONDIA, removed outliers and data errors. We summarized numerical and categorical variables to characterize the sample; baseline demographic, clinical, and social characteristics are shown in Table 1. More detailed distributions and missing-data summaries by cohort, together with boxplots of VAS across categorical variables and kernel density plots for numerical variables, are provided in the Supplementary Material (Descriptive.docx; Boxplot.docx).

Given non-trivial missingness, we used multiple imputation to create m = 20 completed datasets (13). We rejected MCAR via Little's test and proceeded under a Missing-At-Random assumption (14). To study social determinants of pain, we built a Social Factors Index (SFI) from education, employment/occupation, income source, housing, and Graffar. In each imputation, models used penalization with out-of-fold predictions to avoid optimism; when group-lasso was unavailable, we used elastic net with internal cross-validation (15). Individual SFI scores were averaged across imputations. We evaluated SFI performance via correlation and discrimination for high pain (AUC for VAS ≥ 7), reporting per-imputation and pooled metrics (Fig. 1).

To quantify associations with VAS beyond linear effects, we computed mutual information (MI) between VAS and candidate predictors—including SFI—within each imputed dataset, following recent applications in rheumatology (16–19). We used the standard definition

$$\mathrm{MI}(X, VAS) = H(VAS) - H(VAS \mid X),$$

where $H(\cdot)$ is the Shannon entropy function. Practically, we discretized VAS into five quantile-based bins, encoded categorical variables, scaled continuous predictors, and estimated MI with bootstrap uncertainty (100 resamples per imputation), then pooled ranks across imputations (Fig. 2). MI guided the choice of pre-exposure covariates for causal analyses.

We estimated the average treatment effect on the treated (ATT) by contrasting high (Q4) vs. low (Q1) SFI using propensity score matching with logistic-model scores, overlap trimming (20), 1:1 nearest-neighbor matching without replacement on the logit scale, and a 0.2 SD caliper (with 0.3 SD as a sensitivity check) (21). Balance was assessed using standardized mean differences, and ATT estimates with standard errors were pooled across imputations via Rubin's rules. We additionally fit doubly robust outcome regressions for covariates with residual imbalance. The pooled ATT (95% CIs) is shown in Figure 3 (forest plot), covariate balance before and after matching in Figure 4 (love plot), and paired post-matching VAS differences in Figure 5 (distribution of matched differences). Full specifications, diagnostics, and sensitivity analyses are provided in the Supplementary Methods (dataanalysis.docx); Supplementary Material also includes: SFI validation outputs and metrics (social_index_eval.xlsx), full PSM results with per-imputation and pooled ATT (psm_results.xlsx), detailed balance tables (balance_tables.xlsx), doubly robust regression outputs (regression_psm_dr.xlsx), caliper sensitivity (0.2 vs 0.3 SD; caliper_sensitivity.xlsx), and stratified Spain vs non-Spain analyses (country_meta.xlsx).

RESULTS

In total, 2,042 patients were included (REGISPONSER n = 1,514; RESPONDIA n = 528). A concise summary of baseline demographic, clinical, and social characteristics is provided in Table 1. Patients from REGISPONSER were on average older (48.1 years, SD 12.9) than those from RESPONDIA (44.7 years, SD 14.7), and the proportion of males was higher (74.7% vs. 67.0%). Regarding disease activity and pain, the mean VAS score was 3.20 (SD 2.13) in REGISPONSER and 3.72 (SD 2.58) in RESPONDIA, while ASDAS was 2.58 (SD 1.12) and 2.03 (SD 1.01), respectively. Socioeconomic differences were also observed: for example, university education and employment were more common in RESPONDIA. Comparative boxplots of all variables can be consulted in the supplementary file boxplots.xlsx, where it can also be observed that patients in the Graffar upper class reported lower VAS levels. Detailed descriptive statistics (means, SD, quartiles, n and %, missing data) by registry are provided in descriptive.docx.

The Social Factor Index (SFI) was defined as a weighted sum of nine social factors each compared to the middle baseline of its domain:

$$
\begin{aligned}
SFI = {} & +0.0467\,[Employment\ Status\_student] \\
& + 0.0177\,[Employment\ Status\_Unemployed] \\
& - 0.0175\,[Employment\ Status\_Employed] \\
& + 0.0179\,[Education\_Illiterate] - 0.0236\,[Education\_University] \\
& + 0.0414\,[Housing\ Conditions\_Inadequate] \\
& - 0.0224\,[Housing\ Conditions\_Deficients] \\
& + 0.0246\,[Graffar\ Scale\_Lower\ class] \\
& - 0.0211\,[Graffar\ Scale\_Upper\ class].
\end{aligned}
$$

Positive coefficients indicate higher expected pain (VAS) relative to baseline. The index showed a modest but consistent association with VAS (Pearson r≈0.17) and modest discrimination for high pain (VAS ≥ 7, AUC ≈ 0.56). However, the boxplot of VAS across SFI quartiles (Fig. 1) showed an upward shift in both medians and interquartile ranges, suggesting higher pain levels with increasing SFI.

Mutual-information ranking (Fig. 2) showed that the Social Factors Index captured more information about VAS than any single social variable. The pooled MI for the SFI index was 0.0239, placing it 20th overall and ahead of other social indicators such as the Graffar scale (0.0124), Employment status (0.0134), Housing conditions (0.0144), Source of income (0.0102), Race (0.0151), and Country (0.0257). As expected, clinical/activity measures dominated the top ranks (e.g., ASDAS 0.579, ASQoL 0.237, Weight 0.104, SF-12 Physical 0.156). As we described in the previous section, we also used these MI results to select those variables high ranked and plausibly relate to both social risk and pain. This led to the final set of baseline covariates: Age (0.074), Diagnosis Age (0.066), Disease Duration (0.059), Diagnostic Delay (0.044), Weight (0.104), Height (0.051), Race (0.015), and Country (0.026).

Regarding the results of the propensity score matching analysis, we compared pain outcomes between patients with high social factor scores (Q4 of the SFI) and those with low scores (Q1). Before matching, Q1 and Q4 included, on average, 918 and 829 patients per imputation, respectively, reflecting roughly one quarter of the cohort. The pooled average treatment effect on the treated (ATT) was +0.38 VAS points (SE = 0.17), with a 95% confidence interval of 0.05 to 0.71 and a p value of 0.023 across the 20 imputations (Fig. 3). After overlap trimming and 1:1 nearest-neighbour caliper matching, an average of 745 matched pairs per imputation was retained (range 731–761), corresponding to about 90% of eligible patients in the extreme quartiles. Matching quality was excellent, with a mean absolute standardized difference of 0.041 across covariates. Figure 4 displays the pooled standardized mean differences (SMDs) across imputations for all variables in the propensity model, weighted by matched pairs. Before matching (grey dots), some covariates showed imbalance, particularly Race (SMD = –0.16) and Country (SMD = –0.27). After matching, these imbalances were reduced to –0.09 and –0.04, respectively, while continuous covariates such as age, diagnosis age, disease duration, delay, weight, and height all achieved post-match SMDs between –0.07 and 0.04. Thus, every covariate met the conventional threshold of |SMD| < 0.10. In the matched sample, the distribution of pairwise differences in VAS remained shifted to the right, with a pooled mean difference of +0.38 points (95% CI: 0.05–0.71), confirming that pain severity was consistently higher in the high-SFI group.

The results of the caliper sensitivity analysis (see caliper_sensitivity.xlsx, Supplementary Material) were highly consistent across specifications. When widening the caliper to 0.3 SD, results were almost identical to those obtained with 0.2 SD (ATT = +0.41; SE = 0.16; 95% CI: 0.09–0.72; p = 0.011), with a similar number of matched pairs (802) and a mean absolute standardized mean difference after matching of 0.022.

Finally, in the stratified analysis to observe differences between Spain and Latin American countries, we obtained an average treatment effect on the treated of +0.31 (SE = 0.20; 95% CI: 0.08 to 0.71) across 20 imputations for Spain, and +0.62 (SE = 0.43; 95% CI: −0.22 to 1.45) across 20 imputations for Latin America.

DISCUSSION

Some previous studies have primarily attributed the impact of socio-economic factors not to pain, but to disease activity. For example, Capelusnik et al.[7] suggest that lower educational levels and single marital status were associated with higher ASDAS. Other studies[23, 24] found that lower levels of education and income were associated with a higher likelihood of receiving a chronic pain diagnosis.

Our Social Factors Index (SFI) shows a clear gradient: students and unemployed individuals, those living in inadequate housing, and patients from lower social classes have higher predicted pain, whereas those with university education and from upper social classes have lower predicted pain. These patterns align with evidence that social determinants are linked to worse outcomes in axial spondyloarthritis[7]. Mendelian-randomization work indicates that higher educational attainment causally reduces multisite chronic pain risk, plausibly via healthier behaviours and greater psychological resources[25]. Daily diary research also shows that financial hardship and unemployment amplify day-to-day pain reactivity, particularly among women with chronic musculoskeletal disease who worry about finances when unemployed[26]. Taken together, this supports aggregating correlated social exposures to reflect cumulative disadvantages.

Despite the modest performance of the SFI in predicting VAS (Pearson r ≈0.17; AUC ≈ 0.56), a low overall correlation does not rule out meaningful group differences. When we group SFI into quartiles, Figure 1 shows that median VAS (and its spread) increases stepwise from Q1 to Q4. In addition, the SFI ranked first among all social indicators in the mutual-information analysis, ahead of Education, Graffar scale, Employment status, Housing conditions, and Source of income (Figure 2). This means that, while social context explains less variance in pain than clinical activity, the combined index captures more information about VAS than any single social variable, supporting its use as a compact summary of cumulative social factors.

Using propensity-score matching to compare patients in the highest SFI quartile (Q4) with those in the lowest (Q1), the pooled average treatment effect on the treated (ATT) for VAS was +0.38 points (SE = 0.17; 95% CI: 0.05–0.71; p = 0.023) across 20 imputations (Figure 3). Covariate balance was excellent after matching (mean absolute SMD ≈ 0.041) with ~745 matched pairs per imputation (Figures 4, 5), and results were stable to caliper changes (0.2–0.3 SD), consistent with prior methodological work[27-29]. In plain terms, patients with high SFI reported ~0.4 points more VAS pain than comparable patients with low SFI. Although modest, this effect was consistent and clinically relevant. The gradient was more pronounced in Latin America (ATT +0.62; 95% CI: −0.22 to 1.45) than in Spain (ATT +0.31; 95% CI: −0.08 to 0.71), suggesting stronger social disparities in the former.

Our methodology has several strengths. We used multiple imputations (20 datasets) and derived a stable, interpretable SFI via sparse-group lasso with bootstrap resampling. Mutual information was applied only to assess the SFI's informativeness for VAS, not for confounder selection. Prespecified baseline covariates (age, age at diagnosis, disease duration, diagnostic delay, weight, height, race, country) achieved excellent post-match balance (mean absolute SMD ≈ 0.041). Propensity scores were estimated with logistic regression, prioritizing balance over prediction; effects were robust to caliper variation on the logit scale (0.2–0.3 SD) and were larger in Latin America than in Spain.

However, several limitations should be noted. First, the study is observational and cross-sectional rather than longitudinal, which limits the ability to establish causal relationships and raises the possibility of reverse causation. Second, although we adjusted for a wide range of demographic and clinical covariates, residual confounding cannot be excluded, particularly from unmeasured psychosocial factors such as depression, coping styles, health literacy, or differences in healthcare access. Third, the SFI was developed primarily in a Spanish cohort, and while the effect appeared stronger in Latin America, where social inequalities are more pronounced, external validation and larger datasets from these countries are needed to confirm and generalize these findings.

CONCLUSIONS

This study suggests that social factors could contribute independently to pain severity in patients with spondyloarthritis (SpA). We summarized education, employment status, Graffar scale and housing conditions into a single Social Factors Index (SFI), showing a clear gradient in pain: VAS increased stepwise from the lowest to the highest SFI quartile. With propensity-score matching, patients in the highest SFI quartile had modest yet consistent increases in VAS compared with otherwise comparable patients in the lowest quartile (ATT = +0.38; 95% CI: 0.05–0.71; $p = 0.023$), supporting an effect beyond measured disease activity. Although the individual-level effect size is small, such gradients can be meaningful at the population level and have policy relevance, helping identify groups who may benefit from targeted support. Clinically, patients with lower education, unstable employment, or inadequate housing may benefit from targeted assessment and support (e.g., pain self-management support, social work referral). Future work should externally validate the SFI and matched findings, use longitudinal designs to establish temporality, and test interventions that address the identified social pathways.

REFERENCES

1. Mease, P. J., Bhutani, M. K., Hass, S., Yi, E., Hur, P., Kim, N. (2022). Comparison of Clinical Manifestations in Rheumatoid Arthritis vs. Spondyloarthritis: A Systematic Literature Review. Rheumatology and Therapy, 9(2), 331–378. https://doi.org/10.1007/s40744-021-00407-8

2. Lee, Y. C., Malfait, A. M., Ogdie, A. R. (2024). Unmet needs in spondyloarthritis: Understanding and managing chronic pain. The Journal of Rheumatology, 51(12), 1235–1240. https://doi.org/10.3899/jrheum.2024-0936

3. Kiltz, U., Baraliakos, X., Regel, A., Bühring, B., & Braun, J. (2017). Causes of pain in patients with axial spondyloarthritis. Clin Exp Rheumatol, 35(Suppl 107), 102-107.

4. Pathan, E. M., Inman, R. D. (2017). Pain in spondyloarthritis: A neuro–immune interaction. *Best Practice & Research Clinical Rheumatology, 31*(6), 830–845. https://doi.org/10.1016/J.BERH.2018.07.003

5. Capelusnik, D., Boonen, A., Ramiro, S., Nikiphorou, E. (2025). The role of social determinants of health on disease outcomes in axial spondyloarthritis: A narrative review. *Autoimmunity Reviews, 24*(5), Article 103762. https://doi.org/10.1016/j.autrev.2025.103762

6. Santos, H., Henriques, A. R., Branco, J., Machado, P. M., Canhão, H., Pimentel-Santos, F. M., Rodrigues, A. M. (2023). Health-related quality of life among spondyloarthritis and chronic low back pain patients: Results from a nationwide population-based survey. Quality of Life Research, 32(2), 383–399. https://doi.org/10.1007/s11136-022-03274-0

7. Capelusnik, D., Zhao, S. S., Boonen, A., Ziade, N., Medina, C. L., Dougados, M., Nikiphorou, E., Ramiro, S. (2022). Individual-level and country-level socio-economic factors and health outcomes in spondyloarthritis: Analysis of the ASAS-perSpA study. Rheumatology (Oxford, England), 61(5), 2043–2053. https://doi.org/10.1093/rheumatology/keab638

8. Ruiz-Vilchez, D., Ladehesa-Pineda, L., Puche-Larrubia, M. Á., Ábalos-Aguilera, M. C., Font-Ugalde, P., Escudero-Contreras, A., López-Medina, C. (2024). The socioeconomic status of patients with ankylosing spondylitis and its association with the burden of the disease and permanent disability: A cross-sectional cluster analysis. Therapeutic Advances in Musculoskeletal Disease, 16, Article 1759720X241272947. https://doi.org/10.1177/1759720X241272947

9. Cheung, J. P. Y., Cheung, P. W. H., Wong, C. K. H., Chung, H. Y., Tsang, H. H. L. (2023). Propensity-matched comparison between chronic nonspecific low back pain and axial spondyloarthritis: impact on patient-perceived quality of life. Spine, 48(8), 577-589.

10. Collantes, E., Zarco, P., Muñoz, E., et al. (2007). Disease pattern of spondyloarthropathies in Spain: Description of the first national registry (REGISPONSER) – extended report. Rheumatology (Oxford), 46(8), 1309–1315. https://doi.org/10.1093/rheumatology/kem084

11. Vazquez-Mellado J, Font Ugalde P, Muñoz Gomariz E, et al. Ibero-American Spondylarthritis Registry (RESPONDIA): what is, how it came about, who we are and what we do. Reumatología Clínica 2008; 4:S17–22.

12. Graffar, M. (1956). *Une méthode de classification sociale d'échantillons de population*. *Courrier de l'Office International d'Hygiène Publique*.

13. Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: John Wiley & Sons.

14. Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. Journal of the American Statistical Association, 83(404), 1198–1202. https://doi.org/10.1080/01621459.1988.10478722

15. Friedman, J., Hastie, T., Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, 33*(1), 1–22. https://doi.org/10.18637/jss.v033.i01

16. Castro Corredor, D., Calvo Pascual, L. Á., Collantes-Estévez, E., López-Medina, C. (2025). *Inflammatory activity levels in patients with anti-TNF therapy: Most important factors and a decision tree model. Data from REGISPONSER and RESPONDIA registries*. Therapeutic Advances in Musculoskeletal Diseases, 17, https://doi.org/10.1177/1759720X251332224

17. Calvo Pascual L. A., Castro Corredor D., Garrido Merchán E. C., Machine learning classification of vitamin D levels in spondyloarthritis patients. Intelligence-Based Medicine. Volume 9. 2024. 100125. ISSN 2666-5212. https://doi.org/10.1016/j.ibmed.2023.100125.

18. Castro Corredor D, Calvo Pascual L. A. (2023) Imbalanced machine learning classification models for removal biosimilar drugs and increased activity in patients with rheumatic diseases. PLoS ONE 18(11): e0291891. https://doi.org/10.1371/journal.pone.0291891

19. Curtis, J. R., Su, Y., Black, S., Xu, S., Langholff, W., Bingham, C. O., Kafka, S., Xie, F. (2022). Machine learning applied to patient-reported outcomes to classify physician-derived measures of rheumatoid arthritis disease activity. ACR Open Rheumatology, 4(12), 995–1003. https://doi.org/10.1002/acr2.11499

20. Crump, R. K., Hotz, V. J., Imbens, G. W., Mitnik, O. A. (2009). *Dealing with limited overlap in estimation of average treatment effects*. **Biometrika**, 96(1), 187–199. https://doi.org/10.1093/biomet/asn055

21. Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Statistics in Medicine, 28(25), 3083–3107. https://doi.org/10.1002/sim.3697

22. DerSimonian, R., Laird, N. (1986). Meta-analysis in clinical trials. Controlled Clinical Trials, 7(3), 177–188. https://doi.org/10.1016/0197-2456(86)90046-2

23. Jöud, A., Petersson, I. F., Jordan, K. P., Löfvendahl, S., Grahn, B., & Englund, M. (2014). Socioeconomic status and the risk for being diagnosed with spondyloarthritis and chronic pain: A nested case–control study. Rheumatology International, 34, 1291–1298. https://doi.org/10.1007/s00296-014-3039-6

24. Prego-Domínguez, J., Khazaeipour, Z., Mallah, N., & Takkouche, B. (2021). Socioeconomic status and occurrence of chronic pain: A meta-analysis. Rheumatology, 60(3), 1091–1105. https://doi.org/10.1093/rheumatology/keaa758

25. Wang, H., Liu, M., Li, H., Xu, S. (2025). Association between educational attainment and chronic pain: A mediation Mendelian randomization study. *Journal of Pain Research, 18*, 1793–1804. https://doi.org/10.2147/JPR.S515921

26. Rios, R., Zautra, A. J. (2011). Socioeconomic disparities in pain: The role of economic hardship and daily financial worry. *Health Psychology, 30*(1), 58–66. https://doi.org/10.1037/a0022025

27. Lee, B. K., Lessler, J., Stuart, E. A. (2010). Improving propensity score weighting using machine learning. Statistics in Medicine, 29(3), 337–346. https://doi.org/10.1002/sim.3782

28. Pirracchio, R., Petersen, M. L., van der Laan, M. J. (2015). Improving propensity score estimators' robustness to model misspecification using super learner. American Journal of Epidemiology, 181(2), 108–119. https://doi.org/10.1093/aje/kwv290

29. Zhao, P., Su, X. G., Ge, T., & Fan, J. (2016). Propensity score and proximity matching using random forest. Contemporary Clinical Trials, 47, 85–92. https://doi.org/10.1016/j.cct.2015.12.012

FIGURE LEGENDS

Figure 1. Distribution of pain scores (VAS) across quartiles of the Social Factors Index (SFI)

Figure 2. Mutual information with VAS for the 30 clinical and social features, computed across 100 bootstrap iterations. We highlight the SFI index in red to show its relative position with respect to the other variables.

Figure 3. Forest plot of ATT estimates for VAS across imputations and pooled results.

Figure 4. Love plot depicting covariate balance before and after PSM

Figure 5. Distribution of paired differences in pain scores after matching

TABLE LEGENDS

Table 1. Baseline characteristics of patients with spondyloarthritis in the joint pooled dataset, REGISPONSER and RESPONDIA cohorts.

SUPPLEMENTARY FILES

- **social_index_eval.xlsx –** Evaluation metrics and outputs of the Social Factors Index (SFI), including per-imputation performance, pooled summaries, out-of-fold predictions, and calibrated VAS scores.

- **psm_results.xlsx –** Full outputs of the propensity score matching analysis (Q4 vs. Q1 of the Social Factors Index), including covariate balance and ATT estimates across imputations.

- **Balance_tables.xlsx–** Detailed covariate balance tables before and after propensity score matching, reported separately for each imputation and pooled across imputations

- **Boxplot.docx –** Visualizations of VAS pain by categorical and numerical predictors, with boxplots and kernel density estimates.

- **caliper_sensitivity.xlsx –** Sensitivity analysis of propensity score matching using alternative caliper widths (0.2 vs. 0.3 SD) on the logit of the propensity score.

- **country_meta.xlsx –** Results of stratified propensity score matching analyses by region (Spain vs. non-Spain), including stratum-specific pooled ATT estimates.

- **dataanalysis.docx –** Extended Data Analysis methods, including multiple imputation, SFI construction/validation, mutual information setup, PSM specifications (trimming/calipers/matching), balance diagnostics, doubly robust models, and stratified analyses.

- **Descriptive.docx –** Summary tables of baseline demographic, social, and clinical characteristics by registry (REGISPONSER and RESPONDIA), including distributions and missingness.

- **regression_psm_dr.xlsx** – Results of multivariable regression analyses performed after propensity score matching. For each imputed dataset, the file reports the treatment effect estimate (β for Q4 vs. Q1 of the Social Factors Index) with cluster-robust standard errors by matched pair, 95% confidence intervals, and the list of covariates included when residual imbalance was observed

Acknowledgements: The Authors would like to thank all the investigators from the REGISPONSER and RESPONDIA study groups. We also like to acknowledge the support of technological tools such as ChatGPT, whose assistance has helped improve the fluency and clarity of English language expression in some sections of this document.

Variable	Dataset (N=2042)	REGISPONSER (N=1514)	RESPONDIA (N=528)
Demographics			
Age, years	47.3 ± 13.4	48.1 ± 12.9	44.7 ± 14.7
Diagnostic delay, years	7.5 ± 9.3	7.7 ± 9.4	6.8 ± 9.1
Disease duration, years	19.5 ± 13.1	21.1 ± 13.1	14.8 ± 11.6
Symptom duration, years	11.7 ± 10.3	13.2 ± 10.4	7.4 ± 8.5
Age at diagnosis, years	35.4 ± 12.7	34.9 ± 11.9	37.0 ± 14.9
Weight, kg	73.8 ± 14.2	74.2 ± 14.2	72.4 ± 14.2
Height, cm	166.4 ± 9.3	166.7 ± 9.1	165.4 ± 10.0
Clinical measures			
Chest expansion, cm	3.6 ± 2.1	3.8 ± 2.2	3.3 ± 1.9
Schober test, cm	3.0 ± 1.9	3.0 ± 1.8	3.2 ± 2.1
Finger-to-floor distance, cm	19.0 ± 15.0	18.6 ± 14.2	20.2 ± 16.8
Occiput-to-wall distance, cm	4.5 ± 6.2	4.4 ± 6.0	4.9 ± 6.6
Swollen joint count	0.8 ± 2.9	0.4 ± 1.8	1.9 ± 4.8
VAS (pain 0–10)	3.3 ± 2.3	3.2 ± 2.1	3.7 ± 2.6
ASQoL (0–18 higher=worse)	6.6 ± 5.2	6.4 ± 5.1	7.4 ± 5.3
SF-12 Physical Component	37.0 ± 7.5	37.2 ± 7.5	36.3 ± 7.5
SF-12 Mental Component	50.2 ± 5.5	50.6 ± 5.5	49.3 ± 5.5
ASDAS-CRP	2.4 ± 1.1	2.6 ± 1.1	2.0 ± 1.0
Form			
Axial	1233 (60.9%)	1055 (69.8%)	178 (34.6%)
Mixed	744 (36.7%)	433 (28.7%)	311 (60.5%)
Peripheral	39 (1.9%)	19 (1.3%)	20 (3.9%)
Enthesitic	9 (0.4%)	4 (0.3%)	5 (1.0%)
Family history			
No	1513 (79.5%)	1117 (79.6%)	396 (79.2%)
Yes	390 (20.5%)	286 (20.4%)	104 (20.8%)
Gender			
Men	1485 (72.7%)	1131 (74.7%)	354 (67.0%)
Women	557 (27.3%)	383 (25.3%)	174 (33.0%)
Education			
High school	606 (41.6%)	359 (38.1%)	247 (47.9%)
Elementary school	556 (38.1%)	433 (46.0%)	123 (23.8%)
Illiterate	20 (1.4%)	13 (1.4%)	7 (1.4%)
University	276 (18.9%)	137 (14.5%)	139 (26.9%)
Marital status			
Married	1023 (70.2%)	713 (75.6%)	310 (60.2%)
Single	364 (25.0%)	194 (20.6%)	170 (33.0%)
Divorced/Separated	48 (3.3%)	25 (2.7%)	23 (4.5%)
Widowed	23 (1.6%)	11 (1.2%)	12 (2.3%)

Profession			
Unskilled worker	377 (25.9%)	238 (25.3%)	139 (27.0%)
Skilled worker	349 (24.0%)	283 (30.1%)	66 (12.8%)
Employee	309 (21.3%)	201 (21.4%)	108 (21.0%)
Technician	219 (15.1%)	129 (13.7%)	90 (17.5%)
University profession	200 (13.8%)	88 (9.4%)	112 (21.7%)
Source of income			
Salary	1022 (70.3%)	643 (68.5%)	379 (73.6%)
Donations	216 (14.9%)	150 (16.0%)	66 (12.8%)
Profits	113 (7.8%)	61 (6.5%)	52 (10.1%)
Rents	103 (7.1%)	85 (9.1%)	18 (3.5%)
Housing conditions			
Good housing	735 (50.5%)	463 (49.2%)	272 (53.0%)
Excellent (no luxury)	588 (40.4%)	411 (43.6%)	177 (34.5%)
Deficient	86 (5.9%)	47 (5.0%)	39 (7.6%)
Excellent with luxury	24 (1.6%)	9 (1.0%)	15 (2.9%)
Inadequate	22 (1.5%)	12 (1.3%)	10 (1.9%)
Graffar scale			
Middle-lower class	594 (41.0%)	411 (43.9%)	183 (35.7%)
Middle class	443 (30.6%)	288 (30.8%)	155 (30.2%)
Upper-middle class	285 (19.7%)	166 (17.7%)	119 (23.2%)
Lower class	95 (6.6%)	59 (6.3%)	36 (7.0%)
Upper class	32 (2.2%)	12 (1.3%)	20 (3.9%)
Race			
White	1193 (81.4%)	930 (98.3%)	263 (50.6%)
White-Indigenous	201 (13.7%)	2 (0.2%)	199 (38.6%)
White-Black	23 (1.6%)	3 (0.3%)	20 (3.8%)
Other	18 (1.2%)	7 (0.7%)	11 (2.1%)
Black	13 (0.9%)	1 (0.1%)	12 (2.3%)
Indigenous	10 (0.7%)	1 (0.1%)	9 (1.7%)
Black-Indigenous	5 (0.3%)	1 (0.1%)	4 (0.8%)
Indigenous-Asian	3 (0.2%)	1 (0.1%)	2 (0.4%)
Employment status			
Employed	1038 (51.6%)	738 (49.5%)	300 (57.6%)
At home	580 (28.8%)	525 (35.2%)	55 (10.6%)
Unemployed	309 (15.4%)	182 (12.2%)	127 (24.4%)
Student	19 (0.9%)	5 (0.3%)	14 (2.7%)
Disability			
No disability	1417 (72.0%)	1047 (71.0%)	370 (75.2%)
Permanent disability	444 (22.6%)	367 (24.9%)	77 (15.7%)
Temporary disability	106 (5.4%)	61 (4.1%)	45 (9.1%)
Country			
Spain	1514 (74.1%)	1514 (100.0%)	0 (0%)
Argentina	147 (7.2%)	0 (0%)	147 (27.8%)

Mexico	103 (5.0%)	0 (0%)	103 (19.5%)
Portugal	88 (4.3%)	0 (0%)	88 (16.7%)
Chile	76 (3.7%)	0 (0%)	76 (14.4%)
Venezuela	38 (1.9%)	0 (0%)	38 (7.2%)
Peru	32 (1.6%)	0 (0%)	32 (6.1%)
Uruguay	28 (1.4%)	0 (0%)	28 (5.3%)
Costa Rica	16 (0.8%)	0 (0%)	16 (3.0%)

*Values are presented as mean ± standard deviation or n (%). Percentages use the variable-specific denominator due to missing data. Abbreviations: VAS = Visual Analogue Scale; ASQoL = Ankylosing Spondylitis Quality of Life; ASDAS = Ankylosing Spondylitis Disease Activity Score.